Published in final edited form as:
Clin Neuropsychol. 2011 May ; 25(4): 670–684. doi:10.1080/13854046.2011.564209.

Developing a Short Form of Benton's Judgment of Line Orientation Test: An Item Response Theory Approach

Matthew Calamia[1,*], **Kristian Markon**[1], **Natalie L. Denburg**[1,2], and **Daniel Tranel**[1,2]

[1] Department of Psychology, University of Iowa

[2] Department of Neurology (Division of Behavioral Neurology and Cognitive Neuroscience), University of Iowa College of Medicine

Abstract

The Judgment of Line Orientation (JLO) test was developed to be, in Arthur Benton's words, "as pure a measure of one aspect of spatial thinking, as could be conceived." The JLO test has been widely used in neuropsychological practice for decades. The test has a high test-retest reliability (Franzen, 2000), as well as good neuropsychological construct validity as shown through neuroanatomical localization studies (Tranel, Vianna, Manzel, Damasio, & Grabowski, 2009). Despite its popularity and strong psychometric properties, the full-length version of the test (30 items) has been criticized as being unnecessarily long (Straus, Sherman, & Spreen, 2006). There have been many attempts at developing short forms; however, these forms have been limited in their ability to estimate scores accurately. Taking advantage of a large sample of JLO performances from 524 neurological patients with focal brain lesions, we used techniques from Item Response Theory (IRT) to estimate each item's difficulty and power to discriminate among various levels of ability. A random item IRT model was used to estimate the influence of item stimulus properties as predictors of item difficulty. These results were used to optimize the selection of items for a shorter method of administration which maintained comparability with the full form using significantly fewer items. This effectiveness of this method was replicated in a second sample of 82 healthy elderly participants. The findings should help broaden the clinical utility of the JLO and enhance its diagnostic applications.

Among the most notable of Arthur Benton's numerous contributions to the field of neuropsychology are the tests he developed to measure various visuospatial abilities. Measurement of these abilities had been underserved, perhaps because of the long entrenched focus on language-related abilities in neuropsychological assessment, and Benton's tests were a welcome addition and were rapidly embraced by the field. The tests remain widely used in contemporary practice (Franzen, 2000). One such measure is the Judgment of Line Orientation (JLO) test. This test ranks within the overall top 100 instruments used by neuropsychologists, and was ranked 11th in the category of "return to work" instruments in one survey (Rabin, 2001). In another survey, 42% of neuropsychologists reported administering the JLO "regularly" (Butler, Retzlaff, & Vanderploeg, 1991).

The JLO test was developed to be "as pure a measure of one aspect of spatial thinking as could be conceived" (Benton, Sivan, Hamsher, Varney, & Spreen, 1994). Two partial line segments are presented together on one page, and the examinee is asked to match the orientation of these segments to those on a multiple choice response card (Figure 1,

[*] Corresponding author: Matthew Calamiam, Department of Psychology, E11 SSH, Iowa City, Iowa, 52242, matthew-calamia@uiowa.edu.

modified from Lezak, Howieson, & Loring., 2004, p. 390). The response options are made up of 11 full lines, all 18 degrees apart from one another, arranged in a semi-circle. The stimulus lines—partial line segments—represent either the proximal (“low”, “L”), middle (“M”) or distal (high, “H”) segment (one-third) of the full lines. The examinee is presented with five sample items, on which erroneous responses are corrected, followed by 30 test items presented without feedback (Benton, Varney, & Hamsher, 1978).

In patient populations, the test-retest reliability of the JLO has been found to be as high as 0.90 (Franzen, 2000). The level of internal consistency of the JLO has also been found to be high across relevant samples: α=0.84 in a neuropsychiatric sample (Winegarden, Yates, Moses, & Faustman, 1997); α=0.90 in a neuropsychological rehabilitation sample (Qualls, Bliwise, & Stringer, 2000); and α=0.85 in a mixed neurological and psychiatric sample (Woodard, Benedict, Roberts, Goldstein, Kinner, Capruso, & Clark, 1996).

The JLO test was originally developed to detect “right hemisphere dysfunction” (Benton et al., 1978), although the test is often used for broader (or narrower) purposes in contemporary neuropsychological practice. A large-scale lesion study found that failure on the JLO was most strongly associated with lesions in the right posterior parietal and occipitoparietal regions, areas within the so-called “where” dorsal visual stream (Tranel, Vianna, Manzel, Damasio, and Grabowski, 2009). Because it measures a relatively basic level of visuospatial ability, the JLO can be useful interpreting a patient’s performance on more complex tasks of visual reasoning and visuoconstruction (cf. Lezak et al., 2004). It also has predictive validity in assessing complex skills such as driving-related abilities (Mathias & Lucas, 2009). Clinicians and researchers working with patients with movement disorders find the test useful as it only requires verbal responses (Montse, Pere, Carme, Francesc, & Eduardo, 2001), avoiding contamination from constructional and motor speed factors

As noted, the JLO test is widely used and has strong psychometric properties and clinical utility; nonetheless, it has been criticized for being long. In one survey of neuropsychologists, the total time required to administer, score, interpret, and report on the results of the JLO was estimated to be roughly 20 minutes (Lundin & DeFilippis, 1999). A strong form of criticism notes that the 30 item length is simply unnecessary, given that the test measures a basic visuospatial ability (Straus et al., 2006). Such criticisms may have some validity, especially given the fact that there is increasing pressure on neuropsychologists to conduct efficient, economic assessments that maximize the information gained from testing (Sweet, Westerberg, & Moberg, 1995). Moreover, in a time when neuropsychologists often do not receive full reimbursement for their hourly services (Kanauss, Schatz, & Puente, 2005), there is an even more pressing need to complete assessments in an efficient and economic manner. Administering tests—or items from tests —that do not have high diagnostic yield is simply a luxury that neuropsychologists cannot afford. For these and related reasons, there have been many attempts at developing short forms of the JLO; however, these forms have been limited in their ability to estimate scores accurately.

Based on evidence of high split-half reliability, both Woodard et al. (1996) and Vanderploeg et al. (1997) used odd-even splits of the JLO items to create short forms, and these two forms predicted the full form score within two points in 75% and 77.7% of the cases, respectively. However, a frequency distribution of errors in Woodard et al. (1996) showed prediction errors of up to six points, and given that the total possible correct score on the JLO test is 30 points, this is a large margin of error (20% of all of the test items). The distribution of errors was not reported in Vanderploeg et al. (1997).

Winegarden, Yates, Moses, and Benton (1998) compared 15-item short forms of JLO based on odd-even splits of the items, a 10-item form based on items 1–10, and two 20–item forms based on items 1–20 and 11–30, respectively. Although the distribution of errors was not reported, the authors recommended the form with items 11–30 "for clinical use in situations in which employment of the full form may not be advisable." The short forms developed by Qualls et al. (2000), forms "Q" and "S," which used the item difficulty levels in the patient sample to create two parallel forms, misclassified 10% of study participants. The authors concluded that their short forms "do not categorize severity of visuospatial impairment in equivalent fashion as the original," and therefore recommended the forms be used as screeners of visuospatial impairment.

With one exception (Qualls et al., 2000), these short forms have been created on the assumption that JLO items are all comparable or at least increase in difficulty in a linear way as the test progresses. However, the stimulus characteristics of individual items, which likely contribute to differences in difficulty, may very well not be distributed in this manner. For example, the orientation of the two lines presented in a stimulus item is one potential source of differences in difficulty across items. There is empirical support for this—in a modified version of the JLO, Collaer and Nelson (2002) found that performance decreased for lines midway between the horizontal and vertical lines. It is easier to discriminate the orientation of lines that fall directly on the horizontal or vertical axis (lines 1, 6, and 11 of the JLO test) than those that lie between those axes. This effect, known as the oblique effect, has long been noted in the visual processing literature (Appelle, 1972). Also, lines on the left side of the page are easier to discriminate than those on the right. This has been attributed to the specialization of the right hemisphere for processing visuospatial material (Eden, Stein, Wood, & Wood, 1996), although others have argued this may be an artifact of the particular items on the test (Treccani, Torri, & Cubelli, 2005).

These item differences may influence not only item difficulty, but also item discrimination, or the ability of an item to differentiate among different levels of the trait(s) underlying JLO performance. For example, an item may be found to be difficult such that only 10% of participants answer it correctly. However, if that 10% included both participants who scored low and high on the JLO, then the item would be a poor discriminator of ability. In contrast, if the 10% included only the top performers, then the item would be highly discriminating. It matters not only how many participants answer an item correctly, but also how those participants perform on the rest of the test. Should discriminations differ significantly across items, they may be a useful criterion to consider because maximizing discrimination leads to more accurate estimates of ability. Also, when items do not differ significantly in their discrimination, other criteria can be used to create short forms (see Methods).

In this study, we used item response theory (IRT) to develop a short form of the JLO. We utilized contemporary IRT methodologies that allow for modeling of item psychometric properties in terms of stimulus characteristics. Stimulus characteristics of items, such as the line orientation of the two lines presented in each stimulus item, were investigated to determine to what extent they account for differences in item difficulty. This allowed us to create an abbreviated form of the JLO that was significantly shorter but produced scores highly similar to the full form.

Methods

Participants

The sample used to create the short form consisted of participants tested under the auspices of the Iowa Neurological Patient Registry in the Division of Behavioral Neurology and Cognitive Neuroscience at the University of Iowa (the "Patient Sample"). For inclusion in the Patient Registry, participants had to have a focal, stable brain lesion. Most of the patients' lesions resulted from stroke, with fewer due to other etiologies including surgical resection, anoxia/ischemia, and herpes simplex encephalitis. This patient sample is neuroanatomically homogeneous, however, in the sense that all of the patients have focal lesions with distinct borders that are not changing over time. The sample does not include patients with diffuse, progressive, or otherwise non-focal or non-stable brain damage. All participants were tested by trained neuropsychology technicians during the chronic epoch of their lesion onset (3 months or more post lesion onset). Records from 524 patients from the Patient Registry (47% women) were analyzed. The average age of these participants was 47 years (SD=16), and the average level of education was 13.4 years (SD=2.6). The mean JLO score in the sample was 24.0 (SD=4.7). The sample contained patients with a range of ability levels, with scores ranging from 4 to 30.

A second sample (the "Elderly Sample") was drawn from a study of older adults ranging in age from 60–85 (n=82; mean age, 73.3 (SD=6.8); mean education, 13.8 (SD=2.4); 55% female) (Denburg, 1997). The sample was composed of community-dwelling adults recruited through churches and community organizations. All adults lived independently and can be considered "able" elderly (mean Mini-Mental State Examination score of 28 (SD=1.6) (MMSE; Folstein, Folstein, & McHugh, 1975)). Individuals were excluded if they had a history of cerebrovascular disease, closed head injury with loss of consciousness greater than five minutes, or greater than mild depression as assessed by the Geriatric Depression Scale (GDS; Yesavage et al., 1983). The mean JLO score in the sample was 24.8 (SD=4.2, Range: 15–30).

Assessments

The JLO test (Benton et al., 1978) had been previously administered to both the patient and elderly samples, as described above. Responses to the 30 items on the JLO were manually entered into a computerized database for analysis. Consistent with test procedures, both stimulus lines have to be correctly identified in order to receive a raw score of 1 for each item (total possible score = 30 points). A raw score of 0 for an item is given when either one or none of the stimulus lines in the item is correctly identified.

In the patient sample, if a participant completed the JLO test on more than one occasion, the first chronic epoch administration was used. Because Forms V and H of the JLO test include the same items, both were combined for analysis. Participants in the elderly sample were administered the JLO on only one occasion.

Analysis

On many neuropsychological tests, items are summed together to create a total score. Typically, this total score then serves as a measure of a person's ability on the construct being measured. On the JLO, correct responses to each of the 30 individual items are summed to create a total performance score. The underlying (but untested) assumption is that each item is a good measure of the underlying construct.

Item Response Theory (IRT) explicitly models the relationship between answering specific test items and a person's standing on the ability being measured by a test. The ability being measured is conceptualized as a latent variable which influences the person's pattern of test performance; persons with higher ability levels have higher probabilities of answering questions correctly than those with lower ability levels. The likelihood of responding to a test item in a certain way (e.g., "correct") is modeled as a mathematical function of two components: 1) item characteristics, such as difficulty, and 2) person characteristics, such as a person's ability level on the trait being measured. Depending on the model, different techniques can be used to estimate these components (Embretson & Reese, 2000).

IRT models vary in how they estimate the relationship between test items and the underlying ability being measured. In a one-parameter model, items differ only in their estimated difficulty. Difficulty can be defined as the point on the underlying ability continuum at which the probability of successfully answering an item is 50%; items that are more difficult require more of the ability to answer correctly. In contrast, a two-parameter model allows items to vary in difficulty as well as discrimination. Discrimination is a parameter, analogous to a factor loading, reflecting the ability of an item to distinguish among trait levels (Embretson & Reese, 2000). Discrimination can be useful in identifying problematic test items. For example, one may want to eliminate very difficult but poorly discriminating items. These items are correctly answered by few individuals (high difficulty), but the individuals who obtain correct answers are not the same individuals who go on to achieve the highest overall scores on the test (poor discrimination). In our study, both one- and two-parameter item response theory models were estimated for the JLO using the patient sample. This was done using the 'ltm package' in the statistical program R (Rizopoulos, 2006). The models were compared using the Bayesian Information Criterion (BIC) which allows for comparisons of models with different numbers of parameters (Schartz, 1978).

We used a random item IRT model, to model item characteristics as predictors of difficulty. This was implemented using a generalized linear mixed model, estimated with a different R package ('lme4'; Bates, 2007). Participants and items were modeled as random effects, and item characteristics, such as the orientation of stimulus lines, were considered fixed effects. This is equivalent to a one-parameter IRT model where items may vary randomly (e.g., as a function of stimulus characteristics), as opposed to traditional IRT models where items are considered fixed (De Boeck, 2008).

Results

Short Form Creation

The 1-paramater model (BIC=14452) fit the data better than the 2-parameter model (BIC=14560). The estimates from the 1-parameter model are reported in Table 1 and are used in all subsequent analyses. The test information curve is shown in Figure 2. This curve displays how well the JLO test can measure different levels of the latent ability underlying test performance. In contrast to other methods which use single estimates of measurement precision for a test (e.g., coefficient alpha), in IRT measurement precision is conceptualized as a continuous variable (information) which can differ across ability levels. Higher information reflects better measurement precision, and is inversely related to the standard error of measurement at a given ability level. Because the JLO was designed to assess impairment in a basic ability (as opposed to aptitude), it is not surprising that the test best measures trait scores at or below the mean ability of the sample (which is set to 0)—in fact, the measurement precision of the test is best in exactly the range one would want for neuropsychological assessment, roughly one to two SDs below the mean. The information curve reflects the fact that almost all of the individual JLO items have difficulties below the mean. Only 2 of the 30 JLO items (items 24 and 27) have difficulty estimates above the mean (Table 1).

Because the 1-parameter model was used, only *item difficulty* was used to create a short form. After arranging items in order of difficulty (see Appendix A), various basal and ceiling rules were applied in order to determine the optimal combination needed to create a short form comparable to the full version. Basal rules set a "floor" for an examinee's performance. An examinee is required to answer a specified number of items correctly and once this criterion is reached, the examinee receives credit for early easier test items not presented. For example, for a basal rule of two items, if an examinee were to miss the first item administered, administration of test items would precede in reverse order until two items are successfully answered. Ceiling rules define the examinee's expected maximum performance. After an examinee incorrectly answers a specified number of items, test administration is terminated, with the expectation that the examinee would not successfully answer any of the remaining more difficult items. Basal and ceiling criteria help to economize assessment by avoiding administration of many items that are too easy or too difficult for a given examinee.

By starting with the 16th most difficult item (item 19 of Form V), and having basal and ceiling rules set at 6 items, comparable full form estimates were obtained using an average of 20.4 items (SD=5.4)—a reduction of almost 10 items, or approximately a third, from the full-length 30-item JLO test. Applying the basal and ceiling rules to the reordered form, the average difference between full-form and short-form scores was 0.60 (SD=1.09) (note that this represents a fraction of an item). The Pearson correlation between the short form of administration and full form was 0.97. The same item reordering and basal and ceiling rules determined in the patient group were applied to the elderly sample. In this sample, the level of item reduction (20.7 items, SD=4.65) and the average full-form versus short-form score difference (0.59 item, SD=1.14) were very similar to those of the patient group. Also, the Pearson correlation between the short form of administration and full form in this sample was 0.96, nearly identical to the patient sample.

In clinical use, a neuropsychologist may only want to know whether a patient can pass a threshold for intact performance on the JLO, and may not be interested in differentiating between "average" and "above average" performance. One commonly used cut-off score for intact vs. impaired performance on the JLO test is 21 (e.g., Lezak et al., 2004; Strauss et al., 2006). When this cut-off was applied in a lesion study, impairment was associated with damage to right posterior parietal and occipitoparietal regions, areas associated with the type of visuospatial processing the JLO was designed to measure, indicating that this cut-off has criterion validity (Tranel et al., 2009). Applying this cut-off to the proposed short form, correctly answering just the first 6 items (beyond the starting point of the 16th item) would yield enough information to predict a non-impaired score. About 28% of the patient sample had this response pattern. To put it another way, determining that these patients (the approximately 28%) had unimpaired performance required only 6 items, rather than the 30 that were administered. This illustrates the sizable savings of administration time that can be obtained with this short form.

Comparison with Previous Short Forms

Although other shorts forms have been shown to correlate fairly highly with the full form of the JLO, the actual score differences can be large, and this limits their ability to categorize impairment as accurately or in the same way as the full form (Straus et al., 2006). In the patient sample, with the cut-off score of 21, classifications based on the newly proposed short form of administration differed from the full form in only 3% of cases. The majority of these differences were ones in which the predicted score was just slightly above the cut-off score for normal performance while the actual score was a point or two below the cut-off. In contrast, scoring according to other published short forms yielded differences in classifications ranging from 8 to 11%, with many false positives (Column 1 in Table 2). To ascertain how well our current short form fared, compared to other previously developed short forms, we calculated an exact test for correlated proportions. Short form outcomes were compared to the full form only for those cases in which the two short forms (ours versus the other) reached opposite conclusions. Our new short form was statistically superior to each of the previous short forms (p <.05).

In the elderly sample, with the newly developed short form, short form vs. full form scores yielded different classifications for only 2% of cases. Similar to the patient sample, these were participants whose estimated score was one or two points above the cut-off while their obtained score was at or slightly below the cut-off. By contrast, classifications based on other short form scores differed from full form classifications in 4 to 10% of cases (Column 2 in Table 2). The new short form was statistically superior to two of the previous short forms (“even items” and “form ‘Q’”) (p<.05) using the exact test for correlated proportions. Although the new short form obtained more correct full form classifications than the three remaining short forms, these differences were not statistically significant (p>.05).

Modeling Predictors of Item Difficulty

Differences in the difficulty of the JLO items were hypothesized to vary as a function of item characteristics, such as the specific line orientations presented within each item. As noted previously, the JLO stimuli are drawn from 11 possible line segments. Based on previous research using a similar task (Collaer & Nelson, 2002), it would be expected that starting with line 1, the difficulty is relatively easy, increases as lines become more oblique, and decreases as one approaches line 6, a vertical line. This pattern would repeat on the right side of the stimulus figure. Because of this type of nonlinear relationship between the angle and difficulty, a cosine transformed function was used to predict difficulty of each individual stimulus line. This deviation of both stimulus lines of a JLO item from the nearest horizontal or vertical axis, and the position of the lines on the left or right side of page, were both significant predictors of item difficulty. The height of the line (proximal, middle, or distal) was not a significant predictor. The results for the model (not including height) are presented in Table 3. These three significant predictors accounted for 45% of the variance in item difficulty. This level of variance accounted for is similar to other studies which have modeled item characteristics as predictors of difficulty (Embretson & Daniel, 2008; Hornke & Habon, 1986).

Discussion

Originally developed in 1978, the JLO remains a popular test in neuropsychological assessment (Rabin, 2001). It is used to assess visual spatial reasoning ability at a fairly basic level, and it can also be helpful in the interpretation of a patient’s performance on more complex tasks of visual reasoning and visuoconstruction. Despite its clinical utility, the length of the JLO has been criticized—with 30 items, the test can be tedious especially in poor-performing patients, and may not provide efficient, economic assessment of visual spatial reasoning. A shorter test would likely lead to increased use and more efficient assessment. Such a test would undoubtedly be welcomed by patients, some of whom find the length of the current test fatiguing.

In the current study, we used Item Response Theory (IRT) to create a short form of the JLO test that would have high precision and utility. In IRT, one can estimate an item’s difficulty and discrimination. Choosing highly discriminating items for a short form is one method of maximizing the measurement of ability. However, we found that JLO items did not differ significantly in their discriminations, leading us to consider other criteria.

Rather than creating a fixed short form in which the same items are administered to all participants, as has been proposed in the past for short forms of the JLO test, we used basal/ceiling rules to create a shorter method of test administration. Basal/ceiling rules are used in many popular neuropsychological tests, such as the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1983), and they are used in the Peabody Picture Vocabulary Test-4 (Dunn & Dunn, 2006) and in the newest version of the Wechsler Adult Intelligence Scale (WAIS-IV; PsychCorp, 2008). In the JLO short form we created, this flexible method of administration is used: the examiner begins at the 16th most difficult item and establishes a basal and ceiling of 6 items passed (basal) or failed (ceiling). With these criteria, nearly a third less items would need to be administered in order to obtain scores highly comparable to those on the full form. Furthermore, this flexible method of administration can be modified depending on the investigator's clinical or research goals. For example, because the JLO measures a basic visuospatial ability, a clinician may not be interested in distinguishing among scores above some cut-off value for intact performance. It may not be necessary to know whether a patient has "above average" or "superior" performance; rather, the clinician may simply want to know whether a patient is impaired on the test or not. With this goal in mind, once a basal is established, and a patient obtains an estimated score above the cut-off value for intact performance, administration can be discontinued. For example, if the clinician is using a cut-off score of 21, and a patient correctly answers the first six items on the new form, the patient's predicted score will be at least 21, and administration could be terminated. The same information would be obtained with six instead of thirty items. Conversely, if a patient misses the first item, and continues to miss items as administration proceeds in reverse order, the patient may reach a point where a score above 21 is unobtainable, and administration could be terminated—again, this would have been accomplished with substantially fewer than all 30 items.

Several attempts have been made to create a short form of the JLO test; however, the failure of these forms to predict performance on the full JLO adequately has led most authors to advise their use only for screening purposes (e.g., Qualls et al., 2000). In the current study, these short forms were found to differ from the full form in their classification of impairment in 4 to 11% of cases, depending on the sample and specific short form examined. In contrast, using the reordered form and basal and ceiling rules proposed in the current study, the difference in classification (compared to the full-length version) was only 2 to 3%. The superior performance for the new proposed short form over previous short forms was statistically significant for all comparisons in the patient sample and for two of five comparisons in the elderly sample. This newly proposed flexible method of administration maximizes the balance between shortening test administration and maintaining measurement precision.

Two stimulus characteristics, the orientation of the two lines presented in a stimulus item and the location of the lines on the left or right side of the page, partially accounted for differences in difficulty across items. The use of different partial line segments (i.e., proximal, middle, or distal) did not account for variability in item difficulty. When the JLO was devised, partial line segments were chosen for the stimulus lines because full line segments were considered too easy (Benton et al., 1978). However, the proximal, middle, and distal segments were not hypothesized to differ in difficulty.

Decomposing test performance as a function of item stimulus characteristics allows for the estimation of item difficulties for novel items. These values could also be used to generate new items in the context of computerized adaptive testing to efficiently estimate trait levels (Embretson & Reise, 2000). They could also be used to create tests of varying difficulty depending on the research or assessment goals. For example, although the majority of the items on the current JLO have difficulties below the mean, one could use the item stimulus information to generate more items with a higher level of difficulty. Or, if one wanted to maximize the ability of the score to distinguish performance at or below some cut-off, items could be generated which maximize the measurement precision at that value.

One limitation to this study is that the shorter form of administration was not implemented with a new set of participants—e.g., persons who had never received the JLO test before. Both the patient sample used for the initial IRT analysis and the elderly sample contained individuals who completed the full version of the JLO in the standard presentation order. It is possible that an examinee's performance may be different when the items are presented in the new order with the basal and ceiling rules. However, because the JLO contains several practice items to familiarize the examinee with the test, we suspect the reordering would not have much of an effect—nonetheless, this remains an empirical question for future research.

Although the measurement of cognitive abilities is central to the field of neuropsychology, the use of IRT or other latent trait models has not been widely adopted. This study illustrates one potential use of such methods for improving upon current assessment practices through broadening the clinical utility and diagnostic applications of existing measures, and maximizing their efficiency.

Acknowledgments

Supported by APA DPN Fellowship 5T32 MH18882 (M.C.) and NINDS P50 NS19632 (D.T.)

References

Bates, Douglas. lme4: Linear mixed-effects models using S4 classes. R package version 0.99875–9. 2007. URL http://CRAN.R-project.org

Benton, AL.; Sivan, A.; Hamsher, K.; Varney, N.; Spreen, O. Contributions to Neuropsychology Assessment: A Clinical Manual. 2. New York: Oxford University Press; 1994.

Benton AL, Varney N, Hamsher K. Visuospatial judgment: a clinical test. Archives of Neurology. 1978; 35:364–367. [PubMed: 655909]

Butler M, Retzlaff P, Vanderploeg R. Neuropsychological Test Usage. Professional Psychology: Research and Practice. 1991; 22:510–512.

De Boeck P. Random item IRT models. Psychometrika. 2008; 73:533–559.

Denburg, N. Doctoral Dissertation. Michigan State University; 1997. Attentional abilities in the able elderly: Examination of structure, correlates, and self-report.

Dunn, LA.; Dunn, LM. Peabody Picture Vocabulary Test—4. Circle Pins, MN: American Guidance Service; 2006.

Eden GF, Stein JF, Wood HM, Wood FB. Differences in visuospatial judgement in reading-disabled and normal children. Perceptual and Motor Skills. 1996; 82:155–177. [PubMed: 8668471]

Embretson, SE.; Reise, SP. Item response theory for psychologists. Lawrence Erlbaum; 2000.

Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research. 1975; 12:189–198. [PubMed: 1202204]

Frazen, M. Reliability and Validity in Neuropsychological Assessment. 2. Kluwer Academic/Plenum Publishers; New York: 2000.

Kanauss K, Schatz P, Puente AE. Current trends in the reimbursement of professional neuropsychological services. Archives of Clinical Neuropsychology. 2005; 20:341–353. [PubMed: 15797170]

Kaplan, E.; Goodglass, H.; Weintraub, S. The Boston Naming Test. Philadelphia: Lea & Febiger; 1983.

Lezak, MD.; Howieson, DB.; Loring, DW. Neuropsychological Assessment. 4. Oxford University Press; USA: 2004.

Lundin KA, DeFilippis NA. Proposed schedule of usual and customary test administration times. The Clinical Neuropsychologist. 1999; 13:433–436. [PubMed: 10806455]

Mathias JL, Lucas LK. Cognitive predictors of unsafe driving in older drivers: A meta-analysis. International Psychogeriatrics. 2009; 21:637–653. [PubMed: 19470197]

Montse A, Pere V, Carme J, Francesc V, Eduardo T. Visuospatial deficits in parkinsons disease assessed by judgment of line orientation test: Error analyses and practice effects. Journal of Clinical and Experimental Neuropsychology. 2001; 23:592–598. [PubMed: 11778636]

PsychCorp. Technical and interpretive manual. 4. San Antonio, TX: Pearson; 2008. Wechsler Adult Intelligence Scale.

Qualls C, Bliwise N, Stringer A. Short forms of the Benton Judgment of Line Orientation test: development and psychometric properties. Archives of Clinical Neuropsychology. 2000; 15:159–163. [PubMed: 14590559]

Rabin LA, Barr WB, Burton LA. Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA division 40 members. Archives of Clinical Neuropsychology. 2005; 20:33–65. [PubMed: 15620813]

Rizopoulos D. ltm: An R package for latent variable modeling and item response theory analyses. Journal of Statistical Software. 2006; 17:1–25.

Schwarz G. Estimating the dimension of a model. The Annals of Statistics. 1978; 6:461–464.

Strauss, E.; Sherman, EMS.; Spreen, O. A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. Oxford University Press; USA: 2006.

Sweet JJ, Westergaard CK, Moberg PJ. Managed care experiences of clinical neuropsychologists. The Clinical Neuropsychologist. 1995; 9:214–218.

Tranel D, Vianna E, Manzel K, Damasio H, Grabowski T. Neuroanatomical correlates of the benton facial recognition test and judgment of line orientation test. Journal of Clinical and Experimental Neuropsychology. 2009; 31:219. [PubMed: 19051129]

Treccani B, Torri T, Cubelli R. Is judgement of line orientation selectively impaired in right brain damaged patients? Neuropsychologia. 2005; 43:598–608. [PubMed: 15716150]

Vanderploeg R, LaLone L, Greblo P, Schinka J. Odd-even short forms of the Judgment of Line Orientation test. Applied Neuropsychology. 1997; 4:244–246. [PubMed: 16318474]

Winegarden B, Yates B, Moses J, Benton AL, Faustman W. Development of an optimally reliable short form for judgment of line orientation. The Clinical Neuropsychologist. 1998; 12:311–314.

Winegarden BJ, Yates BL, Moses JA, Faustman WO. Development, validity and reliability analysis of short-forms of three bentonian perceptual tests. Archives of Clinical Neuropsychology. 1997; 12:430.

Woodard J, Benedict R, Roberts V, Goldstein F, Kinner K, Capruso D, Clark A. Short-form alternatives to the Judgment of Line Orientation test. Journal of Clinical and Experimental Neuropsychology. 1996; 18:898–904. [PubMed: 9157113]

Yesavage J, Brink T, Rose T, Lum Q, Huang Q, Adey V, Leirer V. Development and validation of a geriatric depression screening scale: A preliminary report. Journal of Psychiatric Research. 1983; 17:37–49. [PubMed: 7183759]

Appendix A. Revised JLO Ordering for New Short Form Administration

Table A-1

In order to administer the new short form, items from the original form (Form V) must be rearranged in the order given in Table A-1 below. Administration should START with item 16 of the new form (see arrow below). Six correct responses are required to establish a basal. Six incorrect responses are required to establish a ceiling.

New Item Number	Original Form V Item Number
1	4
2	11
3	3
4	6
5	5
6	16
7	9
8	2

	New Item Number	Original Form V Item Number
	9	12
	10	17
	11	10
	12	13
	13	7
	14	30
	15	8
→	16	19
	17	20
	18	14
	19	15
	20	1
	21	21
	22	22
	23	18
	24	23
	25	26
	26	28
	27	25
	28	29
	29	27
	30	24

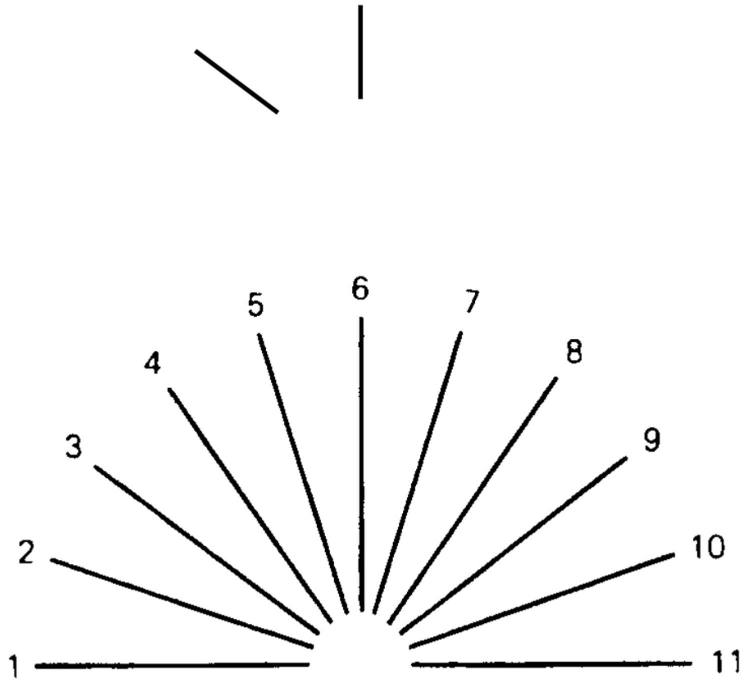

Figure 1.
Example of a JLO item.

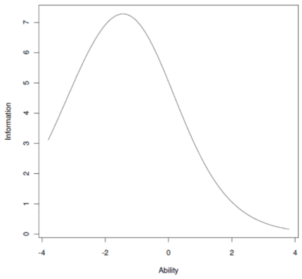

Figure 2.

Test Information Curve for the JLO Test.

“Ability” refers to the trait underlying JLO performance. “Information” reflects how well the JLO measures the level of ability. Higher levels of information reflect higher measurement precision.

Table 1

Difficulty estimates for the JLO items (Form V).

Item Number	Difficulty
1	−1.023
2	−2.237
3	−2.988
4	−3.172
5	−2.770
6	−2.921
7	−1.668
8	−1.381
9	−2.318
10	−1.898
11	−3.132
12	−2.008
13	−1.751
14	−1.104
15	−1.043
16	−2.585
17	−1.913
18	−0.963
19	−1.278
20	−1.267
21	−1.023
22	−0.973
23	−0.913
24	0.235
25	−0.467
26	−0.750
27	0.028
28	−0.557
29	−0.350
30	−1.412

Difficulty is defined as the point at which the probability of successfully answering an item is 50%. Higher negative numbers indicate easier items.

Table 2

Differences in Classification (Impaired vs. Unimpaired) Based on Other Short Forms

Study	Form	% Disagreement (Patient Sample)	% Disagreement (Elderly Sample)
Woodard et al., 1996	Odd items	10	5
Woodard et al., 1996	Even items	8	7
Winegarden et al., 1998	Items 11–30	8	4
Qualls et al., 2000	Form “Q”	10	10
Qualls et al., 2000	Form “S”	11	8

Table 3

Fixed effects of the GLMM for Item Characteristics as Predictors of Difficulty

	Estimate	Standard Error	Z Value	P Value
(Intercept)	2.251	0.360	6.261	<.01
Line 1 Location	0.654	0.202	3.236	<.01
Line 2 Location	0.395	0.199	1.980	<.01
Midangle	−0.009	0.004	−2.404	<.01

Note: Line 1 and 2 Location are equal to cosine(4 * ((angle made by the line and the horizontal axis*π)/180)). Midangle refers to the angle that bisects the angles created by lines 1 and 2 and the horizontal axis.