An analysis of the VOSP Silhouettes Test with neurological patients

THOMAS MERTEN

Abstract
An item analysis of the Silhouettes, part of the Visual Object and Space Perception Battery, was performed using the test protocols of 266 German-speaking neurological patients with a mean age of 54.8 years, all of them presenting some sort of brain pathology. The sample yielded a mean test score of 17.0 ($SD = 4.6$). The two subsets of 15 animals and 15 objects were only moderately correlated (0.45), so the inclusion into a single scale is questionable. Other reliability estimates were also rather low (0.62 to 0.77). Moreover, gross deviations in item difficulty were obtained with this sample; scoring rules were found to be insufficiently explicit. Despite moderate rank correlations with other instruments (Hooper VOT: 0.65; WAIS-R Block Design: 0.57; neuropsychological screening battery SKT: -0.45), the psychometric properties obtained with this sample must be considered to be insufficient.

Key words: Psychological Assessment; Neuropsychology; Object Perception; Reliability; Test

1 Correspondence concerning this article should be addressed to Thomas Merten, Klinikum im Friedrichshain, Klinik für Neurologie, Landsberger Allee 49, D-10249 Berlin, Germany. Tel. +49 30 4221 1663. Electronic mail may be sent via Internet to thomas.merten@vivantes.de
Despite an extensive body of knowledge about test construction and test analysis available to neuropsychologists, it appears that a number of instruments have never been thoroughly submitted to those procedures. As a consequence, item revisions based on empirical rather than rational analysis do not seem to be very common. Moreover, a verdict given by Snaith (1981, p. 512) continues to be true: “It is very rare for any further work to be undertaken in an attempt to improve a scale once it has been published, and work to compare the merits and drawbacks of various scales is an even rarer event.” Regular and critical test users, however, often know about the shortcomings of their instruments, sometimes much better than the authors themselves do.

In the context of some work done on the Hooper Visual Organization Test – VOT (Merten & Beal, 1999; Merten, 1999, 2002) it was predicted that a test like the Silhouettes from the Visual Object and Space Perception Battery – VOSP (Warrington & James, 1986, 1991a, b) would resort to very similar mental functions and, thus, yield high correlations with the VOT. Both tests contain a component of visual object perception and internalized object manipulation (recalling what Lorenz, 1943, called “Hantieren im Vorstellungsraum”, or handling in the imagination space).

Also, in both tests the answers are given verbally, so a component of naming ability is involved although research points to a rather minor influence of language-related skills on test performance (Paolo, Cluff, & Ryan, 1996; Paul, Cohen, Moser, Ott, Zwacki, & Gordon, 2001).

In a factor-analytic study (Merten, 2005) performed on data from a comprehensive neuropsychological test battery, numerous measures of visuospatial processing and attention loaded significantly on the first strong factor, which was interpreted as a global dimension of non-verbal cognitive functions. The Silhouettes test also loaded highest on this first factor. This underlined the assumption that naming abilities are of relatively minor importance for this task.

When further analyzing the data of that study, it was found that:

1) The correlations between VOT and the Silhouettes amounted to only 0.64 and were not as high as predicted, whereas VOT and WAIS-R Block Design correlated at 0.71. Although this difference of correlations is not significant (z = -1.59, according to Olkin’s procedure as described in Bortz, 1999) one has to bear in mind that, in contrast to Block Design, both VOT and the Silhouettes neither involve an overt motor performance, nor are they speed tests.

2) During test sessions it was repeatedly observed that there were individuals who performed quite well in one of the two Silhouettes subsets, animals or objects, but quite badly on the other one, without noticeable differences in motivation or other behavioral variables. In fact, the two sets of 15 items each (15 animals and 15 commonly known objects) which make up the test did not correlate sufficiently with each other (r = 0.41) to justify the inclusion into one test implicitly thought to be homogeneous. It has to be emphasized that this correlation is a measure of internal consistency, alias reliability.

3) According to the test manual, item position is determined by their difficulty, separately for the two subsets of items. However, with a sample of neurological patients certain items were considerably easier than their relative position would indicate, whereas others turned out to be very difficult.
These three observations, then, led to the decision to perform an extended item analysis of the Silhouettes test which, to the author’s knowledge, has never been done before.

The test consists of 30 drawings of animals and inanimate objects. They differ in the angle of view and the degree to which distinctive features can be identified. The task to be performed is to name the items. In order to minimize language difficulties, descriptions or gestures or other means of identification are explicitly permitted. Moreover, guessing is encouraged. In the manual (Warrington & James, 1991a, p. 11) it is stated that the test should be abandoned after five consecutive failures, separately again for the two subsets. There is, however, no indication of whether or not this rule was tested empirically and applied to the subjects whose data are presented in the manual. This would have a considerable impact on the establishment of the item order.

Low test scores are not limited to individuals with right hemisphere damage. Although impairment in visual recognition is stressed for patients with lesions in posterior regions of the right hemisphere, the data presented by the authors demonstrate a considerable overlap between different groups (normal sample, patient groups with right hemisphere vs. left hemisphere lesions).

The test is used as part of the Visual Object and Space Perception Battery which allows identifying two different domains of visual processing, but the tests can also be used singly. The authors state that neuropsychological assessment is incomplete without an assessment of object and space perception.

Published studies mostly focus on case reports and series of patients with smaller N (e.g., Bodenburg, 2000; Hittmair-Delazer, Sailer, & Benke, 1995; Sprengelmeyer, Young, Sprengelmeyer, Calder, Rowland, Perrett, Hömberg, & Lange, 1997; Uttner, Bliem, & Danek, 2002; Warrington & James, 1988; Warrington, 2000). A more recent analysis of the VOSP testing the performance of 111 healthy older persons centered on validating Warrington’s theory of visual processing as a two-dimensional issue, which is space perception as opposed to object perception (Rapport, Millis, & Bonello, 1998).

With the same sample, the psychometric properties of the VOSP were analyzed (Bonello, Rapport, & Millis, 1997). Some tests were found to have low internal consistency scores, in particular Progressive Silhouettes (0.27), Incomplete Letters (0.54), and Object Decision (0.58). Cronbach’s alpha for the Silhouettes subtest amounted to 0.78.

Nonetheless, as a result of their analyses, it was stated that “the VOSP exemplifies a good test: It began as a theory-driven approach to assessment, and studies of both group differences and the latent structure of the battery are consistent with the theory on which it was based.” (Rapport et al., 1998, p. 219) The authors expect from their results that, in a psychometrical perspective, the “object perception subtests show considerable promise in measuring visuoperception” (p. 219). This is to be investigated in the following analysis, albeit limited to only one of the VOSP subtests and to a heterogeneous population of neurological patients with very diverse sorts of cerebral pathology.
Method

Sample

The VOSP Silhouettes test protocols of an unselected group of 266 German-speaking neurological in-patients were retrospectively analyzed. The patients had undergone neuropsychological assessment. The factor-analytic study presented by Merten (2005) was performed on a subsample of those Ss who were included in the present study. The group was composed of 168 males and 98 females with a mean age of 54.8 years ($SD = 15.2$; range: 16 to 86) and, on average, 13.6 years of education ($SD = 3.6$). The primary pathology of the patients was traumatic in 14% of the cases, cerebrovascular in 53%, neoplastic in 5%, inflammatory in 8%, degenerative in 6% of the cases; the remaining 15% was made up of patients with various causes of brain pathology. According to the patient profile in an acute neurological ward, the sample was heterogeneous with regard to etiology, quality, and degree of neuropsychological dysfunction at the time of the test administration.

Lateralized brain damage could be assumed for 131 patients of the original sample of 266. 62 patients showed evidence of right-hemisphere brain damage, 69 were shown to have suffered left-hemisphere brain damage. The test scores of these two subsamples (which did not differ significantly in terms of age, gender and education) were compared in a separate analysis. The heterogeneity of the sample did not allow for more detailed analyses of such factors as specific etiologies or chronicity of symptoms.

Procedure

Patients were given the standard administration as described in the test manual. The discontinue rule after five consecutive failures was not applied. Testing was performed individually, the answers being transcribed by the examiner. When several answers were offered (e.g., # 15: a goose, or a duck, or a swan), patients were asked to decide on the most likely.

The manual does not contain specified scoring rules, except for a list of the answers as they were originally conceived, so all answers given in its Appendix 2 are scored as correct. The published German version of the test (Warrington & James, 1992) is a strict translation of the original, with no changes made to the test material or scoring rules, as the translators emphasize. However, for adequate scoring a few principles have to be stated explicitly for this analysis: Answers were scored as correct if they were virtually synonymous with the listed ones, with respect to everyday German language (e.g., rabbit and hare), if names of young animals or diminutive expressions were employed (e.g., lamb or calf), if subcategories of the correct responses were given (e.g., winter shoe, polar bear or sun glasses). For item #8, other large reptiles (e.g., alligator and caiman) were accepted. Altogether, allowances with respect to the correct answer were highly constrained. Thus, for item # 15, only duck was scored as correct, as indicated by the manual, but not goose, turkey, or any other bird. If the correct category name (such as bird for # 15, reptile for # 8, or vehicle for # 23) was given, the patients were encouraged to specify their answer.

Following a flexible battery approach to neuropsychological assessment, a large number of additional test results were available. For those 200 patients whose test protocols were included in the factor analysis described by Merten (2005), scores on a complete core battery
of psychological tests were analyzed. To investigate item validity in the context of the pre-
sent analysis, the results for two other measures of visuospatial functions and for an unspec-
cific screening instrument were examined.

(1) The *Hooper Visual Organization Test* (Western Psychological Services, 1983), as out-
lined before, was thought to be conceptually closely related to the Silhouettes. After the
development of a VOT short version which was based on an empirical item analysis
(Merten, 2002), the 15-item version substituted the full scale in the neuropsychological
assessment done by the author. Full-scale results were available for 223 patients; the
short version was given to 42 patients. As the versions correlate at 0.95, full-scale score
predictions were computed for these 42 patients using the regression equation given by
Merten (2002). Thus, VOT scores for 265 individuals could be used.

(2) In contrast to the VOSP and the VOT, *WAIS-R Block Design* (Wechsler, 1981) is a timed
test which is considered to reflect visuospatial abilities to a high degree. Results were
available for 265 patients.

(3) 262 of the patients were given the *Syndrom-Kurztest* (SKT: Erzigkeit, 1989; cf. Overall
& Schaltenbrand, 1992) which is a short screening instrument yielding a rough estimate
of cognitive deficits, combining a number of speed tests and memory tasks. High SKT
scores reflect a high level of cognitive impairment.

**Results**

The sample yielded a mean total score for the Silhouettes test of 17.0 (*SD* = 4.6) with a
range from 6 to 28 points. The mean of the first half (animals) was 9.8 (*SD* = 2.8; range: 2 to
15), that of the second half (objects) was 7.2 (*SD* = 2.6; range: 0 to 15).

To investigate the effect of sex on total scale scores, a one-way analysis of variance was
performed. No significant effect appeared (*F*(1, 264) = 2.45, ns). A correlation of 0.24 (*p* <
.05) was obtained between the Silhouettes total score and years of education, and -0.45 (*p* <
.05) between the total score and age.

<table>
<thead>
<tr>
<th></th>
<th>Total test</th>
<th>1st half (animals)</th>
<th>2nd half (objects)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inter-item correlations</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.10</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td>Range</td>
<td>-0.11 to 0.34</td>
<td>-0.06 to 0.34</td>
<td>-0.06 to 0.28</td>
</tr>
<tr>
<td>Cronbach’s alpha</td>
<td>0.77</td>
<td>0.73</td>
<td>0.62</td>
</tr>
<tr>
<td>Split-half consistency</td>
<td>0.76&lt;sup&gt;1&lt;/sup&gt;</td>
<td>0.77&lt;sup&gt;2&lt;/sup&gt;</td>
<td>0.60&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Correlation between first and second half (animals vs. objects): 0.45
Full-scale split-half consistency (animals vs. objects): 0.62<sup>1</sup>

Notes: <sup>1</sup>Spearman-Brown formula, corrected for equal length; <sup>2</sup>corrected for unequal length
Table 2: Item analysis of the VOSP Silhouettes.

<table>
<thead>
<tr>
<th>Item</th>
<th>Per cent correct</th>
<th>Item-total correlation</th>
<th>Mean VOT scores (SD) for Ss who gave wrong responses</th>
<th>Mean VOT scores (SD) for Ss who gave correct responses</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 Camel</td>
<td>58</td>
<td>0.18</td>
<td>11.8 (6.1)</td>
<td>18.9 (5.6)</td>
<td>3.02*</td>
</tr>
<tr>
<td>02 Elephant</td>
<td>58</td>
<td>0.19</td>
<td>12.2 (4.6)</td>
<td>18.8 (5.7)</td>
<td>2.59*</td>
</tr>
<tr>
<td>03 Penguin</td>
<td>54</td>
<td>0.24</td>
<td>14.6 (7.2)</td>
<td>19.0 (5.5)</td>
<td>3.09*</td>
</tr>
<tr>
<td>04 Pig</td>
<td>51</td>
<td>0.29</td>
<td>17.2 (5.4)</td>
<td>19.1 (5.8)</td>
<td>2.12*</td>
</tr>
<tr>
<td>05 Cow</td>
<td>45</td>
<td>0.17</td>
<td>18.3 (5.8)</td>
<td>19.3 (5.7)</td>
<td>1.45</td>
</tr>
<tr>
<td>06 Rabbit</td>
<td>65</td>
<td>0.37</td>
<td>14.6 (6.2)</td>
<td>19.8 (6.3)</td>
<td>4.82*</td>
</tr>
<tr>
<td>07 Snail</td>
<td>69</td>
<td>0.44</td>
<td>16.5 (8.2)</td>
<td>20.1 (5.2)</td>
<td>4.30*</td>
</tr>
<tr>
<td>08 Crocodile</td>
<td>61</td>
<td>0.42</td>
<td>16.9 (7.8)</td>
<td>20.3 (5.1)</td>
<td>4.28*</td>
</tr>
<tr>
<td>09 Frog</td>
<td>43</td>
<td>0.39</td>
<td>17.1 (5.8)</td>
<td>20.9 (4.9)</td>
<td>5.55*</td>
</tr>
<tr>
<td>10 Bear</td>
<td>79</td>
<td>0.38</td>
<td>15.2 (5.3)</td>
<td>19.6 (5.5)</td>
<td>5.36*</td>
</tr>
<tr>
<td>11 Kangaroo</td>
<td>48</td>
<td>0.35</td>
<td>17.3 (6.0)</td>
<td>20.3 (5.1)</td>
<td>4.38*</td>
</tr>
<tr>
<td>12 Rhinoceros</td>
<td>54</td>
<td>0.37</td>
<td>16.9 (5.8)</td>
<td>20.3 (5.2)</td>
<td>5.07*</td>
</tr>
<tr>
<td>13 Sheep</td>
<td>62</td>
<td>0.27</td>
<td>17.5 (5.4)</td>
<td>19.5 (5.8)</td>
<td>2.65*</td>
</tr>
<tr>
<td>14 Seal</td>
<td>50</td>
<td>0.48</td>
<td>16.8 (5.4)</td>
<td>20.6 (5.5)</td>
<td>5.68*</td>
</tr>
<tr>
<td>15 Duck</td>
<td>10</td>
<td>0.22</td>
<td>18.6 (5.6)</td>
<td>19.9 (6.7)</td>
<td>1.14</td>
</tr>
</tbody>
</table>

16 Cup | 51               | 0.34                   | 15.3 (6.2)                                          | 19.0 (5.6)                                           | 3.07*    |
17 Corkscrew | 46         | 0.35                   | 16.9 (5.6)                                          | 20.3 (5.1)                                           | 5.95*    |
18 Dustpan | 67           | 0.29                   | 18.4 (5.6)                                          | 22.3 (6.6)                                           | 2.83*    |
19 Bike | 42               | 0.09                   | 18.1 (5.6)                                          | 19.6 (5.8)                                           | 2.12*    |
20 Shoe | 60               | 0.15                   | 16.4 (5.4)                                          | 19.3 (5.7)                                           | 3.33*    |
21 Ladder | 57            | 0.18                   | 17.9 (6.2)                                          | 19.4 (5.3)                                           | 2.08*    |
22 Spanner | 48           | 0.19                   | 18.5 (5.6)                                          | 18.9 (5.9)                                           | 0.49     |
23 Tractor | 69            | 0.34                   | 15.9 (5.5)                                          | 20.0 (5.4)                                           | 5.81*    |
24 Key | 24               | 0.31                   | 18.3 (5.7)                                          | 20.0 (5.8)                                           | 2.08*    |
25 Deckchair | 29          | 0.31                   | 17.4 (5.8)                                          | 22.1 (3.9)                                           | 6.55*    |
26 Scissors | 45           | 0.34                   | 17.1 (5.6)                                          | 20.8 (5.2)                                           | 5.50*    |
27 Pick axe | 21           | 0.21                   | 18.2 (6.0)                                          | 20.5 (4.6)                                           | 2.68*    |
28 Watch | 25               | 0.20                   | 18.1 (5.7)                                          | 20.5 (5.5)                                           | 2.82*    |
29 Binoculars | 77         | 0.12                   | 17.5 (6.4)                                          | 19.1 (5.5)                                           | 1.91     |
30 Glasses | 61            | 0.42                   | 16.4 (5.8)                                          | 20.9 (4.8)                                           | 6.79*    |

Notes: 1 Test scores corrected for the credit given to the item in question; *$p < 0.05$. 2 Per cent correct = number correct / total number of responses. Item-total correlation = correlation of the item with the item-total scores. Mean VOT scores (SD) for Ss who gave wrong responses = mean VOT score (standard deviation) for subjects who gave wrong responses. Mean VOT scores (SD) for Ss who gave correct responses = mean VOT score (standard deviation) for subjects who gave correct responses. $p$ value = significance level for the test of significance.
Inter-item correlations and reliability estimates are presented in Table 1. A more detailed item analysis given in Table 2 comprises the percentage of correct responses, the item-total correlations as a reliability estimate, and the VOT scores which corresponded to wrong and correct item responses. The differences as a measure of item validity were tested using t test statistics.

Differences between the two Silhouettes subsets (individual scores for animals minus scores for objects) reached from -7 to +9. Spearman rank correlations between the complete Silhouettes test and the two subsets with scores obtained for the Hooper VOT, WAIS-R Block Design, and the SKT short neuropsychological test battery are presented in Table 3.

Moreover the use of cutoffs was tested. Cutoff scores were applied for the Silhouettes subtest according to the VOSP test manual (for ages < 50, patients with scores < 16 were classified as impaired, for the older patients a score of < 15 was regarded es impaired). For the VOT, the cutoff for minimizing false-positive classifications given by the manual was < 21. For ages older than 65 years, the cutoffs recommended for geriatric patients (Nabors, Vangel, Lichtenberg & Walsh, 1997) were used (scores < 16). The resulting crosstable is presented in Table 4. Despite some discrepancy in the respective classifications, the contingency of the two measures is significant ($X^2 = 35.4$, $DF = 1$, $p < 0.05$), with a corresponding correlation (phi) of 0.37.

Table 3:
Rank correlations (Spearman) between the Silhouettes and scores in three other neuropsychological tests.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Silhouettes Total Score</th>
<th>1st Half (Animals)</th>
<th>2nd Half (Objects)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hooper VOT</td>
<td>265</td>
<td>0.65</td>
<td>0.54</td>
<td>0.56</td>
</tr>
<tr>
<td>WAIS-R Block Design</td>
<td>265</td>
<td>0.57</td>
<td>0.48</td>
<td>0.49</td>
</tr>
<tr>
<td>SKT Test Battery</td>
<td>262</td>
<td>-0.45</td>
<td>-0.38</td>
<td>-0.38</td>
</tr>
</tbody>
</table>

Notes: Abbreviations: VOT = Visual Organization Test, WAIS-R Wechsler Adult Intelligence Scale-Revised, SKT Syndrom-Kurztest.

Table 4:
Comparison of the Results of Two Tests for Visuospatial Impairment.

<table>
<thead>
<tr>
<th>Classification according to VOT cutoffs</th>
<th>Impaired</th>
<th>Not impaired</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impaired Silhouettes cutoffs:</td>
<td>69</td>
<td>18</td>
<td>87</td>
</tr>
<tr>
<td>Not impaired</td>
<td>72</td>
<td>106</td>
<td>178</td>
</tr>
<tr>
<td>Total</td>
<td>141</td>
<td>124</td>
<td>265</td>
</tr>
</tbody>
</table>

Note: Abbreviations: VOT = Hooper Visual Organization Test.
Table 5: Comparison of patients with right (n = 62) vs. left (n = 69) brain lesions.

<table>
<thead>
<tr>
<th></th>
<th>Right Brain Lesions</th>
<th>Left Brain Lesions</th>
<th>U Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Silhouettes (Total Score)</td>
<td>16.2</td>
<td>3.9</td>
<td>16.8</td>
</tr>
<tr>
<td>Silhouettes (Animals)</td>
<td>9.4</td>
<td>2.3</td>
<td>9.7</td>
</tr>
<tr>
<td>Silhouettes (Objects)</td>
<td>6.7</td>
<td>2.6</td>
<td>7.1</td>
</tr>
<tr>
<td>Hooper VOT</td>
<td>16.4</td>
<td>6.3</td>
<td>19.5</td>
</tr>
<tr>
<td>WAIS-R Block Design</td>
<td>15.6</td>
<td>12.1</td>
<td>20.9</td>
</tr>
<tr>
<td>SKT Test Battery</td>
<td>7.3</td>
<td>5.7</td>
<td>8.0</td>
</tr>
</tbody>
</table>

Notes: Abbreviations: VOT = Visual Organization Test, WAIS-R Wechsler Adult Intelligence Scale—Revised, SKT Syndrom-Kurztest.
*p < 0.05

In Table 5, the results of the subgroups with damages of the right vs. left cerebral hemispheres are summarized. While patients with right-hemisphere damage scored significantly below those with left hemisphere damage both in the Hooper VOT and the Block Design test, the same was not found for the Silhouettes. The non-significant differences in the SKT may be taken as an indication that the general level of mental dysfunction – having in mind the limitations of such a concept – was comparable for both groups.

In addition, some major findings of a more detailed analysis based on an item-by-item inspection are summarized below:

(1) Some items are very easy even for a neuropsychologically impaired population. This is particularly true for # 1 (camel) and # 2 (elephant). For the purpose of presenting a warming-up phase, however, this might be fully acceptable.

(2) For some items, specifying concrete scoring rules is necessary in order to secure inter-rater concordance. How are answers like bull, ox, or calf to be treated in # 5 (cow), what about sea-lion or walrus in # 14 (seal), to give just a few examples. Some scorers might take motorbike as a correct answer in # 19 (bike) while others do not. The same may be true for hammer in # 27 (pick axe), a very common response. The rules that were applied for this analysis are outlined above, but they were only established for the present analysis because of their absence in the manual.

(3) Some items lack item validity if the Hooper VOT test scores are employed as an external criterion. No significant validity coefficients were obtained for # 5 (cow), # 15 (duck), # 22 (spanner), and # 29 (binoculars).

(4) For a number of items, some wrong responses are very common. Such responses have to be analyzed separately because they may be indicative of significant item errors. This is illustrated by the answer bone (n = 71) for item # 22 (spanner) for which the correct response occurred in 129 cases. In Table 6, the scores of the two patient groups are presented. Neither do patients who gave the wrong response bone have significantly lower VOSP total scores, nor does any of the validity markers indicate that the two groups differ in visuospatial skills or cognitive functions in general.
Table 6:
Examination of item quality, item #22 (spanner). Patients who gave the correct answer do not differ in any of the four variables from patients who answered bone (means and standard deviations; all differences n. s.).

<table>
<thead>
<tr>
<th></th>
<th>Patients who gave the correct answer (spanner) (n = 129)</th>
<th>Patients who gave the wrong answer (bone) (n = 71)</th>
<th>t (DF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total VOSP score¹</td>
<td>17.3 (4.7)</td>
<td>16.5 (4.0)</td>
<td>1.22 (198)</td>
</tr>
<tr>
<td>Hooper VOT²</td>
<td>18.9 (5.9)</td>
<td>19.9 (4.8)</td>
<td>1.16 (197)</td>
</tr>
<tr>
<td>Block Design</td>
<td>20.4 (12.1)</td>
<td>20.2 (11.3)</td>
<td>0.09 (198)</td>
</tr>
<tr>
<td>SKT error score³</td>
<td>6.6 (5.4)</td>
<td>7.7 (5.6)</td>
<td>1.33 (195)</td>
</tr>
</tbody>
</table>

Notes: ¹Total VOSP scores corrected for the credit given to the item in question; ²Hooper Visual Organization Test, one missing case; ³Syndrom-Kurztest, three missing cases

Similarly, answers like iron for item #18 (dustpan), with 82 respondents, in contrast to only 19 who named the item correctly, or sawhorse (n = 47) for #25 (deckchair), or pump (n = 45) for #17 (corkscrew) are other examples of frequently occurring errors.

Discussion

Although Lezak (1995) described the VOSP as incorporating “experimental techniques for exploring visual perception” (p. 408), the normative data as well as the cutoffs provided in the manual suggest the use of these tests at a level of individual assessment. However, for the subtest Silhouettes it was demonstrated that its psychometric properties call for caution in this respect. Most importantly, reliability estimates are below those expected for test measures. In particular, the correlation of only 0.45 between the two subsets calls into question whether the combination of the two sets into one score is justified. As an alternative explanation, the recognition of living and nonliving objects may represent neuropsychologically different functions (as is suggested by Bunn, Tyler, & Moss, 1998, and Forde, 1999).

Contrary to what was found for the Hooper VOT and the Block Design test, no significant disadvantage of patients with right-hemisphere damages could be found for the Silhouettes. This can be seen as an indication for the importance of left hemisphere functions for object identification.

The analyses presented here reveal a number of other shortcomings. Item-total correlations (see Table 2) vary between 0.09 and 0.48, with a number or correlations in the range of 0.20. These results have to be judged unsatisfactory. A similar analysis of VOT items with a comparable sample of German-speaking neurological patients (Merten & Beal, 1999) yielded item-total correlations in the range between 0.18 and 0.62, with the median at 0.44.

The scores of item difficulty suggest that the empirically established item order for this sample shows considerable deviations from that given in the manual. This is important for the application of the discontinue rule proposed by the authors. For some items, the deviations might be due to cultural differences (possibly for the dustpan, #18), for others it is difficult to find an explanation in language or cultural differences (e.g., #13, sheep, or #29,
binoculars). Major problems were found in particular with # 15 (duck), which was not only one of the most difficult items, but regularly produced commentaries regarding test acceptance and statements of this figure being “nonsensical”. In the same vein, Müller (1997), in her test review, noted similar problems with the subtests Silhouettes, Object Decision, and Progressive Silhouettes.

The test performance in neurological patients was significantly correlated with age and education. Similarly, Herrera-Guzmán, Peña-Casanova, Lara, Gudayol-Ferré, and Böhm (2004) found with a sample of healthy normal elderly persons that age and sex were modifying variables in the Silhouettes subtest as well as education was. This, however, was not true for a number of other VOSP subtests.

One important issue to be discussed is whether an item analysis may be based on a patient sample at all. It is argued that since this is the population for which the test construction was developed, such a data basis for item analysis seems particularly relevant. In contrast, it appears questionable to develop a test for clinical purposes whose test and item characteristics are acceptable for a healthy population but not for the clinical population for which it was intended. An ideal solution would be to base test analyses and test revisions on the results of both normal and target populations. Often, however, no sound item analyses have ever been performed with samples of either population.

In her review, Müller (1997) also regretted a complete lack of any reliability data in the VOSP test manual. However, she concludes that a sufficiently high reliability is probably to be expected because of the nature of the tasks. As mentioned before, this could not be found for the Silhouettes with the sample studied here.

The present analyses were limited to the Silhouettes subtest. Although it would have been preferable to have comprehensive data on the complete VOSP battery, it is scarcely feasible to do the complete set of tests on a routine basis for neuropsychological assessment because this would be quite a time-consuming procedure. A more stepwise approach similar to that proposed by Lasogga and Michel (1994) was considered to be more appropriate: to use easy-to-perform screenings on all patients and to perform comprehensive assessment of visual functions only in cases when deficits become visible or a pre-established cutoff is reached. Moreover, it is a frequently observed practice in patient populations to give the subtest on its own or in combination with one other subtest, both in research and for single-case evaluations (e.g., Bodenburg, 2000; Uttner et al., 2001). The test authors themselves wrote: “Any number of the eight individual tests may be administered and there is no prescribed order.” (Warrington & James, 1991, p. 7)

Altogether, the specific task employed by the test appears to appeal to a number of researchers and clinicians and meet their practical needs. This, then, should stimulate efforts to perform some form of test revision to improve the instrument and secure that it will meet minimal psychometric requirements.

Acknowledgements

Thanks are due to Harriet Heier, Minden, Germany, for her linguistic supervision of the manuscript.
References


