T-SCORE AND RAW-SCORE COMPARISONS IN DETECTING BRAIN DYSFUNCTION USING THE BOOKLET CATEGORY TEST AND THE SHORT CATEGORY TEST

SAMUEL T. GONTKOVSKY AND GARY T. SOUHEAVER

Department of Veterans Affairs Medical Center
Little Rock, Arkansas

Summary.—Often considered one of the more discriminating tests between normal and brain-damaged individuals, the Category Test has been described as a complex measure of new problem solving, logical analysis, concept formation, abstract reasoning, and mental efficiency. Since publication of a booklet format by DeFilippis and McCampbell in 1979, researchers have attempted to develop other versions to reduce time required for administration. The present study compared the sensitivity of the Short Category Test, Booklet Format with the Booklet Category Test. Subjects were 22 male veterans seen for comprehensive neuropsychological evaluation and were classified into a brain-damaged or a control group (n=11) according to independent neurological or neuroradiological evidence. Comparison of performances based on T scores yielded no significant differences between the groups on either format. Analyses of performance based upon suggested raw score cut-offs for the tests, however, yielded a statistically meaningful difference, but caution is in order when using T scores derived from age- or education-based normative data to suggest the presence of brain dysfunction. Raw score cut-offs yielded valid differentiation of individual patients in both groups.

Since the inception of Halstead's original battery to assess "biological intelligence," the Category Test has been included in the Halstead-Reitan Neuropsychological Test Battery as an important measure for evaluating cerebral functioning (10). Often considered one of the more discriminating tests between normal and brain-damaged individuals (6, 11), the Category Test has been described a complex measure of new problem solving, logical analysis, concept formation, abstract reasoning, and mental efficiency (10, 13) and has been sensitive to both prefrontal lobe disturbance as well as diffuse dysfunction (2, 7). Initially, the test had 360 items organized into nine subtests but was later modified by Reitan to 208 items within seven subtests (14). Given the balky and heavy equipment required for administration, however, the measure was essentially limited to evaluations conducted within larger offices or laboratories. In addition, the cost of such extensive apparatus resulted in its exclusion from some neuropsychological evaluations. To overcome these obstacles, an easily portable booklet version of the test was developed

Address enquiries to Samuel T. Gontkovsky, Psy.D., Clinical Development Department, INTEGRIS Jim Thorpe Rehabilitation Center, 4219 South Western, Oklahoma City, OK 73109.
by DeFilippis and McCampbell (3). Scores on this booklet format were correlated with the original slide version, and no effect of test version was documented for clinical patients or normal persons administered both forms (12).

While the Booklet Category Test alleviated the issue of costly and unwieldy instrumentation, the time required for administration of the measure remained about two hours for more impaired patients (14). To address this concern, numerous researchers developed shortened versions. One, developed by Wetzel and Boll, is the Short Category Test, Booklet Format, which has five subtests and retains the first 20 items in sequence of subtests 2, 3, 4, 5, and 6 of the full test. It is administered in a manner consistent with the booklet version developed by DeFilippis and McCampbell (for a more complete description of development and standardization, refer to 14). The present study compared the sensitivity of this 100-item short form with the 208-item Booklet Category Test.

**METHOD**

Subjects (\(N=22\)) were men from a large Department of Veterans Affairs Medical Center seen for comprehensive outpatient neuropsychological evaluation. Eleven patients (\(M \text{ age} = 42.1 \text{ yr.}, SD = 14.5; M \text{ education} = 11.9 \text{ yr.}, SD = 2.3\)) were classified as brain-damaged based on independent neurological or neuroradiological examination and served as one group. The remaining 11 patients (\(M \text{ age} = 37.3 \text{ yr.}, SD = 10.7; M \text{ education} = 12.5 \text{ yr.}, SD = 2.6\)) presented with complaints of cognitive difficulties, e.g., memory problems, but no evidence of abnormality was demonstrated on neurological or neuroradiological evaluation. Therefore, these patients served as a pseudo-neurological control group. No significant difference was noted between the groups on mean Full Scale Intelligence Quotient based on the Wechsler Adult Intelligence Scale-Revised (\(F_{1,21} = 1.23, p > .05\); see Table 1). Specific diagnoses for the externally validated group were 1 generalized cortical atrophy, 1 epilepsy, 3 other seizure disorder, 1 cerebral vascular disease, 1 malignant parietal lobe neoplasm, 1 brain hemorrhage, and 3 brain injury not elsewhere classified.

All subjects were administered the booklet form as part of a flexible neuropsychological evaluation. Protocols were rescored subsequently by the authors using the items included in the short form. Raw scores from the full form were converted to T scores using the norms developed by Heaton, Grant, and Matthews (8). Short-form raw scores were converted to T scores using the manual (14). Subjects were treated in accord with the ethical principles of the American Psychological Association (1).

**RESULTS AND DISCUSSION**

Comparison of performances based on T scores yielded no between group differences on either the booklet (\(F_{1,21} = 1.82, p > .05\)) or short form
(\(F_{1,21} = 1.56, \ p > .05\)). Analyses using raw-score cut-off for the tests, however, yielded a statistically significant difference in mean performances between groups for both the booklet form (\(F_{1,21} = 9.25, \ p < .001\)) and short form (\(F_{1,21} = 5.21, \ p < .01\)). Means and standard deviations of performances are presented by group in Table 1. Using the established raw score cut-off of 51 errors, 10 of 11 subjects from the externally validated group and 9 of 11 subjects from the pseudoneurological control group were correctly classified with the booklet version. On the other hand, using cut-offs of 41 or more errors for individuals \(\leq 45\) years and 46 or more errors for individuals over age 45 as suggested by Wetzel and Boll (14), only 3 of 11 subjects from the externally validated group were correctly classified with the short version. Yet, all 11 subjects from the control group scored below the cut-offs. Adjusting the short-form raw score cut-off to 41 errors for all ages increased the hit rate, so 7 of 11 subjects of the externally validated group and all 11 subjects of the control group were correctly classified. Consistent with the conclusions of Horn and Kelly (9), these results suggest that the cut-off scores established by Wetzel and Boll (14) may require downward adjustment to increase sensitivity. Overall, results are concordant with previous findings, e.g., 4, 5, indicating caution in the classification of patients with brain dysfunction with using regression-based norms of a single test to correct for various demographic influences.

<table>
<thead>
<tr>
<th>Scale/Score</th>
<th>Externally Validated Group</th>
<th>Pseudoneurological Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Scale Intelligence Booklet Form</td>
<td>95.2 11.1</td>
<td>104.3 10.0</td>
</tr>
<tr>
<td>T Score</td>
<td>37.4 7.8</td>
<td>44.8 5.8</td>
</tr>
<tr>
<td>Raw Score</td>
<td>73.0 13.5</td>
<td>42.3 9.7</td>
</tr>
<tr>
<td>Short form</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T Score*</td>
<td>61.4 7.3</td>
<td>51.4 5.9</td>
</tr>
<tr>
<td>Raw Score</td>
<td>40.7 11.6</td>
<td>23.7 7.6</td>
</tr>
</tbody>
</table>

*On the short form, higher T scores correspond with a greater likelihood of impairment.

While these findings are consistent with use of the short form as a valid measure of neurocognitive status in patients with cerebral dysfunction, results must be interpreted in light of several practical and methodological limitations. First and foremost, the sample was quite small and comprised a heterogeneous clinical population with respect to diagnosis. In addition, because group classification was based upon the presence of neurological or neuroradiological evidence of impairment, it is possible that groups may
have been equivocally deficient with respect to cognition. Therefore, some caution is warranted with respect to generalization of findings. The fact that short-form scores were derived from performances on the booklet version rather than on separate administration increased the likelihood that scores between the measures would be concordant in classification. Such a design, however, prevented practice effects and was necessitated by practical reasons, given that this sample was comprised of patients referred for clinical evaluation not of subjects recruited specifically for research. Finally, although raw-score performances were converted to T scores on both measures, comparison must be considered in light of the fact that T scores derived from different populations may not be entirely comparable.

REFERENCES