TEST-RETEST STABILITY
OF THE BEHAVIOURAL ASSESSMENT
OF THE DYSEXECUTIVE SYNDROME
IN A SAMPLE OF PSYCHIATRIC
PATIENTS

MARKO JELICIC*, CÉCILE E. C. HENQUET*,
MAYKE M. A. DERIX and JELLE JOLLES*

*Academic Neuropsychology Department, Psychomedical Centre Vijverdal,
Maastricht Brain and Behaviour Institute and European Graduate
School of Neuroscience (EURON), University of Maastricht,
The Netherlands; Department of Clinical Psychology,
Twenteborg Hospital, Almelo, The Netherlands

(Received 12 March 2001)

The Behavioural Assessment of the Dysexecutive Syndrome (BADS) is a relatively new
test battery designed to measure disorders of executive functions. We studied the
temporal stability of the BADS in a sample of 22 adult psychiatric patients. All
patients were administered the BADS twice with an interval of 3 weeks. Test-retest
correlations for the BADS tests ranged from .22 to .85. On the repeat administration,
patients obtained higher scores on one test as well as on the total BADS. Our results
suggest that the BADS should not be administered on two occasions a few weeks
apart.

Keywords: Executive functions; Neuropsychological testing; Psychiatric patients

In clinical neuropsychology, it is common practice to assess patients' cognitive functions on different occasions (Lezak, 1995). Traumatic
brain injury patients, for instance, are tested repeatedly to determine

*Address for correspondence: Academic Neuropsychology Department, Psychomedical Centre Vijverdal, P. O. Box 88, 6200 AB Maastricht, The Netherlands. e-mail: m.jelicic@np.unimaas.nl
recovery of brain functions. Elderly with suspected dementia are given repeat assessments in order to establish further cognitive decline. Psychiatric patients are administered neuropsychological tests on more than one occasion, to examine the effects of psychotropic medications. To be used repeatedly, neuropsychological tests should have high test-retest stability. It is difficult to interpret changes in performance on tests with questionable temporal stability: A higher score on the repeat administration of a test may reflect genuine cognitive improvement, a mere practice effect, or a combination of both improvement and practice.

Neuropsychological tests designed to measure executive functions (planning, monitoring, and organization of behavior) often have low test-retest reliability. In a study by Paolo et al. (1996) healthy older adults were administered the Wisconsin Card Sorting Test (WCST) twice. The authors found that all of the WCST scales had low stability coefficients (but see Ingram et al., 1999, for slightly different results). De Zubicaray et al. (1998) studied the temporal stability of the Modified Card Sorting Test (MCST) in a healthy older sample, and reported a low MCST test-retest reliability. Why do card sorting tests have low temporal stability? Burgess (1997) argued that in most executive tests subjects are confronted with novel problems which they will have to solve. When administered a second time, the novelty is gone and the test no longer measures problem-solving abilities. The Behavioural Assessment of the Dysexecutive Syndrome (BADS) is a relative new instrument developed to assess disorders of executive functions (Wilson et al., 1997; 1998). In contrast with traditional executive tests, the BADS is considered to be ‘ecological valid,’ i.e., the instrument aims to measure deficits in everyday life. Wilson et al. (1998) studied the temporal stability of the BADS. A group consisting of 29 normal subjects was administered the BADS twice, with an interval of 6–12 months. Although the test-retest correlations were moderate, no significant improvement in test performance was found on the repeat administration. The present study was initiated to investigate the temporal stability of the BADS in a clinical population. Since executive dysfunction is a feature of many psychiatric disorders (Grant and Adams, 1996), we decided to examine the instrument’s test-retest reliability in a group of psychiatric patients.
METHODS

Subjects

Our sample consisted of 22 adult psychiatric patients with different diagnoses (e.g., schizophrenia, affective disorders, obsessive compulsive disorder etc.). Mean age of the subjects was 40 yrs (SD 8.8), mean IQ was 101 (SD 16.7). All patients were in remission.

Procedure

All patients were administered the Dutch version of the BADS (Krabbendam and Kalff, 1998) twice with an interval of 3 weeks. The BADS consists of six tests (Wilson et al., 1996). For each of the tests a summary profile score is obtained (with a minimum of 0 and a maximum of 4) and these are added up to obtain a total profile score (with a maximum of 24). The Rule Shift Card Test uses playing cards and evaluates the ability of subjects to respond correctly to a rule and to rule shifting. In the Action Program Test, subjects are asked to solve a practical problem, i.e., to remove a cork from a small tube using a few props. The Key Search Test is used to assess search strategies. Subjects are presented with a piece of paper containing an imaginary square and requested to draw a path to find a lost key. In the Temporal Judgement Test subjects are asked to make time judgments, for example “How long do most dogs live for?” The Zoo Map Test examines route planning. Subjects are given a map of a zoo and instructed to visit locations of the zoo while obeying certain rules. In the Modified Six Elements Test the ability to plan, organize and monitor behavior, is measured. Subjects are given six tasks which they have to do in 10 minutes. It is impossible to complete all tasks within the time limit.

Data Analysis

We first calculated Spearman test-retest correlations of the six BADS tests and the total BADS score. To find out if the BADS scores at the initial test session differed from the repeat administration, Wilcoxon matched-pairs tests were performed. We opted for nonparametric
statistical methods because the BADS scores were not normally distributed.

RESULTS

The scores on the first and second administration of the BADS are presented in Table I. The Action Program Test had a test-retest correlation coefficient of .22. For the Rule Shift Card Test, the Temporal Judgement Test, The Zoo Map Test, and the Modified Six Elements Test, correlations ranged from .48 to .60. Both the Key Search Test and the total BADS had a test-retest correlation over .80. Compared to the initial test session, patients had higher scores on the repeat administration of both the Action Program Test and the total BADS ($p < .01$).

DISCUSSION

In line with research using card sorting tasks (cf. de Zubicaray et al., 1998), most of the test-retest correlations of the BADS tests failed to reach the minimum of .80 required for a clinical psychometric instrument (Anastasi and Urbina, 1997). The test-retest correlation of the total BADS was higher than .80, suggesting that the overall score may be used in repeat measurements. However, the second test session yielded significantly higher total BADS scores than the initial test administration, indicative of a practice effect. In most of the BADS tests subjects are asked to solve a new problem. It seems likely

<table>
<thead>
<tr>
<th>Test</th>
<th>1st administration mean (SD)</th>
<th>2nd administration mean SD</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule Shift Card</td>
<td>2.86(1.42)</td>
<td>3.00(1.11)</td>
<td>.60</td>
</tr>
<tr>
<td>Action Program*</td>
<td>3.27(1.03)</td>
<td>4.00(0)</td>
<td>.22</td>
</tr>
<tr>
<td>Key Search</td>
<td>2.27(1.55)</td>
<td>2.45(1.30)</td>
<td>.84</td>
</tr>
<tr>
<td>Temporal Judgement</td>
<td>1.77(1.31)</td>
<td>1.91(1.81)</td>
<td>.50</td>
</tr>
<tr>
<td>Zoo Map</td>
<td>1.95(1.43)</td>
<td>2.27(1.39)</td>
<td>.48</td>
</tr>
<tr>
<td>Modified Six Elements</td>
<td>3.00(1.31)</td>
<td>3.27(1.94)</td>
<td>.48</td>
</tr>
<tr>
<td>Overall BADS*</td>
<td>15.00(5.35)</td>
<td>16.82(3.90)</td>
<td>.85</td>
</tr>
</tbody>
</table>

* Difference between 1st and 2nd administration: $p < .01$. 
that patients obtained higher scores on the repeat administration of
the BADS because the tests had lost their novelty (cf. Burgess, 1997).
This especially would have been the case for the Action Program Test.

Wilson et al. (1998) studied the temporal stability of the BADS in
normal subjects. They found moderate test-retest correlations, but no
significant differences in performance between the first and second test
administration. In contrast, our subjects demonstrated improved
performance on the Action Program Test and the overall BADS.
How to explain these discrepant findings? Our sample consisted
of psychiatric patients, who as a group exhibited poor executive
functioning. Krabbendam et al. (1999) administered the BADS to
schizophrenic patients and normal controls. The mean total BADS
score was 19.1 for the control subjects and 16.1 for the patients. At the
first test session, the mean overall BADS score of our study group was
1.1 point below that of the schizophrenic group of Krabbendam and
her colleagues. Thus, our subjects had substantial “room for
improvement” on the second administration of the BADS. Possibly,
Wilson et al. (1998) did not find practice effects in their study, because
normal subjects have little to gain on a repeat administration of the
BADS.

Our findings suggest that, in clinical psychiatric practice, the BADS
should not be administered on two occasions a few weeks apart.
Further research is necessary to determine the temporal stability of the
BADS in psychiatric patients using longer time intervals, and in other
patient populations.

References

Prentice Hall.

Rabbitt, P. (Ed.), Methodology of frontal and executive function (pp. 81–116).
Hove: Psychology Press.

Card Sorting Test: Test-retest stability and relationships with demographic
variables in a healthy older adult sample. British Journal of Clinical Psychology,
37, 457–466.


stability of the Wisconsin Card Sorting Test in an untreated patient sample. British
Journal of Clinical Psychology, 38, 209–211.


