RELIABILITY AND VALIDITY OF THE DIGIT CANCELLATION TEST, A BRIEF SCREEN OF ATTENTION

Takeshi HATTA1), Kazuhito YOSHIZAKI2), Yasuhiro ITO3), Mitsuhito MASE4), and Hidehiro KABASAWA5)

1)Kansai University of Welfare Sciences, Japan
2)Aichi Shukutoku University, Japan
3)Nagoya University, Japan
4)Nagoya City University, Japan
5)Nagoya City Rehabilitation Center, Japan

The reliability and validity of a screening test called the D-CAT (Digit Cancellation Test) were evaluated across two studies. The D-CAT was developed to provide a highly practical and user-friendly assessment of various aspects of prefrontal cortex functioning, including information processing speed, the ability to focus attention, and executive functioning. Participants perform the D-CAT by deleting given target numbers on a sheet of randomly arranged possibilities. In Study 1, the reliability of the D-CAT was evaluated using a test-retest paradigm. Reasonably high correlations between scores on the two test sessions were obtained. In Study 2, construct validity was examined using a sample of participants with traumatic brain injury. TBI participants showed significantly lower D-CAT performance than age and education level matched healthy controls. On the basis of these findings, the D-CAT can be regarded as a reliable and valid screening test for attentional functioning.

Key words: screening test for attention, digit cancellation test, traumatic brain injury, test-reliability, test-validity

“Higher-order cognitive functioning” consists of several different facets, including attention, memory, language, and problem solving. Each of the facets interrelates with the others in complex ways, although attention is regarded as a basic aspect of executive function in so-called higher-order cognitive functioning. The assessment of attention can be regarded as one of the most important basic aspects of any neuropsychological evaluation. Various assessments of human attentional abilities have been developed, including the Paced Serial Addition Test (PASAT), Brief Test of Attention (BTA), Continuous Performance Test (CPT), Symbol Digit Modalities Test (SDMT), and Visual Search and Attention Test (VSAT). Each of these tests possesses advantages and shortcomings (Spreen & Strauss, 1998). For example, the PASAT (Fisk & Archibald, 1998) is too difficult for elderly and brain damaged individuals, requiring participants to add the currently presented digit to the one immediately preceding it when aurally presenting a sequence of digits. The test also generates high levels of negative affect in

Authors are much indebted to Drs. Abe, J., Nagano, T. and Kanari, A. for their kind help in conducting study.

Correspondence concerning this article should be addressed to Kazuhito Yoshizaki, Department of Psychology, Aichi Shukutoku University, 9, Katahira, Nagakute, Aichi, 480-1197, Japan (e-mail: yoshizak@asu.aasa.ac.jp).
examinees (e.g., “I failed the final trial and then the examiner ended the testing!”). Other tests such as the BTA and CPT involve similar shortcomings, and also require the use of various equipments (e.g., tape recorders or computers) that render them unsuitable for group administration. The SDMT, Concentration Endurance Test (d2Test) and VSAT are paper and pencil tests, but are not necessarily appropriate measures to use in all settings, in terms of cost, administration times and difficulty levels (e.g., for the elderly; Ballard, 1997; Brickenkamp, 1981; Schretlen, Bobholz, & Brandt, 1996; Smith, 1991; Trenerry, Crosson, DeBoe, & Leber, 1990).

Hatta, Ito, and Yoshizaki (2001) developed a digit cancellation test (the D-CAT) that is suitable for group administration, to meet staff requirements at a public health service center. Staff had requested development of a screening test high in practical utility, including such features as no expensive equipment required for administration, quick administration without special training, clear quantitative results yielded, possibility of repeated administration for longitudinal examinations, and no or minimal triggering of negative affect in examinees. Clinical neuropsychologists have also stressed the importance of developing short screening measures, given the major demands that test batteries such as the Halstead-Reitan battery place on examinees (length, fatigue, test difficulty level; Spreen & Strauss, 1998).

The D-CAT is a pencil and paper test that does not require special or expensive equipment to administer, requiring only five minutes for administration, including instructions and practice opportunities. Examinee performance is quantitatively analyzed. The D-CAT is suitable for repeated administration (e.g., every month) and does not typically elicit frustration, disappointment, or other forms of negative affect (participants are unaware of missed digits). Finally, the D-CAT can be administered individually or in groups.

The idea of a digit cancellation test to assess attention is not particularly novel or innovative. Digit cancellation tests have been used to assess hemi-neglect patients and those with Alzheimer’s disease (e.g., Egelko, Riley, Simon, & Diller, 1988; Geldmacher, 1998; Geldmacher, Doty, & Heilman, 1995; Geldmacher, Fritsch, & Riedel, 2000; Geldmacher & Hills, 1997; Weintraub & Mesulam, 1988). However, such tests have not been validated in a public health examination setting. Although there are many medical screening tests in a doctor’s office, there are few simple and practically valid screening test in medical public health settings. MMSE may be an exceptional simple screening test but it is too simple and some claimed limitation (Iwahara, Hatta, Ito, & Nagahara, 2006). The WAIS-IV/WISC-IV may be reliable and possess high scientific validity but it is too heavy for the screening test in medical public health settings.

Among the various models of attention, the D-CAT is basically consistent with the model proposed by Sohlberg and Mateer (1989), in which five hierarchical classifications are proposed; each component of the hierarchy requires the effective functioning of the one below. Among their five hierarchical classification model, focused attention, sustained attention or concentration, and selective attention are the most basic classes of the hierarchy. Therefore, we regarded that the lower class of attention hierarchy must be the basis of sound functioning of higher class attention hierarchy. Then, it must be
suitable to focus on the lower three classes to simplify as a screening test. Then, the D-CAT aims to evaluate three levels of attention: *Focused attention, sustained attention or concentration*, and *selective attention*. The D-CAT has been widely adopted in Japan, since revised norms based on a sample of more than 2000 (18- to 89-year olds) were published in 2006 (Hatta, Ito, & Yoshizaki, 2006; e.g., the D-CAT has been recommended for inclusion in evaluations of traumatic brain injury (TBI) patients). However, further evidence of the reliability and validity of this test is required.

The present investigation aimed to provide a thorough examination of the reliability and validity of the D-CAT. Study 1 focused on reliability and Study 2 examined validity.

**STUDY 1: RELIABILITY OF THE D-CAT**

Study 1 assessed test-retest reliability. The D-CAT was administered twice to the same group of participants, and correlations between the scores at times 1 and 2 were calculated. Charter (1999) recommends using a large sample for such an analysis, as was done here. Although use of a fairly diverse sample is ideal when evaluating reliability, for the present purposes it was not practical to administer the D-CAT to a large mix of different examinees (e.g., clinical, non-clinical, etc.). We instead administered the D-CAT twice to a sample of college students at different academic levels. In our view, obtaining consistent results across the two administrations is the crucial issue with regards to reliability.

**METHOD**

**Participants**

Three hundred and ten students (18–22 years old) from three universities and two junior colleges participated in this study. The students completed the test twice with the same examiner, and with a 2-week interval between administrations.

**Materials**

The D-CAT test sheet consists of 12 rows of 50 digits. Each row contains 5 sets of numbers from 0 to 9, arranged in a random order: Any one digit appears five times in each row, along with randomly determined neighbors. The D-CAT as a whole consists of three such sheets, as described by Hatta et al. (2001, 2006).

**Procedure**

Participants were instructed to search for specified target number(s) and to delete each one with a slash mark, as quickly and accurately as possible, until the experimenter provided a signal to stop. The experimenter instructed participants to start searching at the left end digit in the uppermost row, and to then move onto the next row when required. There are 3 trials in total, the first involving a single target number (6), a second with two target numbers (9 and 4), and a third with three (8, 3, 7). One minute was allowed for each trial, with the total time required to complete the D-CAT being roughly 3 minutes. During the second and third trials, it was stressed that all of the target numbers should be cancelled without omission.

After the instructions and practice trials, participants searched for and cancelled, as fast and accurately as possible, the target number 6 during the first trial, numbers 9 and 4 during the second, and numbers 8, 3, and 7 during the third.
RESULTS

Three D-CAT measures were calculated and analyzed according to the test manual: Total Performance, Omission ratio, and Reduction ratio. Total Performance refers to the total number of digits the participant inspected (rather than the digits deleted). This index pertains mainly to cognitive components such as information processing speed, focused attention, and sustained attention. Omission ratio primarily reflects sustained and selective attention. It is calculated using the formula (number of missed targets) ÷ (number of digits inspected) × 100. Reduction ratio is calculated using the formula (number of digits inspected in Trials 2 and 3) ÷ (number of digits inspected in Trial 1). This measure pertains to the relative durability of focused, sustained, and selective attention, as well as tolerance of mental fatigue. False alarms (commission errors) are also used as an index in the original D-CAT. However, false alarms and omission ratios were not analyzed here, as the occurrences of false alarms and omissions in our sample were only 0.1% and 3% respectively.

Pearson’s correlation coefficients were calculated, using the participants’ first and second test administrations. The Total Performance and Reduction ratio indices were analyzed.

Total Performance correlations between Test 1 and Test 2 for Trials 1, 2, and 3 were $r(308) = 0.81$, $r(308) = 0.76$, and $r(308) = 0.75$, respectively. These correlations were all quite large and statistically significant ($ps < .001$).

Reduction ratio correlations between Test 1 and Test 2 for Trials 1, 2, and 3 were $r(308) = 0.79$, $r(308) = 0.86$, and $r(308) = 0.85$, respectively. Again, these are all relatively high and statistically significant ($ps < .001$).

DISCUSSION

Test-retest reliability for the D-CAT appears to be reasonably high. Dikmen, Heaton, Grant, and Temkin (1999) report test-retest reliability data for the Halstead-Reitan Neuropsychological Test Battery as well as other measures of memory, attention, and motor functioning from 384 healthy adults, thereby providing a standard of comparison by which the D-CAT can be evaluated. Almost all measures showed reasonably robust Pearson’s coefficients, ranging from 0.7 to 0.9, though the reliability of most memory measures was relatively poor (i.e., correlations of less than 0.6). Relatively poor test-retest reliability of memory tests has also been reported elsewhere (e.g., Geffen, Butterworth, & Geffen, 1994; Rapport et al., 1997; Rasmussen, Bylsma, & Brandt, 1995; Uchiyama, D’Elia, Dellinger, & Becker, 1995). The correlations obtained for the D-CAT were robust, and we may therefore conclude that the D-CAT is reasonably reliable.
This second study examined the construct validity of the D-CAT. We assessed whether the D-CAT is sensitive enough to evaluate attentional impairments, administering the test to a group of participants who are reasonably suspected to have attentional disorders.

Lezak (2004) wrote that impairments in attention and concentration are the most common neuropsychological problems following traumatic brain injury (or TBI). Mateer (2000) stated that individuals with mild TBI frequently report concentration problems, distractibility, forgetfulness, and difficulties with doing more than one thing at a time. Further, it is well known that both information processing speed and attention are impaired after TBI (Prigatano, 1986; Whyte, Schuster, Polansky, Adams, & Coslett, 2000). Then, in the present study, the D-CAT was administered to two groups: TBI patients and normal, healthy controls. Clear performance differences between the two groups on the D-CAT would provide support for the test’s construct validity.

**Method**

**Participants**

Forty-two TBI participants (34 males and 8 females) and 42 intact individuals (34 males and 8 females) participated in this study. The TBI participants were recruited from the Nagoya City Rehabilitation Center (NCRC). All TBI participants were diagnosed as TBI at their admission phases, however, it is reasonable that they are mostly mild TBI (mTBI) by the recent diagnostic classification (Cassidy et al., 2004). Participant’s original diagnosis was done by MDs at NCRC hospital and their diagnoses were based upon at least one of the brain imaging data such as CT, MRI, fMRI SPECT and Photo-scan depending on admission time. The participants showed a clear brain damage in the data or a discrepancy between both MRI and fMRI (or SPECT) (i.e., no clear damage on MRI but weak activation on fMRI) and MDs have proposed diffuse type dendrites damage as an interpretation of causes in latter cases.

The control participants were from Y-Town. The control participants were the staffs of the public health department. The study was conducted with the ethical approval of the NCRC research community. Participant demographics are shown in Table 1. There were no statistical differences in age or education level between the groups (t = 1.75 and 0.62). Thirty-eight participants had been injured in a traffic accident, and 4 injured in falls. The TBI participants were all outpatients at the NCRC, involved in a cognitive rehabilitation program. These patients were suffering from deficits typically occurring after TBI, but they had no serious language problems (scoring above 90 points on the Japanese Western Aphasia Battery). No participant had a pre-morbid history of substance dependence, intellectual disability, or prior neurological or psychiatric difficulties. Therefore, all participants were suspected diffused axonal injury because of their etiology and symptoms. Specific neuropsychological test results for this sample were reported elsewhere by Ito and Hatta (2003).

**Materials and Procedure**

One examiner individually administered the D-CAT to all of the TBI participants, following the test manual instructions. After the instructions and practice trials were administered, the participants searched for and cancelled the target number 6 during the first trial, the numbers 9 and 4 during the second trial, and 8, 3, and 7 during the third trial (60 s), as fast and as accurately as possible. For the control participants, the D-CAT was administered following the test manual instructions.
We calculated the same D-CAT performance indices that were analyzed in Study 1. Table 2 shows the means (and standard deviations) for Total Performance, Omission ratio, and Reduction ratio, for both intact individuals and TBI participants at each trial condition.

Mean scores on the three indices were each analyzed using a repeated two-way ANOVA (2: group × 3: trial). There was a significant main effect of group on Total Performance scores, $F(1, 82) = 72.72$, $p < .001$. The control group performed significantly better than the TBI group on this measure. The main effect of trial was also significant, $F(2, 164) = 196.76$, $p < .001$, suggesting that Total Performance decreased with increasing target size. Furthermore, post-hoc analyses showed that the group by trial condition interaction was significant for Total Performance, indicating that the performance differences between groups were not consistent across trials.

Table 2. Mean Scores and SDs for Total Performance, Omission Ratio, and Mean Percentages of Reduction Rate in Intact Control and TBI Participants (Study 2).

<table>
<thead>
<tr>
<th>Target Condition</th>
<th>TBI</th>
<th>Normal Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ONE</td>
<td>TWO</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Total Performance</td>
<td>220.1</td>
<td>79.1</td>
</tr>
<tr>
<td>Omission Ratio</td>
<td>1.0%</td>
<td>4.0</td>
</tr>
<tr>
<td>Reduction Ratio</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
interaction was significant, $F(2, 164) = 10.40, p < .001$. This means that the performance difference across the three target conditions was larger in the intact individuals, as compared to the TBI participants.

In the analysis of **Omission ratio**, there was a significant main effect of group, $F(1, 82) = 6.63, p < .01$, such that the **Omission ratio** was larger in the TBI group than in the control group. There was also a significant effect of trial, $F(2, 164) = 24.20, p < .001$. Post-hoc analyses showed that the **Omission ratio** for Trial 1 (one target number) did not differ across the two groups, while this ratio was significantly higher in the TBI participants than in the control group on Trials 2 and 3, where the participants had to use a greater attention load. The group by trial interaction was not significant, $F(2, 164) = 2.40$.

There was a significant main effect of trial on **Reduction ratio**, $F(1, 82) = 151.68, p < .001$, but neither the main effect of group nor the interaction were significant ($F$s$(1, 82) < 2.40$). These results show that, with increasing target size, ratio scores significantly decreased, and that this tendency was common to both groups of participants.

**DISCUSSION**

We examined the construct validity of the D-CAT by comparing the performance of TBI participants with that of healthy controls. Attention difficulties constitute one of the most prominent cognitive disorders in TBI individuals. The **Total Performance** and **Omission ratio** comparisons revealed worse performance in TBI participants, relative to healthy age, education, and gender matched control participants. The **Total Performance** index refers mainly to information processing speed. Mean **Total Performance** of the TBI participants was approximately 40% less than that of the control group participants.

This robust performance decline suggests that searching for and canceling out a memorized target is one of the most reliable and valid techniques, from the standpoint of cost effectiveness and practical convenience, for evaluating attention. Furthermore, for the TBI participants we found more prominent performance decrements in the presence of high cognitive demand, i.e., two and three target trials (Baddeley, 2000; McDowell, Whyte, & D’Eposito, 1997).

The **Omission ratio** of the TBI participants was approximately three times as great as that of the intact individuals, during the two and three target trials (i.e., during a more demanding task), but there was no significant difference between the **Omission ratios** of the TBI participants and intact individuals during the one target trial. These declines of two measures suggest that D-CAT is useful in detecting danger when we consider the presence of greater attentional load to drive home due to the presence of deficit in TBI participants. No differences in **Reduction ratio** suggest that this index might not be sufficiently sensitive for the evaluation of sustaining attention, given that even TBI participants appear capable of sustaining their attention abilities for 3 minutes. The D-CAT might not prove to be a suitable screening test for mental fatigue or the ability to sustain attention to the cognitive demanding effort. Further examinations should focus on the D-CAT as a means to detect sustained attention or vigilance deficits.
In sum, the present results show that the D-CAT possesses some degree of construct validity, and that it could be a useful tool for detecting cognitive disorders, particularly those involving deficits in information processing speed and focusing one’s attention for a brief period of time.

**GENERAL DISCUSSION**

The present study examined the reliability and validity of a relatively new screening test for attention, the D-CAT. This test was developed at the request of community welfare management officers, who aim to reduce the costs of health management while at the same time ameliorating age-related cognitive deficits and/or problems in motor-kinesthetic function. The D-CAT could potentially be included in a low cost management system for elderly people in the community. Hatta (2004) has proposed a cognitive test battery (the NU-CAB) for use in community health examinations, which includes the D-CAT.

In order to test the validity for the D-CAT, we compared the performance of TBI patients with that of the intact individuals. A pile of studies has proposed a strong relation between mTBI and frontal area (especially prefrontal lobe) dysfunction. Recent reviews (Drag, Spencer, Walker, Pangilinan, & Bieliauskas, 2012; Ettenhofer & Barry, 2012; Vasterling & Dikmen, 2012) could introduce present research status around mTBI studies. Frontal area dysfunction was suspected due to the facts that almost TBI patients showed decreased function in attention, working memory, verbal fluency, executive functions based upon neuropsychological behavioral tests such as Boston naming test, VFT, WCST, WMS-R, Stroop-test, WAIS-III, as well as brain imaging test (e.g., Owen, Evans, & Petrides, 1996). Our recent fMRI, SPECT as well as photo-scan data also confirmed a frontal and temporal lobe low activation in TBI patients (Hatta et al., 2009; Kabasawa, 2008).

Before describing the present reliability and validity results in more detail, we should briefly explain why digit/letter cancellation test performance is strongly associated with frontal lobe dysfunction and/or early signs of aging-related cognitive decline. Brain imaging studies have revealed that attention, especially digit/letter finding behavior, activates frontal brain areas (Cabeza et al., 2003; Leonard, Sunaert, Van Hecke, & Orban, 2000; Mellers et al., 1995). For example, Cabeza et al. (2003) used event-related fMRI and reported that frontal areas, the parietal cortex, cingulate gyri, and thalamus were activated during a visual target detection task. Mellers et al. (1995) reported activation in the anterior and posterior parasagittal cortices during a visual letter monitoring task. More precisely, a target (digit/letter) monitoring/searching task activates not only the frontal area, but also the fronto-parietal-cingulate-thalamic network.

The D-CAT possesses reasonably high levels of test-retest reliability (Study 1). Two test administrations set two weeks apart correlated quite highly with one another, providing evidence of temporal stability in D-CATS performance. We also examined construct validity (Study 2). As described previously, attention deficit is one of the most
prominent symptoms in TBI participants (Prigatano, 1986; Whyte et al., 2000), and the D-CAT should be able to detect at least some forms of attention deficit in this population. Indeed, **Total Performance** (D-CAT index) of the TBI participants was roughly 40% less than that of age-matched healthy control participants, while the **Omission ratio** of the TBI participants was three times that of the controls. These results show that the D-CAT serves as a useful attention screen, although it may not be useful in the assessment of vigilance or the ability to sustain effort over time due to the result of the test’s relatively low difficulty level and short administration time.

The D-CAT can be regarded as a clinically efficient test, from the standpoint of the financial cost, time and ease of administration. The D-CAT requires roughly 3 minutes to administer. The cost of the test per person is relatively low and it requires no special facilities or equipment to administer (such as sound proof room, computer, etc.). The D-CAT can be administered by a person without specialized neuropsychological training. The test’s advantages render it suitable for multiple administrations to patients in community settings. Although the present study supports the use of the D-CAT in community settings, further investigations of the test’s operating characteristics and validity are necessary. For example, norms are needed for younger age groups, different educational levels, different occupational groups, and different geographical areas. Furthermore, the cross-cultural validity and utility in the assessment of non-TBI neuropsychological disorders should be examined.

**REFERENCES**


York: Oxford University Press.

(Manuscript received 5 September, 2011; Revision accepted 9 July, 2012)