Cross-cultural Effect on Suboptimal Effort Detection: An Example of the Digit Span Subtest of the WAIS-III in Taiwan

Cheng-Chang Yang1, Chen-Jay Kao2, Ting-Wen Cheng3,4, Chi-Cheng Yang8, Wei-Han Wang1, Rwei-Ling Yu1, Yen-Hsuan Hsu1, Mau-Sun Hua1,4,5,6,7,8

1 Department of Psychology, National Taiwan University, Taipei, Taiwan
2 Department of Clinical Psychology, Taipei City Hospital, Songde Branch, Taipei, Taiwan
3 Clinical Psychology Center, National Taiwan University Hospital, Taipei, Taiwan
4 Department of Neurology, National Taiwan University Hospital, Taipei, Taiwan
5 Department of Psychiatry, National Taiwan University Hospital, Taipei, Taiwan
6 Neurobiology and Cognitive Science Center, National Taiwan University, Taipei, Taiwan
7 Graduate Institute of Brain and Mind Sciences, National Taiwan University, Taipei, Taiwan
8 Division of Clinical Psychology, Department of Occupational Therapy, Chang-Gung University, Taoyuan, Taiwan

*Corresponding author at: Department of Psychology, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan.
Tel.: +886-2-33663101; fax: +886-2-23695438.
E-mail address: huams@ntu.edu.tw (M.-S. Hua).
Accepted 22 August 2012

Abstract

Suppressed Digit Span performance has been proposed as an embedded indicator for suboptimal effort detection in neuropsychological evaluations in Western societies, particularly in the USA. However, its effectiveness in Chinese countries remains unexplored. The purposes of this study were first to explore normative Digit Span performance patterns between the Taiwan and American standardization samples, then to examine performances of patients with traumatic brain injury and with psychiatric diseases on the embedded measures (the Digit Span Scaled Score, Vocabulary minus Digit Span difference score, Reliable Digit Span, and the longest string of digits forward and backward) through retrospective data analysis. The normative Digit Span performance differs between the two cultural populations. Although litigating and nonlitigating participants perform differently on these measures, further prospective studies are needed to explore this issue with comprehensive external corroborating validity data.

Keywords: Wechsler scales; Intelligence tests; Traumatic brain injury; Neuropsychological tests; Incentives; Cross-cultural comparison

Introduction

Neuropsychological assessments can contribute important information on diagnosis, prognosis, and treatment to individuals with brain injury or disease (Lezak, Howieson, & Loring, 2004). However, test results may be useless because of patients’ feigning or suboptimal effort with various grounds, especially when incentives are present (Mathias, Greve, Bianchini, Houston, & Crouch, 2002; Tan, Slick, Strauss, & Hultsch, 2002). Recently, detection of suboptimal effort as an essential component of neuropsychological assessments has become a consensus (Heilbronner, Sweet, Morgan, Larrabee, & Millis, 2009).

Techniques of Suboptimal Effort Detection

In the Western societies, detection of poor effort or response bias in neuropsychological testing has followed two main approaches (Iverson, 2010). One has been the development of specific measures of effort, such as the Test of Memory Malingering (Tombaugh, 1996) which generally follows a forced-choice format. Similarly, the Portland Digit Recognition Test (Binder, 2003) is another early application of forced-choice recognition testing which takes 40–60 min to administer.
Although these dedicated tests are widely used in clinical practice and supported by empirical data, their vulnerability to coaching and time-consuming limitation have been strongly criticized (Grote & Hook, 2007).

Recently, investigators have recommended to employ the standard cognitive tasks to detect noncredible test performance as adjunct indicators of these free-standing symptom validity tests and to circumvent coaching (Babikian, Boone, Lu, & Arnold, 2006). The principle of detection method is identifying cutoff scores or response patterns associated with exaggeration on standard assessment instruments (Larrabee, 2007). Several neuropsychological tests have been examined for their ability to detect malingering, including the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; Mittenberg et al., 2001; Wechsler, 1997a), Wechsler Memory Scale-Third Edition (WMS-III; Killgore & DellaPietra, 2000; Wechsler, 1997b), Trail Making Test (O’Bryant, Hilsabeck, Fisher, & McCaffrey, 2003), Rey-Osterrieth Complex Figure Test (Lu, Boone, Cozolino, & Mitchell, 2003), Wisconsin Card Sorting Test (Bernard, McGrath, & Houston, 1996), Auditory Verbal Learning Test (Binder, Villanueva, Howieson, & Moore, 1993), California Verbal Learning Test (Millis, Putnam, Adams, & Ricker, 1995), and Finger Tapping Test (Arnold et al., 2005), among others. Since the WAIS-III is the most frequently used instrument in clinical settings (Rabin, Barr, & Burton, 2005) and has been developed within a large-scale standardization sample, interest in indices embedded within the WAIS-III has grown in recent years. For example, very poor performance on the Digit Span subtest is relatively uncommon in patients with numerous different types of brain damage (Iverson & Tulsky, 2003). The Digit Span subtest is easily and quickly administered either individually or as part of a battery, so is particularly user friendly in a clinical setting (Jasinski et al., 2011). Therefore, indicators derived from the Digit Span subtest have been recommended as effective measures for identifying suspect effort, such as the Digit Span age-corrected scaled score (DS-SS; Axelrod, Fichtenberg, Millis, & Wertheimer, 2006; Harrison, Rosenblum, & Currie, 2010; Heinly, Greve, Bianchini, Love, & Brennan, 2005; Iverson & Tulsky, 2003), the longest forward/backward span (Heinly et al., 2005; Iverson & Franzen, 1996; Iverson & Tulsky, 2003), Reliable Digit Span (RDS; Babikian et al., 2006; Greffenstein, Baker, & Gola, 1994; Harrison et al., 2010), and Vocabulary minus Digit Span difference score (VDS; Curtis, Greve, & Bianchini, 2009; Mittenberg et al., 2001; Schwarz, Gfeller, & Oliveri, 2006). Because borderline or impaired DS-SS (below 5 or 4) occur in less than 5% of both normal and clinical groups (Iverson & Tulsky, 2003), the Digit Span has been viewed as a promising effort indicator. In addition, Iverson and Tulsky (2003) offered that the longest digits forward and backward should rarely be less than five and three, respectively. Otherwise, RDS is computed by adding the longest string of digits forward and backward in which both trials at that level were passed (Greffenstein et al., 1994). Most investigators (e.g., Greffenstein et al., 1994; Larrabee, 2003) suggested a raw cutoff score of ≤ 7 in differentiating probable malingerers from patients with traumatic brain injury (TBI). Some researchers (Mittenberg et al., 2001) found that the suppressed Digit Span performance relative to Vocabulary subtest performance could be used as a suspicion index for poor effort because large VDS (≥ 5) were rare in the patients with TBI but yet relatively common in the analog simulators. In conclusion, increasing evidence reveals that the Digit Span subtest may have utility as a measure of suboptimal effort.

Cross-Cultural Issues

Although cultural, ethnic, and socio-economic factors should be taken into assessment and diagnostic considerations (Heilbronner et al., 2009; Judd & Beggs, 2008), clinicians assessing patients of a specific cultural group or minority populations are often forced to infer the validity, and justify the use, of common malingering instruments, basing largely on their own unique experience and expertise (Salazar, Lu, Wen, & Boone, 2007). Specifically, the Chinese have been the largest Asian subgroup in America and accounted for about 1% of the total U.S. population (U.S. Census Bureau, 2011a). Considering approximately 70% Chinese Americans are foreign born and about 1.5 m of them speak English less than “very well” (U.S. Census Bureau, 2011b), language and cultural factors may play a large role in their performance on neuropsychological tests developed on individuals of Euro-American background (Wong, 2011). In contrast to the robust data of poor effort detection in the Northern American and European countries, the literature addressing the utility of effort tests with non-Caucasians and individuals who speak other languages is sparse to nonexistent. In addition, few studies focused on patients in other societies, particularly native Chinese individuals, and their performance on measures designed to detect suboptimal effort (Salazar et al., 2007). Some researchers (e.g., Chiu & Lee, 2002; Gao, Liu, Ding, & Lu, 2002) in Hong Kong and China examined the effectiveness of various detection techniques commonly used in the Western countries, particularly the forced-choice paradigm, in identifying noncredible effort. Nevertheless, major limitations of these studies were the use of the simulation design and restricted to the technique of forced-choice paradigm, which markedly reduced the generalizability and increased vulnerability to coaching. Otherwise, in the past two decades in Taiwan, only a few studies (e.g., Chiang et al., 2006) paid attention to feigning and most of them did not focus on the issue of suboptimal effort detection. Therefore, the need to develop measures to objectively document suspect effort and noncredible cognitive impairment for individuals in Chinese societies was highlighted. Since the normative data of the WAIS-III Chinese version (WAIS-IIIC; Chen & Chen, 2002)
was derived from one representative, large-scale population in Taiwan, utilizing embedded indicators of suboptimal effort detection, especially the Digit Span from the WAIS-IIIC, might be beneficial to local clinical practitioners. However, although the atypical performance patterns on the Digit Span subtest were drawn from the Western societies, the Taiwanese normative performance should be examined. On the other hand, participants in most existing studies have been restricted to patients with TBI and few studies have investigated the extent to whether these indices might misclassify individuals with genuine psychiatric disorders as patients with suboptimal effort.

Goal of the Present Study

The aim of the present study was two-fold. The first purpose was to examine the universality of the Digit Span performance patterns through the contrast between Taiwan and American standardization samples. The second aim was to examine the rates of Taiwanese clinical samples identified by the embedded Digit Span-related symptom validity variables, specifically the DS-SS, longest digits forward, longest digits backward, RDS, and VDS.

Study 1: Cultural Difference in Normative Performance on the Digit Span

Methods

Participants

Taiwanese standardization sample. The WAIS-IIIC and the WMS-IIIC (Hua et al., 2005) standardization samples were included in this study. The standardization sample of the WAIS-IIIC was initially collected in 1999 on a stratified random sample of 888 participants in Taiwan. The sampling followed the major demographic distribution of the national population in Taiwan: gender (50.97% men, 49.03% women), age (16–84 years collapsed into 11 age groups), education level (less than grade 6, grades 7–10, high-school diploma, some post-secondary training, or university degree). Each age group consisted of 80 participants, with the exception of the oldest age group (75–84) that enrolled 88 participants. The WMS-IIIC published in 2005 was derived on another standardization sample in Taiwan. The WMS-IIIC sampling followed the WAIS-IIIC in gender, age, and education level, but the stratified random sample included 770 participants equally divided into 11 age groups. The sampling exclusion and inclusion criteria of the WAIS-IIIC and WMS-IIIC were patterned after those of the American versions.

American standardization sample. The WAIS-III standardization sample was presented in the Technical Manual (Psychological Corporation, 1997). The nationally stratified standardization sample consisted of 2,450 subjects divided into 13 age groups. Each age group contained 200 participants with the exception of the two oldest age groups, 80–84 (n = 150) and 85–89 (n = 100). The gender ratio was equal between the eight younger age groups, whereas the other age groups (older than 65) consisted of more women.

Procedures and statistics. Since the WAIS-IIIC and the WMS-IIIC did not provide available information for computing the VDS of the normal population, we only compared the longest string of digits forward and backward in the two Taiwan standardization samples with those in the American normative population.

The WAIS-IIIC and the WMS-IIIC standardization samples were compared by the independent t-test for the longest string of digits forward and backward for examining whether the Digit Span performance was stable across the two groups. These two samples would be combined into one Taiwan standardization sample if they performed similarly on the Digit Span. Then, further comparison of the two standardization samples (Taiwan and American) was performed by the independent t-test.

Results

The means and the standard deviations of the longest string of Digit Span forward and backward for the three groups are presented in Table 1. Since the two Taiwan groups did not differ in the longest digit forward, t(1656) = 0.97, p = .33, d = 0.05 (95% CI: −0.07, 0.21), and digit backward, t(1656) = 0.45, p = .65, d = 0.02 (95% CI: −0.13, 0.21), we merged the two Taiwan standardization samples into one combined group for the following comparison. The mixed Taiwan standardization
sample performed significantly more digits than the American standardization sample in the longest forward span, \( t(4106) = 27.09, p < .001, d = 0.31 \) (95% CI: 0.41, 0.61). In addition, the \( \chi^2 \) analysis indicated that the proportions of subjects performed below the cutoffs significantly differed in the longest digit forward, \( \chi^2(2, n = 4,108) = 76.57, p < .001 \), and digit backward, \( \chi^2(2, n = 4,108) = 35.25, p < .001 \), across the two cultural groups.

### Study 2: Digit Span Performance in Taiwan Clinical Population

#### Methods

**Participants**

**Traumatic brain injury.** One hundred and thirty-two patients with TBI were culled from the archival neuropsychological evaluation records consecutively referred by neurosurgeons or neurologists at National Taiwan University Hospital spanning between May 2005 and August 2011. To avoid confounding educational factors and ensure clinically useful comparisons, patients were excluded from the study if they had <6 years of education, previous history of developmental problems (e.g., mental retardation, autistic spectrum disorders, learning disabilities), or incomplete WAIS-IIIC scores. Finally, 96 TBI patients were included in the study. Thirty-three of these were classified as mild TBI while the remaining 63 were moderate or severe in severity according to the criteria set by the Mild Traumatic Brain Injury Committee of the Head Injury Interdisciplinary Special Interest Group of the American Congress of Rehabilitation Medicine (1993). Seventy-four (77.1%) TBI subjects revealed no presence of external incentive, whereas the other 22 participants (including 11 patients with mild TBI) were involved in litigation or pursuing insurance compensation claims.

**Psychiatric patients.** Participants were culled from the archival records of 253 inpatients or outpatients referred for intellectual evaluation at Taipei City Hospital, Songde Branch from January 2007 through August 2011. These psychiatric patients were diagnosed by psychiatrists according to the Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition, Text Revision (American Psychiatric Association, 2000). Per the aforementioned exclusion criteria, patients who had <6 years of education, with previous history of developmental problems (e.g., mental retardation, autistic spectrum disorders, learning disabilities), or without complete WAIS-IIIC scores were excluded. Finally, 110 psychiatric patients were included in the study. The breakdown for primary diagnosis was as follows: schizophrenia spectrum disorders (i.e., schizophrenia, schizoaffective, and psychotic disorder not otherwise specified) were the most common diagnoses (\( n = 72 \)), followed by depressive disorders (\( n = 22 \)), and bipolar disorders (\( n = 16 \)). Ninety-two (83.6%) psychiatric patients had no presence of external incentive, while the remaining 18 (16.4%) patients were involved in litigation or pursuing insurance compensation claims.

The WAIS-IIIC was administered by trained intern clinical psychologists or licensed clinical psychologists. The study was approved by our research ethics committees.

**Procedures and statistics.** Five Digit Span variables (the DS-SS, VDS, RDS, longest string of digits forward, and longest string of digits backward) collected for following analyses were reproduced in Table 2.

Descriptive analyses were conducted. For comparison of the four subgroups (TBI not in litigation, TBI in litigation, psychiatric disorders not in litigation, and psychiatric disorders in litigation), the analysis of variance with the Scheffe post hoc comparison (\( p < .05 \)) was performed to examine the group difference. Otherwise, if the data violated the assumption of
homogeneity of variance, the nonparametric Kruskal–Wallis test with the Mann–Whitney pairwise comparison was conducted. The statistical significance was set at \( p < .05 \). All statistical procedures were performed by using PASW 18.0 for Windows.

**Results**

Information regarding demographic and performance data for the four groups appear in Table 3. As can be seen, there were significant differences in age, \( \chi^2(3, n = 206) = 25.48, p < .001 \), education, \( F(3, 206) = 3.83, p = .011 \), Verbal IQ, \( \chi^2(3, n = 206) = 16.75, p = .001 \), Performance IQ, \( F(3, 206) = 12.72, p < .001 \), Full-Scale IQ (FSIQ), \( F(3, 206) = 10.29, p < .001 \), DS-SS, \( \chi^2(3, n = 206) = 12.31, p = .006 \), longest forward digits, \( \chi^2(3, n = 206) = 16.22, p = .001 \), and RDS, \( \chi^2(3, n = 199) = 16.79, p = .001 \), whereas there were no significant group differences in longest backward digits, \( F(3, 206) = 2.05, p = .109 \), and VDS, \( \chi^2(3, n = 198) = 1.21, p = .752 \).

Since reduced effort may been found in patients outside medical-legal circumstances (Locke, Smigielski, Powell, & Stevens, 2008), it is impossible to present the sensitivities, specificities, and accurate diagnostic rates in our study. It should also be acknowledged that litigating and nonlitigating patients falling below the cut-offs do not necessarily represent true-positive and false-positive errors, respectively. Cumulative percentages of the DS-SS, VDS, and RDS by groups are presented in Table 4. DS-SS of \( \leq 5 \) in nonlitigating samples were compatible with the WAIS-III norms, occurring in 12 (7.2%) participants. Of those participants with external incentive, 12 (30.0%) were below the cutoff proposed by Iverson and Tulsky (2003). Although the VDS of \( \geq 6 \) were uncommon, occurring in three (1.9%) nonlitigating participants, none of the litigating
participants scored above the cutoff. The incidence of the RDS scores of \( \leq 7 \) were very rare in patients with no external incentives, occurring in only 2 (1.3%) of the 159 participants. In contrast, eight (20.0%) of the litigating participants performed below the cutoff proposed by Greiffenstein and colleagues (1994).

Cumulative percentages for the longest string of digits forward and backward are presented in Table 5. The longest string of digits forward of \( \leq 4 \) was very unusual (0.6%) in nonlitigating patients, occurring in only one (1.4%) nonlitigating patient with TBI and none nonlitigating psychiatric patients. In contrast, the longest string of digits forward of \( \leq 4 \) for those in the litigating group occurred in six (15%) participants. Similarly, only 9 (5.4%) nonlitigating patients performed \( \leq 5 \) on the longest string of digits forward compared with 12 (30.0%) of litigating patients performing such scores. None of the nonlitigating patients performed the longest string of digits backward of \( \leq 2 \). However, two (5.0%) of the litigating patients performed the longest string of digits backward of \( \leq 2 \), and the proportion was compatible with the Taiwan standardization sample.

<table>
<thead>
<tr>
<th>Variable</th>
<th>TBI</th>
<th>Psychiatric group</th>
<th>No litigating (n = 74)</th>
<th>Litigating (n = 22)</th>
<th>No litigating (n = 85(^*))</th>
<th>Litigating (n = 18)</th>
<th>No litigating patients (n = 159(^*))</th>
<th>Litigating patients (n = 40)</th>
<th>Mixed Taiwan standardization sample (n = 1,658)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS-SS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤4</td>
<td>1.4</td>
<td>13.6</td>
<td>1.1</td>
<td>22.2</td>
<td>1.2</td>
<td>17.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8.1</td>
<td>27.3</td>
<td>6.5</td>
<td>33.3</td>
<td>7.2</td>
<td>30.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>12.2</td>
<td>31.8</td>
<td>20.7</td>
<td>50.0</td>
<td>16.9</td>
<td>40.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥7</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥6</td>
<td>1.4</td>
<td>–</td>
<td>3.5</td>
<td>–</td>
<td>2.5</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6.8</td>
<td>–</td>
<td>9.4</td>
<td>5.9</td>
<td>8.2</td>
<td>2.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>12.2</td>
<td>–</td>
<td>20.0</td>
<td>11.8</td>
<td>16.4</td>
<td>5.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤3</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤6</td>
<td>1.5</td>
<td>13.6</td>
<td>–</td>
<td>22.2</td>
<td>0.6</td>
<td>17.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.5</td>
<td>13.6</td>
<td>1.1</td>
<td>27.8</td>
<td>1.3</td>
<td>20.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6.0</td>
<td>18.2</td>
<td>12.0</td>
<td>61.1</td>
<td>9.4</td>
<td>37.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>16.4</td>
<td>31.8</td>
<td>26.1</td>
<td>72.2</td>
<td>22.0</td>
<td>50.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥10</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: TBI = traumatic brain injury; DS-SS = Digit Span Scaled Score; VDS = Vocabulary minus Digit Span difference score; RDS = Reliable Digit Span. 
\(^*\)n = 92 in the analysis of the DS-SS. 
\(^*\)n = 166 in the analysis of the DS-SS.
Discussion

The two studies were conducted to examine the normative performance measured by the Digit Span subtest of the WAIS-III across different cultural groups and to examine how many patients with TBI and psychiatric diseases in Taiwan classified by the embedded Digit Span-related symptom validity measures. Previous studies have less reported results regarding the impact of culture or ethnicity on measures of effort detection. Some findings of these effort detection indicators may be successfully utilized in patients from ethnic minority in the USA (Salazar et al., 2007) and others reporting instruments of the forced-choice format are as sensitive to detect poor effort in the native Chinese population as North American and Europeans (Chiu & Lee, 2002; Gao et al., 2002). Because relatively high sensitivities and specificities of the Digit Span subtest of the WAIS-III for effort detection have been reported in Western societies (Jasinski et al., 2011), it is important to understand the Digit Span-related performance patterns in Chinese societies.

In the first study, neither the longest string of digit forward nor the longest string of digit backward performed by the two standardization samples showed between-group equivalence on average. In addition, since the two Taiwanese normative data were compatible in the longest string of digit forward and backward, the performance pattern of the Digit Span in native Taiwanese population was reliable and consistent. In another study regarding normative data on cross-cultural neuropsychological tests in Mainland China (Hsieh & Tori, 2007), results of the Digit Span subtest of the WAIS-R were comparable with those of the WAIS-IIIC in Taiwan. Furthermore, our results are consistent with cross-language studies (Chan & Elliott, 2011; Chen & Stevenson, 1988; Chen, Cowell, Varley, & Wang, 2009; Stigler, Lee, & Stevenson, 1986) which found that the Chinese speakers outperformed their English counterparts in Digit Span tasks. Given that digit articulation times in Chinese are shorter than those in English (Chen & Stevenson, 1988; Stigler et al., 1986), our findings further support the “word-length effect” (Baddeley, Thomson, & Buchanan, 1975) in large-scale populations. Considering bilingual studies (Chincotta & Hoosain, 1995; Hoosain, 1979, 1984) demonstrated the Digit Span superiority in Chinese-English bilinguals’ mother tongue, our study highlighted the importance of spoken language in administering the Digit Span to Chinese Americans and interpretation of their performances. Generally speaking, the Digit Span performance difference might be based on cultural factors, particularly the linguistic feature, rather than on fundamental difference in cognitive resource between the two groups. Moreover, one recent retrospective study (Young, Sawyer, Roper, & Baughman, 2012) found effort indices from the recently revised Digit Span contribute to suboptimal effort detection with limited sensitivity. Since the “word-length effect” may play an important role in the revised portion of the Digit Span, namely Digit Span sequencing, utilizing these effort detection indicators derived from the new Digit Span version to Chinese Americans with caution is also suggested.

The purpose of the second study was to investigate the rates of Taiwanese clinical samples identified by the embedded Digit Span-related symptom validity variables proposed in the Western societies (low DS-SS, reduced number of the longest digit forward and backward, high VDS, and low RDS). Consistent with the previous studies, the second study indicates that litigating patients performed poorer on the intellectual assessment than nonlitigating ones. The DS-SS of ≤5 was relatively rare in litigating patients, in contrast to nearly one third of the patients in litigation performing scores below this cutoff. Otherwise, since the native Taiwanese produced more digits forward and backward than the U.S. sample did in the first study, applying the cutoffs from forward and backward digits derived from the U.S. standardization sample may take into consideration for the clinical populations in Taiwan. Our results suggest that a cutoff for the longest forward digits may partially differentiate the standardization sample and nonlitigating patients from patients in litigation in comparison with poor discriminability for the backward digits due to compatible proportions below the cutoff among groups. Similarly, since the RDS is associated with the numbers of longest forward and backward digits, a cutoff of ≤7 as proposed by Greiffenstein and colleagues (1994) may also be questioned based on the results of our first study. Our results indicate that RDS of ≤8 for patients with TBI and ≤7 for patients with psychiatric diseases may identify the usual rates in nonlitigating groups. However, contrary to Iverson and Tulskey’s (2003) recommendation, a cutoff for the VDS showed low discriminability for our samples. Interestingly, our results reveal that nonlitigating patients show more discrepancy between the Vocabulary and DS-SS. In fact, recent evidence (e.g., Axelrod et al., 2006; Harrison et al., 2010) suggested that the VDS might not be an effective indicator for effort detection but further studies are needed to explore possible explanations. It is especially noteworthy that our litigating samples performed lower on the WAIS-IIIIC compared with the nonlitigating groups, suggesting that the relatively low performance on the Digit Span indices in our litigating sample may be secondary to low ability. However, the discrepant patterns persisted after controlling for their premorbid IQ (data not shown). Given that the premorbid IQ estimates derived from an unpublished work (Chen, Hua, Zhu, & Chen, 2006) accounted for about only one third of the variance in FSIQ, the possibility of different premorbid function levels between groups contributing to the lower scores on Digit Span performance indices may not be ruled out.

Limitations of the retrospective study must also be noted. Patients with or without incentive recorded from the documents may not represent their real motive and effort level. The sample sizes of the litigating groups are apparently too small to conduct further analysis. In addition, the high exclusion rate raises the question of selection bias and reduces generalizability.
In fact, the high exclusion rate resulted from patients with incomplete WAIS-III scores because they were assessed with abbreviated versions of the WAIS-IIIC (Chen, Hua, Zhu, & Chen, 2008). These tetrad short-forms did not include the Digit Span subtest. Moreover, subjects with <6 years of education and with history of developmental problems were excluded as well, thus limiting the sampling representativeness. Furthermore, clinical information (e.g., medication, duration of illness, age of onset, and severity of psychiatric symptom) of our clinical populations was incomplete or missing.

In summary, findings of the current study reveal that the normative performance pattern of the Digit Span is not compatible between the populations in Taiwan and America. Except for the VDS, most of the embedded Digit Span-related symptom validity measures proposed in the Western societies classify relative few nonlitigating participants below cutoffs in Taiwan. We thus suggest that practitioners in Euro-American regions may pay attention to the cultural-linguistic effect while administering the Digit Span to their bilingual Chinese clients, irrespective of external incentive. In Chinese societies, the Digit Span may be an essential effort detection instrument as an adjunct indicator. Future studies in cross-validation may conduct known-group designs and examine its utility for other populations, such as patients with chronic pain or criminals who claimed mental disability. However, it is important to note that the Digit Span may not be used as the sole instrument for effort detection as Axelrod and colleagues (2006) suggestion. Thus, it would be quite useful in combination with other effort indices for detecting incomplete effort if possible. Since there is no other validated instrument for effort detection in Taiwan, it is urgent to examine the effectiveness of other tests utilized in forensic neuropsychological evaluations in the Western societies.

**Funding**

This work was partially supported by grants from the National Science Council (NSC-98-2410-H-002-026-MY3) and Taipei City Hospital (10101-62-010), Taipei, Taiwan.

**Conflict of Interest**

None declared.

**References**


