Exploration of Malingering Indices in the Wechsler Adult Intelligence Scale-Fourth Edition Digit Span Subtest

Caitlin S. Reese*, Julie A. Suhr, Tara L. Riddle

Department of Psychology, Ohio University, Athens, OH, USA

*Corresponding author: Department of Psychology, Ohio University, Athens, OH 45701, USA. Tel.: +1-740-593-1707; fax: +1-740-593-0579.
E-mail address: cr145708@ohio.edu (C. Reese).

Accepted 23 December 2011

Abstract

Prior research shows that Digit Span is a useful embedded measure of malingering. However, the Wechsler Adult Intelligence Scale-IV (Wechsler, 2008) altered Digit Span in meaningful ways, necessitating another look at Digit Span as an embedded measure of malingering. Using a simulated malingerer design, we examined the predictive accuracy of existing Digit Span validity indices and explored whether patterns of performance utilizing the new version would provide additional evidence for malingering. Undergraduates with a history of mild head injury performed with best effort or simulated impaired cognition and were also compared with a large sample of non-head-injured controls. Previously established cutoffs for the age-corrected scaled score and Reliable Digit Span (RDS) performed similarly in the present samples. Patterns of RDS length using all three subscales of the new scale were different in malingerers when compared with both head-injured and non-head-injured controls. Two potential alternative RDS scores were introduced, which showed better sensitivity than the traditional RDS, while retaining specificity to malingering.

Keywords: Head injury; Malingering; Assessment

Introduction

Malingering of cognitive impairment remains a primary concern for neuropsychological researchers and practitioners alike, given the high base rates of probable malingering, especially in mild head injury (Bender & Rogers, 2002; Larrabee, 2007; Mittenberg, Patton, Canyock, & Condit, 2002). Although there are many well-validated standalone measures of malingering (Bauer & McCaffrey, 2006; Grote & Hook, 2007), access to information about such measures is readily available, making them potentially vulnerable to coaching (Bauer & McCaffrey, 2006; Lees-Haley, 1997; Ruiz, Drake, Glass, Marcotte, & van Gorp, 2002; Wetter & Corrigan, 1995). Embedded measures of malingering (i.e., patterns of performance on traditional neuropsychological measures that are not consistent with neurological injury) may be more robust to coaching, although they may also be less sensitive to malingering (Gutheil, 2003; Suhr & Gunstad, 2007).

One neuropsychological test commonly used as an embedded measure of malingering is Digit Span. Jasinski, Berry, Shandera, and Clark (2011) recently conducted a meta-analysis of two Digit Span malingering indices, including Reliable Digit Span (RDS; Greiffenstein, Baker, & Gola, 1994) and age-corrected scaled score, from the revised or third editions of the Wechsler Adult Intelligence Scale (WAIS) and Wechsler Memory Scale. Based on the results of 22 studies, they found RDS and age-corrected scaled score to be equally effective in their ability to distinguish malingering from non-malingering groups (of note, all control groups were either patient populations or had history of head injury or other neurological condition). As the accuracy of RDS was more consistent across studies, they suggested that it may be a preferable measure of malingering.

However, given recent revisions of the Wechsler scales, further examination of the malingering indices on this measure is important. The WAIS-IV (2008) Digit Span subtest differs from the Digit Span used in all prior studies in two ways. First, individual items in Digit Span Forward and Backward were modified so that the numbers 1–9 were evenly distributed throughout the trials and were also reordered within some items so that numbers with similar sounds (e.g., 5 and 9) did not appear in the...
same trial. Thus, the items within these two subscales are not identical to their previous versions, although this likely has minimal effect on the overall difficulty level of the test. A much larger change was the introduction of a new subscale, Sequencing, which was designed to increase the working memory demands of the task (Wechsler, 2008). In the present study, we examined the accuracy of the age-corrected total scaled score and the traditional RDS index calculated from the new version of Digit Span in the detection of malingering, and we explored whether patterns of performance that include the new Sequencing subtest could be used to identify new embedded measures of malingering.

Method

Participants

Participants were taken from two ongoing studies in the second author’s research laboratory. The first group of participants included 96 undergraduate students who endorsed a history of mild head injury/concussion involving a loss of consciousness that lasted between a few seconds and 20 min, but who were not currently in treatment for or engaged in litigation concerning the injury, who were participants in a simulated malingering study. The second group of participants included 105 undergraduate students who had completed the Digit Span as part of a larger battery of tests being given in order to validate an experimental executive functioning measure. In both groups, participants were excluded if they reported current diagnosis of psychological conditions or neurological conditions other than mild head injury; in the case of participants randomly assigned to malingering cognitive impairment, participants who reported at the end of the experiment that they did not comply with malinger cognitive impairment, participants who reported at the end of the experiment that they did not comply with malingering instructions were also excluded from data analysis (N = 1).

Procedures

All data were gathered in compliance with institutional research standards for human research, in accordance with the Helsinki Declaration. Participants taken from the first study were randomly assigned to one of four conditions: observed malingering, unobserved malingering, observed control, and unobserved control. Control participants were instructed to perform with their best effort, while malingers were asked to realistically portray what they believed traumatic brain injury symptoms might “look like” during neuropsychological testing (see Appendix A for detailed instructions). “Observed” participants were told that their session was going to be videotaped and completed their measures in the presence of a video camera set up in plain sight, off to the side, oriented at a right angle to the plane of interaction between the experimenter and examinee; none was actually video recorded. All participants were administered the WAIS-IV (Wechsler, 2008) Digit Span subtest within the context of a full neuropsychological battery that took ~2 h to complete. Although examiners were aware of the presence/absence of the video camera, they were unaware of malingering/non-malingering status of participants. All participants completed a compliance check following the study, which asked them to (a) restate the instructions they were given, (b) rate the difficulty they had in following the directions, and (c) rate their perceived success at following the directions; those in the malingering group were also asked to rate how frequently they engaged in a number of behaviors commonly endorsed by malingers in prior simulation studies. As noted above, only one participant (from the malingering group) was removed from analyses because he reported the highest difficulty rating and lowest success rating, rated himself as “never” engaging in any of the malingering strategies, and spontaneously reported to the examiner that he would “never” be able to simulate because of his own personal values. Participants were then debriefed about the nature of the study and compensated for their time. Initially, participants received only experimental course credit for participation; following slow recruitment, most participants were offered $10 for their participation (not for their performance or ability to comply with study instructions) in addition to experimental course credit. Initial analyses for the present study showed that observation status, source of participant, and payment status did not affect performance on the Digit Span, and thus, for the present analyses, data were collapsed across these conditions.

Participants in the second study were recruited to complete a 2-h battery of neuropsychological tasks focused on executive functioning, which included the WAIS-IV Digit Span, for experimental course credit. Participants also completed the Word Memory Test (Green, 2005), and none was identified as malingering; therefore, these participants were considered non-malingering controls. Of note, 16 participants reported a history of mild head injury consistent with the definition used in the first study. All reported having sustained their injury as a result of sports, with the exception of one (non-vehicular accident), and all reported loss of consciousness of <5 min, with the exception of one participant with loss of consciousness of 20–30 min. These participants were included in the head-injured control sample of the first study, resulting in three groups for the present analyses: malingering (N = 46), head-injured controls (N = 63), and non-head-injured controls (N = 93).
Results

The mean age of participants was 18.99 years (SD 1.01, range 18–24). Of the whole sample, 49.5% were men and 88.6% of participants identified as Caucasian/Non-Hispanic (5.4% African American, 1.5% Asian American, 4.5% other/biracial). Groups were not different in age, \( p = .55 \), or race/ethnicity, \( p = .76 \). The non-head-injured controls had proportionally more women than the other two groups, \( p = .037 \).

Analysis of Existing Digit Span Malingering Indicators

Jasinski and colleagues (2011) suggested an age-corrected scaled score cutoff of 6 on the old versions of WAIS Digit Span for the assessment of malingering. Analysis of age-corrected scaled scores on the new WAIS suggested that a score of 6 or lower resulted in 56% sensitivity, with 97% specificity to head-injured controls (96% to non-head-injured controls). When the two control groups were collapsed together, positive predictive accuracy was 81%, while negative predictive accuracy was 86%. A score of 7 or lower was 63% sensitive to malingering, with still acceptable specificity (91% in head-injured controls; 94% in non-head-injured controls). For this score, collapsed across the two control groups, positive predictive accuracy was only 69%, whereas negative predictive accuracy was 89%.

Using the Jasinski and colleagues (2011) RDS cutoff of 7 resulted in 65.2% sensitivity, but only 84% specificity (relative to head-injured controls; 95% to non-head-injured controls). Positive predictive accuracy was 70% and negative predictive accuracy was 90%. Although this is comparable with accuracy rates reported in the Jasinski and colleagues meta-analysis, we assessed whether a slightly lower cutoff might result in higher specificity, which has been suggested in some other studies (Babikian, Boone, Lu, & Arnold, 2006; Duncan & Ausborn, 2002; Heinly, Greve, Bianchini, Love, & Brennan, 2006). A cutoff of 6 or lower on the RDS resulted in lower sensitivity (43%), but with 98% specificity to head-injured controls (99% specificity to non-head-injured controls); positive predictive accuracy improved to 91% and negative predictive accuracy was 86%.

Exploration of Potential Malingering Patterns Utilizing the New Sequencing Subtest

In order to explore overall patterns on the three Digit Span subtests within the three groups, we calculated the RDS length for each of the three subtests and conducted a 3 (group) by 3 (subtest) mixed-measures analysis of variance using those data (Table 1). The main effects for subtest, \( F(2, 398) = 261.97, p < .001 \), and group, \( F(2, 199) = 47.98, p < .001 \), were both significant, as was the interaction between subtest and group, \( F(2, 398) = 6.88, p < .001 \). Follow-up repeated-measures analysis for each group showed that, for both control groups, the longest RDS was on the Forward subtest, with a significant decrease in RDS length on the Sequencing subtest \( (p = .002 \text{ for head-injured controls}, p = .004 \text{ for non-head-injured controls}) \), followed by another significant decrease in RDS length on the Backward subtest \( (p < .001 \text{ for both groups}) \). However, for the malingering group, RDS length for Forward and Sequencing subtests was not significantly different from one another \( (p = .36) \).

Table 1. Means and standard deviations (SD) for the RDS lengths and versions of RDS total scores for each subtest, separately by group

<table>
<thead>
<tr>
<th></th>
<th>Malingering (( N = 46 ); mean [SD])</th>
<th>Head-Injured Controls (( N = 63 ); mean [SD], ES)</th>
<th>Non-Head-Injured Controls (( N = 93 ); mean [SD], ES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward length*</td>
<td>4.2 (1.2)*</td>
<td>6.1 (1.2)*, 1.58</td>
<td>5.9 (1.1)*, 1.47</td>
</tr>
<tr>
<td>Backward lengthb</td>
<td>2.7 (1.1)</td>
<td>3.5 (1.1)*, 0.73</td>
<td>4.1 (1.1)*, 1.27</td>
</tr>
<tr>
<td>Sequencing length*</td>
<td>4.0 (1.6)</td>
<td>5.6 (1.0), 1.19</td>
<td>5.6 (1.2), 1.13</td>
</tr>
<tr>
<td>Standard RDS*</td>
<td>6.9 (2.1)</td>
<td>9.6 (1.9), 1.35</td>
<td>10.1 (1.8), 1.64</td>
</tr>
<tr>
<td>Alternative RDSd</td>
<td>8.2 (2.6)</td>
<td>11.7 (1.9), 1.54</td>
<td>11.5 (1.9), 1.45</td>
</tr>
<tr>
<td>Enhanced RDSd</td>
<td>10.9 (3.4)</td>
<td>15.2 (2.6), 1.42</td>
<td>15.7(2.6), 1.59</td>
</tr>
</tbody>
</table>

Notes: ES = effect size relative to malingering group; RDS = Reliable Digit Span.
*Malingering lower than both controls.
bWithin group Backward length significantly different than other subtests.
dWithin groups all three subtests different from one another. Traditional RDS = Reliable Digit Span; alternative RDS = reliable digits forward plus reliable digits sequencing; enhanced RDS = sum of reliable digits forward, reliable digits backward, and reliable digits sequencing.
dAll three groups different from each other.
Difference between the head-injured controls and the non-head-injured controls effect size = −0.55.
although both were significantly different from the Backward subtest (both \( p < .001 \)). When comparing the three groups on each subtest individually, on the Forward and Sequencing subtests, the malingering group had lower RDS length than both control groups (Bonferroni corrected: \( p < .001 \) for both contrasts and both subtests), who did not perform differently than one another (Bonferroni corrected: \( p = .19 \) Forward subtest, \( p = 1.0 \) Sequencing subtest). However, for the Backward RDS length, all three groups performed significantly differently from one another, with the non-head-injured controls having the longest span (Bonferroni corrected: \( p = .002 \)), followed by the head-injured controls, who had a longer span than the malingers (Bonferroni corrected: \( p = .001 \)).

Overall, these analyses suggested that using the RDS for the Sequencing subtest to create alternative RDS scores might be useful in the detection of malingering. We examined two possible alternate scores: (a) replacing the Backward RDS score with RDS length from the Sequencing subtest (alternative RDS), and (b) including the Sequencing RDS score with the traditional Forward and Backward scores (enhanced RDS). Results of several cutoffs for each of these scores appear in Table 2.

**Discussion**

Results suggest that the recently revised Digit Span subtest is still useful as an embedded measure of malingering and that the addition of the Sequencing subtest may actually enhance usefulness of the RDS index in malingering detection.

Overall, our findings for the Digit Span scaled score cutoffs and for the traditional RDS cutoffs are similar to those reported in the Jasinski and colleagues (2011) meta-analysis. However, the traditional RDS showed unacceptable specificity at their recommended cutoff and was less sensitive to malingering when using a lower cutoff. Of note, a recently released supplement to the WAIS-IV (NCS Pearson, 2009) provided detection rates for the RDS in the WAIS-IV clinical samples; a cutoff of 6 or lower on RDS was generally seen in \( \leq 15\% \) of the clinical sample, with unacceptably high false-positive rates in their Traumatic Brain Injury, Schizophrenia, Intellectual Disability, Autistic Disorder, Reading Disorder, and Mathematics Disorder samples, but acceptable specificity in their Temporal Lobectomy, Major Depressive Disorder, Anxiety Disorder, Asperger’s Disorder, and Attention Deficit/Hyperactivity Disorder samples. The cutoff of 7 had unacceptably high false-positive rates in all clinical samples. Thus, our data and the WAIS-IV data suggest that a lower cutoff for the traditional RDS might be needed when it is calculated using the new WAIS.

In our samples, the Digit Span Sequencing subtest appeared to discriminate malingering from non-malingering groups equally as well as the two traditional Digit Span subtests, and perhaps better than the Backward subtest, on which the mild head injury participants performed worse than non-head-injured controls. These analyses prompted the development of the alternative RDS and the enhanced RDS, both of which had cutoffs that showed better sensitivity to malingering (when keeping specificity high), relative to the traditional RDS. Our data suggest that the enhanced RDS at the 11 cutoff is the best of the new measures, although this conclusion is based only on our sample and requires replication in independent samples. Of note, the new indices did not appear to be superior to using age-corrected scaled score cutoffs to detect malingering.

Although the data came from a simulated malingering study, several methodological considerations were incorporated into the study design to strengthen the internal validity, including using participants with a self-reported mild head injury history and administration of a full neuropsychological battery rather than just the measure of interest. In addition, those assigned to the malingering group were given a handout from an attorney’s website (see Appendix), in an attempt to make the instructions

<table>
<thead>
<tr>
<th>Score</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPA (%)</th>
<th>NPA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Head-injured</td>
<td>Non-head-injured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alternative RDS (reliable digit forward plus reliable digit sequencing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 or lower</td>
<td>20</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>7 or lower</td>
<td>40</td>
<td>98</td>
<td>99</td>
<td>89</td>
</tr>
<tr>
<td>8 or lower</td>
<td>52</td>
<td>94</td>
<td>97</td>
<td>77</td>
</tr>
<tr>
<td>9 or lower</td>
<td>78</td>
<td>87</td>
<td>83</td>
<td>72</td>
</tr>
<tr>
<td>Enhanced RDS (sum of reliable digit forward, reliable digit backward, and reliable digit sequencing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 or lower</td>
<td>33</td>
<td>98</td>
<td>99</td>
<td>86</td>
</tr>
<tr>
<td>10 or lower</td>
<td>43</td>
<td>98</td>
<td>99</td>
<td>89</td>
</tr>
<tr>
<td>11 or lower</td>
<td>59</td>
<td>94</td>
<td>97</td>
<td>77</td>
</tr>
<tr>
<td>12 or lower</td>
<td>72</td>
<td>83</td>
<td>88</td>
<td>58</td>
</tr>
</tbody>
</table>

Notes: PPA = positive predictive accuracy; RDS = Reliable Digit Span; NPA = negative predictive accuracy.

Table 2. Cutoffs for new versions of RDS and Detection Rates
comparable with what an attorney might share with a client or to what someone scanning the Internet for information to assist them in malingering might find. However, there were some limitations to our study design. Participants were generally healthy, successful undergraduate students, and even the participants with self-reported head injury were not actively seeking treatment or compensation for their injury, which does limit the generalizability of the findings. In addition, participants from the first study endorsed having a head injury with a brief loss of consciousness, but no other details about the injury itself were collected for the study. However, it should be noted that there were still differences between the head-injured controls and non-head-injured controls on digits backward, with medium effect size. Nevertheless, our head-injured control group should not be considered equivalent to a head-injured patient sample, for whom there is medical documentation of an actual brain injury.

Given these limitations, as well as the data from NCS Pearson (2009) suggesting poor specificity of traditional RDS in clinical groups, further studies should examine the detection accuracy of our proposed revisions to the RDS, especially the specificity of the cutoffs for clinical samples of individuals with more severe head injuries or with other neurological conditions, and in samples with broader age and education range.

**Funding**

There were no sources of financial support for this study.

**Conflict of Interest**

None declared.

**Appendix**

**Directions for Malingering Group in Study 1**

Today you will take a series of neuropsychological tests that assess motor speed, attention, memory, and thinking skills. You are being asked to believably pretend that you have significant problems (e.g., representative of brain damage) with motor speed, attention, memory, and thinking skills tests. In other words, pretend that you are someone involved in a lawsuit, and you want to pretend to have brain damage in order to win a financial settlement. What might you do to indicate (even though you do not) that you have permanent and significant problems with motor speed, attention, memory, and thinking skills while taking neuropsychological tests? There is no wrong answer. Supplemental materials have also been provided for you to read and give you additional ideas regarding how to pretend to have significant problems with motor speed, attention, memory, and thinking skills. The examiner who gives you the tests does not know that you will be pretending to have significant problems representative of brain damage, and of course, you do not want to get caught pretending; therefore, it is important to remind you to believably pretend to have such problems however you see fit. Please take a few moments to review the supplemental materials that you have received (full supplement provided at: http://www.braininjury.com/symptoms.shtml). Please do not share your instructions with anyone else, including the individual who gives you your tests, as different people have different instructions. If you have any questions, you may independently ask the individual who handed you these instructions. You will be taken to the testing environment in which you will take the tests in a moment.

**Directions for Head-Injured Control Group in Study 1**

Today you will take a series of neuropsychological tests that assess motor speed, attention, memory, and thinking skills. You are asked to give your best effort on all of these tests. Please do not share your instructions with anyone else, including the individual who gives you your tests, as different people have different instructions. If you have any questions, you may independently ask the individual who handed you these instructions. You will be taken to the testing environment in which you will take the tests in a moment.

**References**
