The Malingering Factor

J. Michael Williams*

Department of Psychology, Drexel University, Philadelphia, PA, USA

*Corresponding author at: Department of Psychology, 3141 Chestnut Street, Drexel University, Philadelphia, PA 19104, USA. Tel: +1-856-503-7079.
E-mail address: jw37@drexel.edu (J.M. Williams).
Accepted 18 January 2011

Abstract

The influence of malingering and suboptimal performance on neuropsychological tests has become a major interest of clinical neuropsychologists. Methods to detect malingering have focused on specialized tests or embedded patterns associated with malingering present in the conventional neuropsychology tests. There are two stages to the study of their validity. The first stage involves whether the method can discriminate malingering subjects from those who are not malingering. In the second stage, they must be examined for their relationship to the conventional tests used to establish impairment and disability. Constantinou, Bauer, Ashendorf, Fisher, and McCaffrey (2005. Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tests? Archives of Clinical Neuropsychology, 20, 191–198.) conducted the only study in which correlations are presented between a commonly used symptom validity test, the Test of Memory Malingering (TOMM) and the subtests of the Wechsler Adult Intelligence Scale-Revised (WAIS-R). A factor analysis was conducted using these correlations. It revealed a clear malingering factor that explained significant variance in the TOMM and the WAIS-R subtests. The relationship of malingering with cognitive tests is complex: some tests are sensitive to malingering and others are not. Factor analysis can summarize the magnitude of variance associated with each test and reveal the patterns of interrelationships between malingering and clinical tests. The analysis also suggested that malingering assessment methods could be improved by the addition of timing the responses.

Keywords: Malingering; Factor analysis; Intelligence; Neuropsychology assessment

Introduction

The assessment of malingering, poor effort, and suboptimal performance has become a necessary part of the neuropsychological evaluation as the field has recognized the significant influence of malingering on test performance, and research investigations are converging on a set of malingering assessment methods (Boone, 2007; Larrabee, 2007a). The development of these methods began with worse-than-chance techniques, in which the malingering was detected when subjects performed significantly worse than chance on recognition memory tasks (Binder & Pankratz, 1987). The next method utilized the recognition memory format but relied on normative, criterion-based probability judgments in order to infer that the subject was malingering. The inference is usually mediated by a cutoff score on the specialized test (Green, 1996; Tombaugh, 1996). The third method involves indicators present in conventional neuropsychological tests. These are used to construct a probability inference that the subject was malingering on the test (Larrabee, 2003; Wolfe et al., 2010).

The validity studies used to examine these methods followed a conventional model in which criterion groups were discriminated using the method. However, malingering validity studies are unique in two ways. First, in contrast to every other construct in neuropsychology, such as aphasia and memory disorder, it is impossible to construct a criterion sample of malingering subjects. Due to the nature of the disorder, true malingering subjects will never volunteer for a validity study. Instead, investigators have relied on two other comparison groups, subjects instructed to maling, and patients involved in some type of litigation or disability determination. It was assumed that these groups represent a facsimile of true malingering (Franzen, Iverson, & McCracken, 1990; Powell, Gfeller, Hendricks, & Sharland, 2004).

© The Author 2011. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
The second unique aspect of validity studies in malingering is the requirement that the results of the malingering assessment method must generalize to the conventional tests used to determine impairment and disability. It is not sufficient to prove that malingering subjects performed as predicted on the malingering assessment. This result must also predict performance on conventional tests. If the predictive power of the malingering test is not established then the validity of the test is still unknown.

There is considerable evidence that the pattern of relationships between malingering and neuropsychological tests is complex (Haines & Norris, 1995; Larrabee, 2007b). Some tests are very sensitive to malingering and others show very little influence. The fact that most specialized symptom validity tests are constructed from recognition memory tests suggests that the recognition memory format is more sensitive to malingering than recall formats (Beetar & Williams, 1995; Bernard, 1990; Wiggins & Brandt, 1988). The creation of a symptom validity measure by subtracting the Digit Span subtest score from the Vocabulary score of the Wechsler Adult Intelligence Scale (WAIS) suggests that Vocabulary is less affected by malingering than Digit Span (Mittenberg, Theroux-Fichera, Zielinski, & Heilbronner, 1995). Finally, numerous studies using discriminant function analysis in which a variety of test scores is set as predictors of malingering status suggest that many tests share variance with malingering status and many do not have a relationship (Babikian & Boone, 2007; Bernard, McGrath, & Houston, 1993).

There have been very few studies of malingering that bear directly on this second aspect of malingering test validity. Among these few, all but one used mean comparisons to demonstrate that malingering subjects performed worse than nonmalingering subjects on a number of neuropsychological tests (Green, 2007). Unfortunately, the mean contrast approach does not clarify the patterns of inter-relationship and the magnitude of variance accounted for by malingering. There are also a number of studies that have a bearing on the validity question but they were not designed to render the variance accounted for in general neuropsychological tests by the malingering method. For example, if an embedded measure of malingering exists in the performance on a traditional test, such as the Bolter Index for the Category Test (Bolter, Picano, & Zych, 1985; Denney, 2007), then the variance attributed to malingering on the test is obtainable from the relationship of the Bolter Index with the other scores on the test. Although marker variables such as the Bolter Index have been studied for many years, none of the studies examining its validity conducted the necessary data analyses, such as multiple regression and factor analysis. These methods would have rendered the amount of variance in the neuropsychological tests that are accounted for by malingering. In general, this observation suggests that many investigators of malingering methods have data sets that were used for simple mean comparisons that could now be analyzed according to a variance-partitioning model.

Explaining the variance associated with malingering contained in the scores on the neuropsychological test battery rests on an understanding of strategies used by malingering subjects when they approach the tests (Beetar & Williams, 1995; Williams, 1998). These strategies were examined in studies that included the self-report of people asked to malinger. The responses of these subjects were summarized into four main strategies: intentional wrong responses, performing slowly, behaving inattentively, and haphazard responding (Meyers, 2007). Since some tests are sensitive to these strategies and others are not, each of these strategies should produce patterns that are detectible using multivariate performance measures. Intentional wrong responding may explain the common finding that memory recognition is significantly lower than memory recall among malingering subjects (Beetar & Williams, 1995; Bernard, 1990). Since the correct and incorrect responses are presented to the subject in a recognition format, and a malingering subject presumably knows the correct answer, it is easy to make incorrect responses. Recall formats on tasks, such as list learning, require the subject to recite the list words presented before; recall on visual-spatial tasks require the subject to draw the figure presented previously. It is more difficult to construct an incorrect response because the subject does not have a clear idea of the responses associated with various score levels. As a result, malingering subjects perform worse than controls on both tasks but relatively worse on recognition tests (Beetar & Williams, 1995).

The strategy of slowness may explain worse performance on any timed test relative to untimed ones. Even tests that are not explicitly timed, such as Digit Span, measure a construct, Sustained Attention, that is very sensitive to the timing of responses and the immediate recall system (Cowan, 1992). If a subject delays responding past the limit of immediate recall, then the score will be low even when the conventional scoring does not involve explicit timing. For example, Beetar and Williams (1995) administered a computer-mediated version of the Digit Span that included timing of individual responses. They discovered that simulated malingering subjects reported the digits slowly, and their Digit Span scores were significantly lower than nonmalingering subjects.

Random or Haphazard responding may uniquely affect tests that rely on multiple responses over many trials. These tests include the Category Test (Reitan & Wolfson, 1993) and the Wisconsin Card Sorting Test (Sweet & Nelson, 2007). Unfortunately, response styles, such as these, have never been examined because they are not part of the conventional test score and they are only available on computer-mediated versions of the tests. Random responding may also explain the relative difference between recall and recognition memory described previously. Memory recognition trials available on conventional tests likely include sufficient trials to detect random responses but these patterns were not examined.
There is one study that examined the relationship between a symptom validity test and tests in the neuropsychological test battery that reported the correlations between the measures rather than simple mean comparisons. Constantinou, Bauer, Ashendorf, Fisher, and McCaffrey (2005) examined the relationship between the WAIS-Revised (WAIS-R: Wechsler, 1981) and the Test of Memory Malingering (TOMM: Tombaugh, 1996). The intent of the study was to examine the relationship of malingering on the TOMM with scores on the WAIS-R and the Halstead-Reitan Neuropsychological Test Battery (HRNB: Reitan & Wolfson, 1993). The subjects were patients who sustained a mild head injury who were referred for neuropsychological evaluations as part of litigation associated with the injuries. The investigators rendered sets of mean comparisons, effect size measures, and correlations between the subtests and summary scores of the WAIS-R and the TOMM. Relationships between these measures were apparent in the significant mean comparisons and correlations. However, the pattern of inter-relationships and the amount of variance accounted for were not as clear as they would be if multivariate analyses had been conducted. The missing analysis was factor analysis. Is there a malingering factor that accounts for variance among malingering tests and the tests used to establish impairment and disability? Will factors discovered by the analysis render patterns that suggest malingering strategies? The present study was designed to re-analyze the correlations in an attempt to answer these questions.

Methods

The correlations used in this analysis were derived from a sample of 69 patients with mild brain injuries involved in litigation or disability determination. They were administered the WAIS-R, the HRNB, and the TOMM as part of an outpatient evaluation. Ages ranged from 18 to 72 with a mean age of 42.41 (SD = 12.45). Education ranged from 7 to 22. The mean education was 12.96 (SD = 2.61). Scores on Trial 1 of the TOMM ranged from 15 to 50. The mean value was 39.70 (SD = 9.98). Trial 2 scores ranged from 13 to 50. The mean Trial 2 TOMM score was 43.22 (SD = 9.76). Twenty-two of the subjects scored below the cut-off score of 45 (M = 32.73; SD = 8.93). The correlation coefficients used in this study were based on the correlation of the WAIS-R subtests and Trial 2 of the TOMM. Full details concerning the sample are presented in Constantinou and colleagues (2005).

Although all the subtests of these batteries were administered, only the correlations between the subtests of the WAIS-R and TOMM were reported. These did not include the correlations of the WAIS-R subtests with each other, such as the correlation of the Vocabulary subtest with the Block Design subtest. Since these correlations are necessary for a factor analysis of the subtests, the unreported correlations were taken from the WAIS-R manual (Wechsler, 1981). The factor analysis was conducted using the Matrix option in SPSS. This option allows for a matrix input into multivariate procedures such as multiple regression and factor analysis. The analysis was conducted to allow up to 25 iterations and eigenvalues $> 1$. The analysis also included Varimax rotation (see the Appendix).

Results and Discussion

The typical factor analysis of the WAIS-R consistently reveals three factors. These cluster the Verbal subtests and Performance subtests and include a sustained attention factor that clusters the Digit Span and Arithmetic subtests (Sherman, Strauss, Spellacy, & Hunter, 1995; Waller & Waldman, 1990; Wechsler, 1981).

The analysis revealed these factors in a new configuration with a very clear malingering factor (Table 1). As expected, the TOMM had a very large loading on this factor. Digit Span, Arithmetic, and all of the Performance subtests of the WAIS-R

<table>
<thead>
<tr>
<th>Table 1. Factor analysis of the TOMM and WAIS-R subtests</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOMM</td>
</tr>
<tr>
<td>Information</td>
</tr>
<tr>
<td>Comprehension</td>
</tr>
<tr>
<td>Arithmetic</td>
</tr>
<tr>
<td>Similarities</td>
</tr>
<tr>
<td>Digit Span</td>
</tr>
<tr>
<td>Vocabulary</td>
</tr>
<tr>
<td>Digit Symbol</td>
</tr>
<tr>
<td>Picture Completion</td>
</tr>
<tr>
<td>Block Design</td>
</tr>
<tr>
<td>Picture Arrangement</td>
</tr>
<tr>
<td>Object Assembly</td>
</tr>
</tbody>
</table>

Notes: TOMM = Test of Memory Malingering; WAIS-R = Wechsler Adult Intelligence Scale-Revised. Loadings $> 0.40$ were considered salient.
loaded on this factor. Verbal subtests of the WAIS-R did not load on the malingering factor. This pattern suggests that among litigating patients a moderate amount of the variance on certain WAIS-R subtests is accounted for by malingering.

The tests that loaded on the malingering factor are presumably the tests from the WAIS-R that are most sensitive to malingering strategies. Of these strategies, which might explain the loadings? The two strategies most often mentioned by malingering simulators were intentional wrong responding and making slow responses. Clinical assessments of litigating subjects also describe slowness as a major feature of malingering and suboptimal performance (Iverson, 1995). When responses were timed by the computer during testing, malingering simulators were slow on virtually every aspect of testing (Beetar & Williams, 1995). Intentional wrong responding is presumably the dominant form of malingering. Although this strategy appears obvious, some tests are probably more affected by the strategy than others. Since the correct and incorrect choices are immediately available, any test in the format of multiple choices will be sensitive to intentional wrong responses. It is more difficult to respond with a clearly wrong response when the subject is required to construct an answer. The subjects may still perform less than they are capable but their relative responses will be more errors on the task that used multiple choices.

Intentional wrong responding and slowness represent clear, parsimonious explanations for the results of the factor analysis. Low performance on the TOMM requires intentional wrong responses. The items are presented in a multiple-choice format of two simple responses, correct and incorrect. The subject who is malingering presumably knows the correct answer and can therefore easily pick the wrong one. This likely explains why the TOMM is so sensitive to malingering.

The WAIS-R subtests are not presented in a multiple-choice format. However, the score on each test may be affected by intentional wrong answers. For example, it would be very easy to report the wrong answers on the Arithmetic subtest items. Picture Arrangement could easily be falsified by incorrectly ordering the picture cards. Although subjects can make wrong answers on the Block Design subtest, this is difficult because the correct model is clearly present while the subject is taking the test. Since the primary Verbal subtests, such as Vocabulary, Comprehension, and Similarities, are presented in a recall format, it is relatively difficult for the subjects to construct incorrect answers that sound credible. In their current format, these tests are apparently not sensitive to malingering. This finding supports the use of the Vocabulary minus Digit Span score as an indication of malingering (Iverson & Tulsky, 2003).

All the Performance tests are timed and therefore sensitive to a slowness strategy. A subject who performs slowly will receive a low score by simply extending the time over the maximum time per item allowed for trials of the Block Design, Picture Arrangement, Object Assembly, and Picture Completion. Digit Symbol is explicitly timed and slowness will dramatically lower the score. Time bonuses are an integral component of the Block Design score. Indeed, the subject can actually attain a zero score for all the Performance subtests by simply delaying responses. Among the Verbal subtests of the WAIS-R, slowness in responding to the Digit Span and Arithmetic subtests will cause the subject to respond past the interval of immediate recall and the digits and other information may be lost (Cowan, 1992; Smyth & Scholey, 1996; Sternberg, 1966). Delayed responses will not affect the scores on the other WAIS-R Verbal subtests.

Of course, it is possible that both strategies were employed, with the degree of influence relative to the sensitivity of each test to the strategies. For example, slowness is probably less likely to influence the TOMM score compared with intentional wrong responding. Time is likely to be the dominant strategy used on the Digit Symbol subtest. Most subjects do not make errors on Digit Symbol since the correct choices are always available at the top of the test form.

The findings also suggest methods to enhance test design and embedded measures to detect malingering. The obvious enhancement would be to include timing of more test items. This can be accomplished on many tests without changing the standard administration. For example, it is very easy to record the response times using computer-mediated tests. Currently, there are computer versions of commonly used clinical tests, such as the Category Test and Wisconsin Card Sorting Test. Recording the response times for these and the specialized effort tests would presumably enhance their validity as malingering markers. Time as an embedded measure may work effectively because subjects are generally unaware of it. For example, if there are false negatives on the TOMM because the subject notices the test is too easy, the subject may still behave consistent with malingering by slowing down and this can be measured by the assessment of response time.

It is important to state that this factor analysis is the first one to include a specialized test as a marker variable for malingering. Although the effect size was large for this study, the correlations between the TOMM and WAIS-R subtests were established using a sample size that was relatively small. Numerous factor analyses, replicating this study and examining many other tests, have yet to be done. These may qualify the results presented here but should clarify the relationship of malingering with the conventional tests even further. Fortunately, there are numerous studies using mean comparisons and classification methods that compiled data sets that could now be analyzed using multivariate procedures like factor analysis.

In general, this factor analysis suggests that methods derived from conventional measurement theory can deal with the second phase of validity studies of malingering assessment. The amount of variance accounted for by malingering in the WAIS-R Performance tests is considerable; the variance accounted for in the Verbal tests is much less. Contrary to the usual clinical use of symptom validity tests, the analysis suggests that the relationships are complex. Simply because a
subject does poorly on a symptom validity test does not mean the subject malingered on every test administered in the clinical battery. When they do engage in a malingering strategy, the effect size will vary with the sensitivity of the tests. In order to completely examine the relationship of malingering assessment with the clinical tests, many more tests should be examined, especially memory tests.

**Conflict of Interest**

None declared.

**Appendix**

IBM-PASW/SPSS Script for the Factor Analysis

```
matrix data variables = rowtype_ tomm inftest comp arith sim dspan vocab dsym pcomp bdesign parr obass.
begin data
  corr 1
  corr 0.35 1
  corr 0.34 0.68 1
  corr 0.55 0.61 0.57 1
  corr 0.34 0.66 0.68 0.56 1
  corr 0.52 0.46 0.45 0.56 0.45 1
  corr 0.3 0.81 0.74 0.72 0.52 1
  corr 0.49 0.44 0.44 0.46 0.42 0.47 1
  corr 0.54 0.52 0.52 0.48 0.37 0.55 0.42 1
  corr 0.45 0.5 0.48 0.56 0.51 0.43 0.52 0.47 0.54 1
  corr 0.44 0.5 0.48 0.46 0.5 0.37 0.51 0.39 0.51 0.47 1
  corr 0.44 0.39 0.4 0.42 0.43 0.33 0.41 0.38 0.52 0.63 0.4 1
n 1880 1880 1880 1880 1880 1880 1880 1880 1880 1880 1880 1880 1880
end data.
list.
FACTOR
/matrix = in(cor = *)
/VARIABLES tomm inftest comp arith sim dspan vocab dsym pcomp bdesign parr obass
/Analysis tomm inftest comp arith sim dspan vocab dsym pcomp bdesign parr obass
/Print Initial Extraction
/Criteria Mineigen(1) Iterate(25)
/Extraction PC
/Rotation VARIMAX
/Method = CORRELATION
/PRINT ROTATION.
```

**References**


