A Comparison of Four Tests of Malingering and the Effects of Coaching

Fredric E. Rose, Stuart Hall, and Allen D. Szalda-Petree

Department of Psychology, University of Montana

Paul J. Bach

Department of Psychology, University of Montana and Department of Neuropsychology, Western Montana Clinic

This study examined the ability of four measures of suboptimal performance to correctly classify four groups of subjects (normal controls, uncoached malingering, coached malingering, and head injured). Only the Portland Digit Recognition Test-Computerized (PDRT-C) identified simulating malingerers with greater than chance accuracy while minimizing false positives. Coached subjects were better able than their uncoached counterparts to avoid detection on all measures. In an additional analysis, a discriminant function using the response latency and total correct scores from the PDRT-C identified 70% of the coached malingerers on cross validation. The three other tests (Nonverbal Forced Choice Test, 21-Item Test, and Dot Counting Test) failed to obtain a satisfactory classification rate for the malingering groups as a whole and coached malingerers in particular.

The detection of malingered deficits in neuropsychological assessment has received considerable attention of late. Research investigating the use of standardized neuropsychological assessment techniques for such detection has met with variable results, suggesting that traditional tests may be of limited value when used for this purpose (cf., Bernard, 1990; Boone & Filskov, 1990; Faust, Hart, & Guilmette, 1988; Faust, Hart, Guilmette, & Arkes, 1988; Franzen, Iverson, & McCracken, 1990; Goebel, 1983; Heaton, Smith, Lehman, & Vogt, 1978;
Trueblood & Schmidt, 1993). Patterns of scores, but not necessarily the scores themselves, tend to separate malingerers from persons with true deficits. Differences on sensorimotor tasks and on tests of immediate attention (e.g., Digit Span) have shown promise as possible indices of effort on standard neuropsychological tests (e.g., Iverson & Franzen, 1996), yet even these indices have interpretive problems. For instance, clinical judgments based on test results alone are less than satisfactory. Statistical procedures are slightly better, but have yet to be adequately cross validated (Franzen et al., 1990). Recognizing this, researchers have begun to turn their focus from the use of standard neuropsychological tests toward the development of objective assessment devices designed specifically for the detection of suboptimal performance.

More recently developed tests of biased responding are based on the symptom validity paradigm described by Pankratz, Fausti, and Peed (1975) and exploit the lay person’s inexperience with the actual clinical presentation of the feigned disorder. Originally designed to assess any functional sensory deficit, symptom validity testing has been extended to detect feigned memory impairment (e.g., Pankratz, 1983; Binder & Willis, 1991; Frederick & Foster 1991; Iverson, Franzen, & McCracken, 1991). Symptom validity testing involves the administration of a simple two-alternative forced-choice task assessing the patient’s alleged deficit. By chance alone, at least 50% of the patient’s answers should be correct if the complaints are valid. The only plausible explanation for below chance performance would be that the patient knew the correct answer and chose to respond otherwise. However, research has shown that normal subjects simulating the effects of head injury as well as persons with mild head injuries with motivation to exaggerate do not reliably respond at below chance levels. Binder and Willis (1991) and Frederick and Foster (1991) found that a majority of their simulated malingerers did not perform below the level of chance despite scoring significantly below their cognitively impaired, honest responding groups. As a result, many current tests have developed decision rules for classification that utilize a cut-off score rather than the more conservative below chance criterion.

Some researchers have begun to explore the utility of more covert measures of biased responding as adjuncts to traditional scores. The advantage of this approach is that it becomes more difficult for the would-be malingerer to monitor performance. Rose, Hall, and Szalda-Petree (1995), for example, converted the Portland Digit Recognition Test (PDRT; Binder & Willis, 1991) to computer administration and added a measure of response latency. They found that the addition of the response latency measure significantly improved the test’s ability to identify simulated malingerers.

The present study compared the efficacy of four tests designed specifically for the detection of biased responding in neuropsychological assessment: The Portland Digit Recognition Test-Computerized (PDRT-C; Rose et al., 1995), Rey’s Dot Counting Test (Lezak, 1995; Rey, 1941), the Nonverbal Forced Choice Test (Frederick & Foster, 1991), and the 21-Item Test (Iverson et al., 1991). We also assessed the extent to which simulating malingerers who were provided with information about the effects of head injuries would be able to escape detection. That is, we examined the effects of coaching subjects on their test performance. Many studies have demonstrated that providing subjects instructed to malinger with such information results in test scores that are more like truly impaired patients than those of noninformed subjects (Boone & Filskov, 1990; Frederick, Sarfaty, Johnston, & Powell, 1994; Kerr et al., 1990; Rogers, Gillis, Bagby, & Monteiro, 1991). We hypothesized that of the four tests, the PDRT-C would be superior to the others in separating persons feigning cognitive impairment from persons with head injuries based on the test’s combined response latency and experimentally derived cut-off scores. Coached malingerers were expected to simulate head injury with greater success than uncoached malingerers, but were still hypothesized to be detected by the covert measurement of response latency on the PDRT-C.
METHODS

Participants

*Uncoached malingerers (UM).* Thirty college students free of neurological and psychological impairment served in this group. These students read the following scenario as part of their instructions:

Three weeks ago you were in a car accident that was not your fault. A car ran a stop sign while you were in the intersection and hit you. You hit your head against the dashboard and were knocked out for about 20 minutes. Afterwards, you felt ‘‘dazed’’ for a while. You had to stay in the hospital for two days with a severe concussion. Because another person caused this accident, you are now in a lawsuit to decide how much money you will get from the person responsible. The more severe your problems, the more money you will get in the law suit.

As part of the law suit a psychologist is about to examine you using several tests of your memory and problem solving ability. The purpose of the testing is to decide if there were any long-term effects from your head injury. Your goal is to convince your examiner, by your performance on these tests, that you have suffered brain damage from the accident.

Note that the scenario did not include a suggested test-taking strategy and there was no description of the effects of head injury. This scenario constituted the uncoached manipulation.

*Coached malingerers (CM).* Another 30 college students, also free of neurological and psychological impairment, served in this group. The CM group received the same instructions and scenario as the UM group, but were also provided with the following description of the cognitive impairment typically associated with head trauma:

[In order to convince the examiner that you have brain damage] you are to try to produce the most severe problems that you can without making it too obvious to the examiner. That means your ‘‘brain damage’’ must be believable. Major exaggerations, such as remembering absolutely nothing, are easy to detect. If the examiner does not believe that you have any problems you will not win your lawsuit and you will not get anything for your injuries. In addition you may be fined or jailed for lying in court. People who have a head injury often have problems paying attention, cannot remember things as well, and do not learn things as easily as they did before their injury. They also think a little slower than they used to. Keep this in mind when taking the tests. Remember, you are to try to produce the most severe problems that you can, mimicking the performance of persons who are truly injured.

Note that this added information suggests both a strategy (i.e., do not exaggerate deficits) and provides basic information about the effects of head injury. This additional information constituted the coached manipulation.

*Head-injured patients (CHI).* The CHI group was made up of 30 persons who had suffered a documented traumatic brain injury of at least moderate severity, defined as a loss of consciousness for more than 30 minutes and/or posttraumatic amnesia greater than 1 hour following the injury (Becker, Grossman, McLaurin, & Caveness, 1979; Jennett & Teasdale, 1981). The average length of unconsciousness for this group was 42 days (range = 1 hour–240 days). The average length of time since the injury was 6.8 years (range = 9 months–26 years). The level of cognitive impairment in this group was demonstrated by comparing scores on the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987) to normal control subjects. The CHI group scored significantly below controls on 10 of the 17 CVLT scores examined (Trial 1, Trial 5, List B, Short Delay Free and Cued Recall, Long Delay Free and Cued Recall, Cued Recall Intrusions, Discriminability, and False Positives),
TABLE 1
Demographic Variables

<table>
<thead>
<tr>
<th>Subject Group</th>
<th>Age</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Uncoached malingering</td>
<td>24.6</td>
<td>11.49</td>
</tr>
<tr>
<td>Coached malingering</td>
<td>22.7</td>
<td>7.07</td>
</tr>
<tr>
<td>Head injury</td>
<td>34.7</td>
<td>8.97</td>
</tr>
<tr>
<td>Normal control</td>
<td>31.8</td>
<td>9.38</td>
</tr>
</tbody>
</table>

Note. Column means with different superscripts were statistically different, \( p < .05 \).

There was a significant difference in the average age of the four groups, with CHI and NC groups being slightly older than the two malingering groups. This was not felt to jeopardize results because the magnitude of the difference was relatively small and previous work (Binder & Willis, 1991) found no effect of age on the malingering scores of the PDRT. There were no significant differences between groups in years of education.

Normal controls (NC). Thirty community volunteers, 16 of whom were friends or spouses of the CHI subjects, served in this group. These 30 participants were free of neurological and psychological impairment. As with the CHI group, all NC participants were instructed to take the tests to the best of their ability and received $10.00 for their participation.

Materials

Portland Digit Recognition Test-Computer. The PDRT-C was developed to closely resemble the manual version with the added benefit of measuring response latency (see Binder & Willis, 1991 and Rose et al., 1995 for a complete description of both versions). The PDRT-C involves viewing a 5-digit number for 5 seconds, performing an interference task for a specified delay period (subjects search for a target letter in groups of three or four letters appearing on the screen), and then identifying the 5-digit number presented previously from among two alternatives. The delay period between presentation and response increased from 5 seconds to 15 seconds, and finally to 30 seconds over the 72-item test. Feedback was provided after each response. Scoring was based on the traditional scores as developed by Binder (Binder, 1993; Binder & Willis, 1991).

Rey’s Dot Counting Test (Rey, 1941). This test, first described by Rey (1941), and discussed by Lezak (1995), consists of two parts: counting the number of dots randomly distributed on six 3 × 5 in (7.6 × 12.7 cm) cards and counting the number of dots that are grouped in an orderly fashion on a similar set of cards. Subjects were told to count the dots as quickly as possible. The answer and time taken to arrive at it are recorded for each card. The cooperative, nonmalingering subject’s time to count dots should increase gradually as the number of dots increases. More than one serious deviation in this pattern, according to Lezak (1995), is evidence that the subject is not acting in good faith. Furthermore, the total time to count grouped dots should be shorter than that in the ungrouped condition. When the time taken...
Comparison of Malingering Tests

The 21-Item Test (Iverson et al., 1991). This test consists of two 21-item word lists. One of the lists is read orally to the participant who is then asked to recall as many of the words as possible. The examiner next reads a list of 21 word pairs in which one word from each pair was presented previously, and the subject indicates which one was on the original list. This test yields several measures of suboptimal performance. First, the free recall portion can be used to analyze serial position effects. Second, subjects answering honestly should show better recognition than recall while the reverse pattern suggests dissimulation. Lastly, because a portion of this test involves symptom validity testing, even if recognition is better than free recall, performance below the chance level of 50% correct on the forced-choice portion suggests a less than honest presentation.

Nonverbal Forced Choice Test (Frederick & Foster, 1991). A modified version of the Test of Nonverbal Intelligence (TONI; Brown, Sherbenou, & Johnsen, 1982), the Nonverbal Forced Choice Test is a nonverbal form of symptom validity testing. Frederick and Foster’s modification involved combining the two alternate forms of the TONI (100 items total) and for each item eliminating all but the correct choice and the most effective distracter. The result was a two-alternative forced-choice test of problem solving ability, with at least two equivalent items at each level of difficulty. All items are administered in random order to eliminate the hierarchical presentation of items in order of difficulty. This test yields three performance indices: (a) total score; (b) slope of performance as a function of item difficulty; and (c) consistency ratio, a ratio of correctly answered pairs at the same level of difficulty to the total possible correct. The assumptions are that malingerers will score below the chance level, that the malingerer’s slope will rise as item difficulty increases (indicating that the subject was missing items early on that could have been answered correctly, changing to a more random guessing pattern as the individual’s ceiling is reached), and that a malingerer’s consistency ratio will be below chance level, indicating an inconsistent response pattern. Previous research (Frederick & Foster, 1991; Frederick et al., 1994) identified a decision rule for this test suggesting malingered performance. If a subject’s slope × consistency ratio is greater than −0.041 and his or her score was below average, malingering was suspected.

Procedure

Participants received their instructions (including the scenario and coached/uncoached information for CM and UM subjects) in a sealed envelope to keep the experimenters blind to the conditions and purposes of the experiment and to randomize subject assignment to the malingering conditions. Instructions for all groups began by stating that the experiment involved taking a variety of memory and problem solving tests, and that they should not tell the examiner the content of the instructions. The instructions for the two malingering groups also stated that if they successfully fooled the examiner into believing they had actual deficits they would be given twice the number of course credits originally promised for participation. These credits are required to pass their course and, therefore, served as an incentive to follow instructions. In fact, all participants who indicated trying at least moderately hard to follow the instructions, regardless of their success, received the bonus credits. Effort was determined by responses to a manipulation check questionnaire administered at the conclusion of the experiment.

Subjects participated on an individual basis. The tests were administered according to their published directions in the following order: the PDRT-C, the Nonverbal Forced Choice
Test, the Dot Counting Test, and the 21-Item Test. Following administration of all tests, participants completed a manipulation-check questionnaire which asked each person to recall their instructions, rate the amount of effort put forth to follow them (on a 5-point scale), and indicate what they felt the purpose of the experiment was. The concluded the experiment.

Data Preparation and Statistical Analyses

In order to be included in the study, each participant must have indicated on their manipulation check questionnaire that they tried at least moderately hard to follow their group instructions. Moderately hard was defined as a rating of 3 or above on a 5-point Likert-type scale, where a rating of 1 indicated not trying at all and a 5 indicated trying very hard. Once all cells had been filled by subjects whose rated effort met this criterion, persons whose average response latencies were greater than 3 standard deviations from their group mean were discarded to avoid biasing results due to outliers. Data from three participants were discarded on this basis. One subject was from the UM group and 2 were from the CHI group. The reasons for these extreme scores were due to a failure to understand the task (e.g., one subject stared at the screen for almost 2 minutes on one item before asking for clarification of instructions, thereby raising his average latency). Thus, for all analyses there were 29 UM subjects, 30 CM subject, 28 CHI subjects, and 30 NC subjects, for a total sample size of 117.

Discriminant function analysis was used to assess jointly the classification accuracy of the PDRT-C total correct and response latency scores. A second stepwise function was also calculated to identify the efficacy of including all measures of suboptimal effort in classification. In calculating the discriminant functions, data from the coached and uncoached malingerers were combined into a single group of malingerers because the question under investigation when using these tests is whether a patient is malingering or truly impaired, and not whether the malingerer was prepared in advance. In addition, none of the data from subjects in the NC group were used in calculating the discriminant function for the same reason. That is, differentiating malingerers from nonimpaired individuals is not a question typically encountered in clinical practice. Including the normal control data would likely result in artificially inflated classification rates. However, after all discriminant functions were calculated, NC subjects were classified on the basis of the obtained functions as a further check for specificity. Subject classification (i.e., malingering vs. not malingering) using the discriminant function scores was calculated after adjusting for prior probabilities. Although the actual base rate of malingering is unknown, we chose a 30% base rate because approximately 27% of Binder’s minor head trauma patients who were seeking compensation scored below the cut-off for malingering on the PDRT (Binder, 1993; Binder & Willis, 1991). Also, when calculating the PDRT-C’s discriminant function, one third of the subjects from the UM, CM, and CHI groups were withheld from the sample and used for cross validation. This was possible because the remaining subject-to-variable ratio was still within the conventional 10:1 limit.

RESULTS

Manipulation Check

Participants from all groups were able to recall their instructions with an appropriate degree of accuracy. The effort ratings were analyzed by one-way analysis of variance (ANOVA). Due to the large number of comparisons in the present study, the alpha level to determine statistical significance was set at .004 using a modified Bonferroni procedure.
Comparison of Malingering Tests

TABLE 2
Malingering Data for All Tests

<table>
<thead>
<tr>
<th></th>
<th>Uncoached M (SD)</th>
<th>Coached M (SD)</th>
<th>Injured M (SD)</th>
<th>Normal Controls M (SD)</th>
<th>F (3, 113)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDRT-C Total Correct</td>
<td>31.41*</td>
<td>40.87*</td>
<td>61.04*</td>
<td>69.60*</td>
<td>69.16*</td>
</tr>
<tr>
<td></td>
<td>(12.55)</td>
<td>(15.65)</td>
<td>(10.93)</td>
<td>(2.51)</td>
<td></td>
</tr>
<tr>
<td>PDRT-C Response Latency (Expressed in Seconds)</td>
<td>2.91*</td>
<td>3.21*</td>
<td>4.24*</td>
<td>1.69*</td>
<td>13.08*</td>
</tr>
<tr>
<td></td>
<td>(0.83)</td>
<td>(1.12)</td>
<td>(2.82)</td>
<td>(0.43)</td>
<td></td>
</tr>
<tr>
<td>Nonverbal Forced Choice Test Total Correct</td>
<td>57.17*</td>
<td>69.70*</td>
<td>71.64*</td>
<td>85.93*</td>
<td>25.79*</td>
</tr>
<tr>
<td></td>
<td>(17.07)</td>
<td>(12.74)</td>
<td>(11.13)</td>
<td>(7.77)</td>
<td></td>
</tr>
<tr>
<td>Nonverbal Forced Choice Test Slope</td>
<td>-.004</td>
<td>-.006</td>
<td>-.007</td>
<td>-.005</td>
<td>4.33</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>Nonverbal Forced Choice Test Consistency Ratio</td>
<td>0.612*</td>
<td>0.748*</td>
<td>0.762*</td>
<td>0.89*</td>
<td>22.62*</td>
</tr>
<tr>
<td></td>
<td>(0.181)</td>
<td>(0.135)</td>
<td>(0.120)</td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>21-Item Test Forced Choice Recognition</td>
<td>14.00*</td>
<td>15.77*</td>
<td>15.38*</td>
<td>18.90*</td>
<td>15.99*</td>
</tr>
<tr>
<td></td>
<td>(3.45)</td>
<td>(3.01)</td>
<td>(2.83)</td>
<td>(1.71)</td>
<td></td>
</tr>
<tr>
<td>Dot Counting Test Ungrouped Correct</td>
<td>3.41</td>
<td>3.67</td>
<td>3.32</td>
<td>4.57</td>
<td>4.19</td>
</tr>
<tr>
<td></td>
<td>(1.50)</td>
<td>(1.52)</td>
<td>(1.70)</td>
<td>(1.31)</td>
<td></td>
</tr>
<tr>
<td>Dot Counting Test Grouped Correct</td>
<td>4.93</td>
<td>5.17</td>
<td>5.36</td>
<td>5.57</td>
<td>1.79</td>
</tr>
<tr>
<td></td>
<td>(1.22)</td>
<td>(1.26)</td>
<td>(1.10)</td>
<td>(0.73)</td>
<td></td>
</tr>
<tr>
<td>Dot Counting Test Ungrouped Total Time</td>
<td>46.16</td>
<td>41.76</td>
<td>45.94</td>
<td>31.99</td>
<td>4.70</td>
</tr>
<tr>
<td></td>
<td>(19.32)</td>
<td>(15.26)</td>
<td>(21.44)</td>
<td>(7.33)</td>
<td></td>
</tr>
<tr>
<td>Dot Counting Test Grouped Total Time</td>
<td>31.93</td>
<td>26.30</td>
<td>26.01</td>
<td>14.97</td>
<td>5.75*</td>
</tr>
<tr>
<td></td>
<td>(17.06)</td>
<td>(13.46)</td>
<td>(23.91)</td>
<td>(4.29)</td>
<td></td>
</tr>
<tr>
<td>Dot Counting Test Ungrouped Trend Reversals</td>
<td>1.07</td>
<td>1.23</td>
<td>1.46</td>
<td>1.00</td>
<td>0.622</td>
</tr>
<tr>
<td></td>
<td>(1.25)</td>
<td>(1.55)</td>
<td>(1.55)</td>
<td>(1.23)</td>
<td></td>
</tr>
<tr>
<td>Dot Counting Test Grouped Trend Reversals</td>
<td>2.10*</td>
<td>2.03*</td>
<td>2.93*</td>
<td>2.10*</td>
<td>6.32*</td>
</tr>
<tr>
<td></td>
<td>(1.26)</td>
<td>(0.928)</td>
<td>(0.851)</td>
<td>(0.31)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Row means with different superscripts were significantly different, p < .05.
*One-way ANOVAs comparing all four groups, p < .004.

(.05/12). Post hoc analyses were calculated using the Newman-Keuls procedure. The ANOVA yielded a significant main effect for group, F(3, 113) = 5.40, p < .004. CM subjects indicated putting forth less effort than the CHI and NC groups, while all other group differences were nonsignificant (M = 4.02, 4.32, 4.47, and 4.62 for the CM, UM, NC, and CHI groups, respectively). Although the difference between groups was statistically significant in this single case, the magnitude of group differences were small and all subjects indicated trying at least moderately hard to follow their instructions.

Group Differences

Group differences were analyzed using separate one-way ANOVAs. The means for these comparisons are presented in Table 2. Group differences were identified for the PDRT-C total correct score, F(3, 113) = 69.16, p < .0001. As can be seen in Table 2, significant differences between all groups emerged, including the UM and CM subjects. Group differences in PDRT-C response latency were also significant, F(3, 113) = 13.08, p < .0001. In contrast to the total correct data, however, the response latencies of the UM and CM groups did not differ, though both groups’ response latencies were significantly shorter than the CHI groups’ and longer than the NC groups’.

Analysis of the Nonverbal Forced Choice Test data revealed that the differences between groups for the slope were nonsignificant, F(3, 113) = 4.33, p > .004. However, differences between the four groups in their total scores and consistency ratio were significant, F(3, 113)
FIGURE 1. Classification accuracy of each malingering test (note that the uncoached and coached malingering group data were collapsed).

= 25.79 and $F(3, 113) = 22.62$, respectively, $p < .0001$. With respect to the total correct score, UM subjects answered significantly fewer items correctly than both the CM group and the CHI patients. These two groups, who did not differ in their number of correct responses, answered significantly fewer items correctly than the NC subjects. Analysis of the consistency ratio scores identified identical results. Specifically, UM subjects were less consistent in their responses to items of identical difficulty than were the CM and CHI groups. The consistency ratios of these latter two groups did not differ from each other but both were significantly lower than those of the NC group.

Group differences on the forced-choice portion of the 21-Item Test were statistically significant, $F(3, 113) = 15.99$, $p < .0001$. NC subjects answered significantly more items correctly than any of the other three groups. In addition, the CM group answered significantly more items correctly than did the UM group. Interestingly, neither of the two malingering groups’ scores differed significantly from the CHI group’s score. As can be seen in Table 2, the mean from the CHI group fell between those of the UM and CM groups.

Of the six ANOVAs from the Dot Counting Test (total correct: grouped, total correct: ungrouped, total time: grouped, total time: ungrouped, trend reversals: grouped, trend reversals: ungrouped), only two achieved statistical significance. Patients with head injuries made significantly more reversals in trend on the grouped dots condition than all other groups, $F(3, 113) = 6.32$, $p < .004$, and the time taken to count grouped dots was shorter for NC subjects than all other groups, $F(3, 113) = 5.75$, $p < .004$.

**Classification Rates**

*Malingering versus honest responding.* To assess further the utility of the measures, classification rates were calculated for each test using published decision rules. The classification accuracy of each test is shown in Figure 1. As can be seen, all tests were able to classify correctly the NC subjects as not malingering (though the Nonverbal Forced Choice Test had one false positive). Using just the total correct cut-off score on the PDRT-C, 39/59 (66%)
of the malingering and 25/28 (89%) CHI subjects were correctly classified, for a total hit rate of 74%. Using Frederick and Foster’s (1991) decision rule for malingering on the Nonverbal Forced Choice Test (a score below average and slope × consistency ratio > −.0041) and a cutoff score of 14 correct on the 21-Item Test (indicating performance at or below chance; Iverson et al., 1991), both tests showed only mild sensitivity and low specificity, such that a number of subjects in the CHI group were incorrectly classified as malingering on both tests. On the Nonverbal Forced Choice Test 20/59 malingering and 21/28 CHI subjects were correctly classified, for an overall hit rate of 47%, which is no better than chance. One of the NC subjects was also misclassified as providing insufficient effort on this test as well. On the 21-Item Test 25/59 malingering and 19/28 CHI subjects were correctly classified, for an overall hit rate of 51%. This hit rate was again no better than chance. The classification accuracy of the Dot Counting test was assessed using the cutoff scores identified by Paul, Franzen, Cohen, and Fremouw (1992). Using these scores, 5/59 (9%) malingering and 27/28 (96%) CHI subjects were correctly classified for an overall hit rate of 37%. This level of accuracy indicates adequate specificity but, again, little sensitivity.

**Discriminant function analysis.** The benefit of adding the response latency measure to the PDRT-C was assessed by discriminant function analysis with UM and CM groups combined into a single malingering group as described above. Because classification rates obtained from the initial discriminant function calculation tend to be spuriously high, the PDRT-C classification data presented here are from the cross validation sample.

The total classification rate using the PDRT-C discriminant function was significant, χ²(2, N = 60) = 50.96, p < .0001, with an overall correct classification rate of 87%. As expected, the overall classification rate dropped slightly to 81% on cross validation. The PDRT-C still demonstrated adequate sensitivity and excellent specificity as 14/19 (74%) of the malingers and 8/8 (100%) of the CHI subjects were correctly classified. For comparison, the chance hit rate was 58%, calculated using the proportional chance criterion described by Huberty (1984; total chance hit rate = sum of each group n squared, divided by the total N²). The improved specificity following the addition of the response latency resulted in better overall classification rates when compared to all other tests included in this study (see Figure 1).

We next sought to assess whether the comparison of the PDRT-C discriminant function classification rates to the ‘cut-off’ classification rates of the remaining tests may have biased results due to the mathematically optimal effects of discriminant function. To do this, a stepwise discriminant function analysis was employed using the scores from all tests. Such an analysis would also assess the cumulative ability of all tests to detect suboptimal performance. Once again, the CM and UM groups were combined into one group, and data from the NC group were not used in the analysis. The resulting function was significant, χ²(5, N = 87) = 81.78, p < .0001. Both the total number correct and response latency scores from the PDRT-C were retained in the discriminant function. Interestingly, the Primacy, Recency, and Middle scores from the 21-Item Test were also included (all possible malingering indicators were included as potential predictors for the stepwise discriminant function). No other scores provided enough group separation to be included in the final function. In using the discriminant function for classification, 48/59 (81%) of the malingers and 28/28 (100%) of the CHI group were correctly classified, for a total correct classification rate of 87%. This rate, though not cross validated on an independent sample, was nearly identical to the function obtained using just the PDRT-C total number correct and response latency scores. Interestingly, when the NC group was classified by the obtained function, seven subjects were misclassified as malingering.
Coached versus uncoached malingerers. The accuracy of each test in identifying coached and uncoached malingerers is demonstrated in Figure 2. On average, 47% of the UM group was correctly identified while only 29% of the CM group was identified as malinger, highlighting the tendency for coached subjects to be better at avoiding detection. The PDRT-C discriminant function demonstrated the best overall level of classification by identifying 7/9 (78%) of the UM group and 7/10 (70%) of the CM subjects. It was the only test to identify more than 50% of the coached subjects. Using just the PDRT-C total number correct, 14/30 CM subjects were correctly identified as providing insufficient effort. Of the remaining tests, the Nonverbal Forced Choice Test correctly identified 10/30 (33%), the 21-Item Test 8/30 (27%), and the Dot Counting Test 3/30 (10%) of the CM group, again demonstrating less than acceptable sensitivity to advanced preparation in our sample.

To summarize the effects of the coaching manipulation, analyses revealed that coached malingerers’ scores on four indices (PDRT-C total correct, the Nonverbal Forced Choice Test total correct, the Nonverbal Forced Choice Test consistency ratio, and the 21-Item Test forced-choice recognition score) were significantly different from the uncoached malingerers’. Scores of the CM group in all cases of significance indicated less impairment than the UM group. No other differences between CM and UM groups were identified, including on the response latency measure of the PDRT-C. In contrast, group differences between the UM and CHI subjects were identified on five indices: PDRT-C total correct, PDRT-C response latency, the Nonverbal Forced Choice Test total correct, the Nonverbal Forced Choice Test consistency ratio, and the 21-Item Test forced-choice recognition score. Scores of CM subjects, however, were statistically different from the scores of CHI subjects on only two indices: PDRT-C total correct and PDRT-C response latency. Note that although the PDRT-C total correct score for both malingering groups was significantly below that of the CHI group’s, only the UM group’s mean score was below the malingering cut-off of 39 correct established by Binder and Willis (1991). The only other malingering index that differentiated CM from CHI subjects was the PDRT-C response latency measure, which did not differentiate UM and CM subjects.
DISCUSSION

This study compared the ability of four measures of suboptimal performance to correctly classify informed and uninformed malingering subjects, persons with documented head injuries, and normal control subjects. An important finding was that only the PDRT-C total correct score resulted in better than chance classification of simulated malingerers on direct comparison within a single study. False positive errors, an important consideration in the use of these tests, occurred with all tests when traditional cutoff scores were used for classification. However, no false positive errors occurred when the PDRT-C total correct and response latency discriminant function was used for classification. It is also interesting to note that when all potential predictors from this study were entered into a stepwise discriminant function analysis, only the PDRT-C scores (total correct and response latency) and the 21-Item Test serial position scores were retained as predictors. The resulting function did not classify subjects to any greater degree than the PDRT-C function alone, and in fact, misclassified 7 of 30 normal control subjects as possible malingerers. A second important result of this study was that coached subjects, while as a group were better able to escape detection than uncoached subjects, were more often correctly identified as malingering using the PDRT-C discriminant function than by using scores from any other test included in the study. Since there is as yet no empirical data on the number of malingerers who prepare in advance versus those who do not, neuropsychology would be best served by the development of tests sensitive to both groups of patients.

Small differences in PDRT-C total correct scores between our sample and those of Binder and Willis (1991) emerged. Our NC and CHI subjects scored slightly higher than the average scores of Binder and Willis’ Nonpatient-No Compensation and Brain Damaged-No Compensation groups, respectively, while the scores of Binder and Willis’ simulated malingerers separated the scores of our UM and CM groups. Because these differences were small and the range of scores overlapped to some degree, this observation may represent normal variation. It is unlikely that these differences reflect differences between the PDRT and the PDRT-C, since previous work (Rose, Petree, & Hall, 1993) found no differences between the tests using simulated malingering and normal control subjects. The cross-study differences between malingering groups, however, may reflect design variables such as instruction set and the use of incentives (our study used incentives for malingering while Binder and Willis’ did not report using such incentives). Again, although the differences were small, more detailed investigation of the effects of incentives on simulated malingering and their relation to the incentives found in actual practice (e.g., large financial gain) would be helpful.

The Nonverbal Forced Choice Test has been found to be an effective aid in the detection of biased responding in other studies (Frederick and Foster, 1991; Frederick et al. 1994), and the classification rate observed in this study was nearly identical to that reported by Frederick et al. (1994). The PDRT-C, however, was slightly better in terms of overall detection rate, when either the total correct cut-off score or the discriminant function was used for classification. Although both the PDRT-C and the Nonverbal Forced Choice Test are based upon the Symptom Validity Testing procedure (Pankratz et al., 1975), there is an important difference between the two measures. The PDRT-C is a two-alternative forced-choice task designed to appear like a memory test, whereas the Nonverbal Forced Choice Test utilizes a nonverbal problem-solving format. Given that the most common complaint found in head injured patients is mnestic in nature (Brandt, 1988), one hypothesis is that subjects instructed to malinger performed worse on the “memory” tests than the problem-solving tests, perhaps by inferring that such a pattern of performance would best achieve their goal. Indeed, the uncoached subjects in this study scored, on average, significantly lower
than both the coached and head-injured groups on the Nonverbal Forced Choice Test, while the coached and head injured subject’s scores were not statistically different from each other. The uncoached group’s strategy may simply have been to do poorly on all tests, whereas the coached group was more selective in their performance. Further study examining the types of deficits typically feigned and the best test to detect such dissimulation may help clarify this issue.

Why, though, did the other tests demonstrate such limited sensitivity in identifying suboptimal performance? Although some studies have found the 21-Item Test to be successful in the detection of simulated memory impairment (e.g., Iverson and Franzen, 1996; Iverson, Franzen, & McCracken, 1994; Iverson et al., 1991), others (Berry et al., 1997; Frederick et al., 1994) found the 21-Item Test to have excellent specificity but poor sensitivity. In the Frederick et al. (1994) study, nearly 100% of their honest responders were correctly classified, but only 40% of their uncoached and 22% of their coached simulated malingerers were correctly identified as malingering. Berry and colleagues (1997) also correctly classified 100% of their honest responders but only 20% of their combined coached and uncoached malinger- ing groups. The findings from the present experiment, therefore, replicate those of other studies that have found weak support for the 21-Item Test in identifying feigned cognitive impairment. The robustness of this finding is evident in the administration differences between studies: Frederick and colleagues used a group-administration format, whereas all subjects participated individually in Berry et al. (1997) and in the present study.

Order effects may have contributed to the present findings and should not be overlooked. As Bernard (1990) observed, administering more difficult tests early on in the testing session may serve to sensitize malingerers to the simplicity of malingering tests administered later in the session. Since the Dot Counting procedure and the 21-Item Test were among the last tests administered in the battery, our simulated malingerers may have limited the amount of “impairment” they produced on these measures due to fatigue or other factors. However, Frederick et al. (1994) administered the 21-Item Test before the Nonverbal Forced Choice Test and Dot Counting procedure but after “true” memory tests, such as the Rey Auditory Verbal Learning Test and found a pattern similar to those observed in the present study. If order effects, including the reduced sensitivity of measures administered after difficult tests, were robust and significant, the Nonverbal Forced Choice Test would likely not have been as successful as Frederick et al. reported. In fact, the Nonverbal Forced Choice Test demonstrated a slightly better ability to detect biased responding in that study than in the present. Moreover, Hall and Parker (1996) found no significant difference in the scores of simulated malingerers on various tests of effort, including the PDRT-C, when the tests were administered either before or after such difficult tests as the Wisconsin Card Sort Test and the Booklet Category Test. As a result, it is suggested that the 21-Item Test and the Dot Counting procedure be interpreted with caution when used to assist in the detection of malingering.

The coaching manipulation resulted in some interesting differences between the uncoached and coached groups. Rose et al. (1994) reported that the average total number correct of the uncoached group was below the cutoff score suggestive of malingering on the PDRT-C. The coached group’s average score, however, was above the cut-off. Using just the cut-off scores, then, Rose et al. (1994) found that only 47% of the coached malingerers were correctly identified, compared to 86% of the uncoached group. In the present study, the Nonverbal Forced Choice Test, 21-Item Test, and Dot Counting Test and similar discrepancies between uncoached and coached groups. Only the PDRT-C, utilizing the total correct score along with the response latency measure resulted in identifying more than 50% of the coached malingering subjects. In fact, the combined PDRT-C method increased the sensitivity of the
measure such that the detection of coached malingerers went up from 47% using the total correct score alone to 70% without increasing the number of false positive identifications.

The results of this study re-emphasize the importance of independent validation of results before new tests can be employed in a clinical setting with confidence. Iverson et al. (1991) found that a full 88% of their sample were correctly identified using the 21-Item Test, while the present study did not do any better than chance at 51%. Similarly, the Dot Counting procedure, found by Paul et al. (1992) to correctly identify 63% of their malingerers, was only able to identify 8% in the present study. The discriminating factor appears to be that the decision rules obtained in the studies by Iverson et al. (1991) and Paul et al. (1992) were developed on their own sample of subjects. As with discriminant function analysis, such methods tend to capitalize on sample-specific variance. The present study attempted to address this problem by validating the discriminant function obtained from the PDRT-C data on a separate, albeit small, sample.

It should be emphasized that only unimpaired college students simulating the effects of head injury were used as our malingering sample, limiting the generalizability. Binder and Willis (1991) looked at the PDRT performance of both nonimpaired subjects instructed to malingering and a sample of mild head trauma patients seeking financial compensation for their injuries. Those researchers found a significant difference between the two groups, with simulators performing more poorly. Although the scores of 26% of Binder and Willis’ minor head trauma group fell below the malingering cutoff, they did not report the extent to which those scores were similar to the simulators. Nonetheless, the degree to which the PDRT-C discriminant function is generalizable to actual malingerers is unknown and needs further investigation.

Also of note is that our patient group had sustained head injuries which were more severe in nature. Because the response latencies of severely head-injured patients tend to be longer than those of less severely injured subjects (Stuss et al. 1989; Van-Zomeren & Deelman, 1978), the discriminant function will need to be validated (or, more precisely, recalculated) using a sample of patients with mild head trauma who are and are not seeking compensation. If the response latency measure is still able to improve the discriminability of the PDRT, only then can it be considered appropriate for clinical use.

The investigation of measures designed to assist in the detection of malingering in neuropsychological assessment is an important area of research. The results of the present study suggest that several tests commonly used to detect malingerers may lack the sensitivity and/or specificity to be clinically useful. At the very least, continued research employing persons with mild head injuries, large sample sizes, and cross validation of results is needed on all of these measures to ensure both reliability and validity. Although only an experimental technique at the present time, the discriminant function using the response latency and total correct score from the PDRT-C resulted in classification rates for both coached and uncoached subjects that were acceptable and substantially better than the three other measures studied. These findings hold promise that with further research neuropsychologists will be able to develop tests that can accurately separate those who are trying to take advantage of the system (i.e., malingerers) from those who are truly in need of ever-dwindling resources.

REFERENCES


Rose, F. E., Petree, A. D., & Hall, S. (1993). The detection of malingering using a computerized version of the...


