Psychological Assessment:
A Journal of Consulting and Clinical Psychology
1991, Vol. 3, No. 4, 596–602

1040-3590/91/$3.00

Multiple Measures of Malingering on a Forced-Choice Test of Cognitive Ability

Richard I. Frederick and Hilliard G. Foster, Jr.
Whiting Forensic Institute
Middletown, Connecticut

This article describes how a 2-alternative, forced-choice response technique was applied to a nonverbal test of cognition in order to generate measures of noncompliance and serve as a means to detect malingering. Three studies were conducted. Study 1 used a 2-alternative, forced-choice test format comprising 100 nonverbal test items of a hierarchy of difficulty presented in a random order. In a simulation study, a combination of decision rules regarding (a) performance curves slope; (b) a measure of response consistency ("consistency ratio"); and (c) the product of the slope and consistency ratio resulted in high discriminability among the test results of normal and cognitively impaired controls and simulating malingerers. A replication, Study 2, yielded similar findings. Study 3 demonstrated the robustness of these decision rules. Most subjects (73%) who simulated malingering after receiving accurate information on how to avoid detection were still classified as malingerers.

Psychologists currently have no established objective methods of detecting malingering on the most commonly used instruments of cognitive, intellectual, or neuropsychological ability. The expanded use of such tests in both criminal and civil forensic evaluations increases the likelihood of impaired performances for secondary gain and necessitates the development of indicators of response validity (Bash & Alpert, 1980; Hiscock & Hiscock, 1989). This article describes how a test of cognitive ability was altered so that it could yield quantitative measures of noncompliance.

Studies of clinical detection of malingering on cognitive tests are not encouraging. Patients are able to produce realistic findings of impairment on neuropsychological tests (Mensch & Woods, 1986), whereas clinicians are generally unable to detect malingered performance at better than chance levels (Faust, Hart, & Guilmette, 1988; Heaton, Smith, Lehman, & Vogt, 1978), even when forewarned that examinees might be faking impairment (Faust, Hart, Guilmette, & Arkes, 1988). Some studies (e.g., Goebel, 1983; Heaton, Smith, et al., 1978) that are often cited (as in Schretlen, 1988) to demonstrate accuracy in clinical detection are compromised by serious methodological flaws (Faust et al., 1988a, 1988b).

A more promising approach has been to objectively analyze response patterns (Goldstein, 1945; Hunt, 1946; Hunt & Older, 1943; Schretlen & Arkowitz, 1990). But, despite the claim by Schretlen (1988) that ". . . examination of the response 'scatter' appears to be the most powerful and well validated detection strategy [for intelligence tests]" (p. 458), the extent of even normal variation in subtest scatter is substantial (Atkinson, 1991; Matarazzo, 1990; Matarazzo & Prifitera, 1989). This, combined with the lack of formal decision rules, makes the current use of response scatter to detect malingering appear unjustified. Using response pattern analysis to detect malingering on most cognitive tests is also hampered by the format of these tests, which were designed with the assumption that subjects would be giving their best effort (Rogers, 1984). They mostly contain tasks that are open-ended. That is, subjects who wish to appear impaired must not give *the* wrong answer, but simply *a* wrong answer. This impedes the formation of a priori hypotheses about recognizable patterns of deception.

In contrast, a forced-response format facilitates the development of several strategies for detecting noncompliance through the analysis of response patterns. For example, on the basis of clinical reports of the investigation of suspicious reports of blindness (Brady & Lind, 1961; Grosz & Zimmerman, 1965; Theodor & Mandelcorn, 1973), Pankratz, Fausti, and Peed (1975) used a two-alternative, forced-choice method to investigate suspicious deafness. They presented 100 tonal-discrimination tasks of equivalent difficulty and concluded that malingering was present when significant variations from chance occurred (i.e., at the .05 level, less than 42 correct discriminations). They argued that the technique has value for the assessment of any sensory loss. In fact, this technique, now termed *symptom validity testing* (SVT; Pankratz, 1979), has been effective in demonstrating true ability in patients who presented with such various symptoms as deafness, anesthesia, and memory deficits (Bickart, Meyer, & Connell, 1990; Binder, 1990; Binder & Pankratz, 1987; Haughton, Lewsley, Wilson, & Williams, 1979; Hiscock & Hiscock, 1989; Pankratz, 1979, 1983).

We thank Brad Waite and Michael Silverstein for facilitating access to research subjects and Marc Hillbrand and William Merrick for reviewing a draft of this article.

Correspondence concerning this article should be addressed to Richard Frederick, Whiting Forensic Institute, Box 70, O'Brien Drive, Middletown, Connecticut 06457.

A Forced-Choice Cognitive Test

We proposed a test of cognitive ability in which each test item required a forced choice between two alternatives. Unlike SVT (which uses stimuli requiring constant levels of ability), we proposed a test with two forms (A and B) of 50 items each, with a parallel hierarchy of item difficulty across forms. From this, 50 "equivalent item pairs" could be generated, comprising two items (one from each form) that were considered equal to each other in difficulty. Ideally, each pair could be assumed to differ in level of difficulty from any other equivalent item pair.

By designing the test in this fashion, it was possible to develop several potential measures of noncompliance, as well as to make predictions about their values on the basis of response style:

1. *Score.* The first prediction was the malingerers (or persons simulating a malingering response style) would be identified by scores of less than 42%, the lower limit of the range of random responding at $p = .05$ for 100 items.

2. *Slope.* The second prediction was that visual plots of performance for simulating malingerers (SM) would differ from the performance curves of both normal control subjects (NC) and cognitively impaired control subjects (CI). Although SM subjects were expected to obtain total scores below 42%, they were expected to perform at or about 50% once they reached their ceiling of true ability. To obtain an overall score below chance, they would have to perform substantially below chance prior to their ceiling. As a consequence, a performance curve describing their responses (arranged from the easiest to the most difficult item) would be positively sloped. On the other hand, NC and CI individuals were predicted to demonstrate a negatively sloped performance curve, as they were expected to perform near 100% early in the test and near 50% as they reached their ceiling (see Figure 1).

3. *Consistency ratio.* The third prediction was that malingerers would perform unreliably: An analysis of equivalent item pairs would reveal a lack of consistent responding. Typical consistency measures for dichotomous variables, such as phi or kappa, were not used because they would be unduly affected by the random responding that occurs when the ceiling is reached. Such measures of response consistency would tend to be spuriously low for our purposes, especially for compliant individuals of low ability. Therefore, to measure consistency, we used the ratio of the number of equivalent item pairs in which both items were answered correctly to the maximum *possible* number of item pairs in which both items could be answered correctly (which is the total number of correct items divided by 2). If a is considered any correct response, and b represents an equivalent item pair in which both items are answered correctly, the equation to compute the consistency ratio (CR) appears as follows:

$$CR = \sum(b)/(\sum(a)/2).$$

In this manner, a test score of 50 can generate a ratio from 0.0 to 1.0, depending on the response style (see Table 1). We hypothesized that most malingerers would obtain a consistency ratio at or below the mean for randomly produced responses. For scores from 42 to 50 (the lower range of random responding), the mean consistency ratio is 0.50. Thus, we chose that value as a cutoff point for malingering. Scores below 50 (except 0 and 1) have a range of value from 0.0 to 1.0, but as scores exceed 50 and approach 100, the minimum possible values for CR rapidly increase. Consequently, the minimum CR value for a score of 67 is 0.50. So, a cutoff value of 0.50, like the 42% for score, compromises sensitivity for malingerers who fabricate only mild impairment. Nonetheless, it was considered potentially useful to discriminate malingerers from compliers among all those whose scores were below 67.

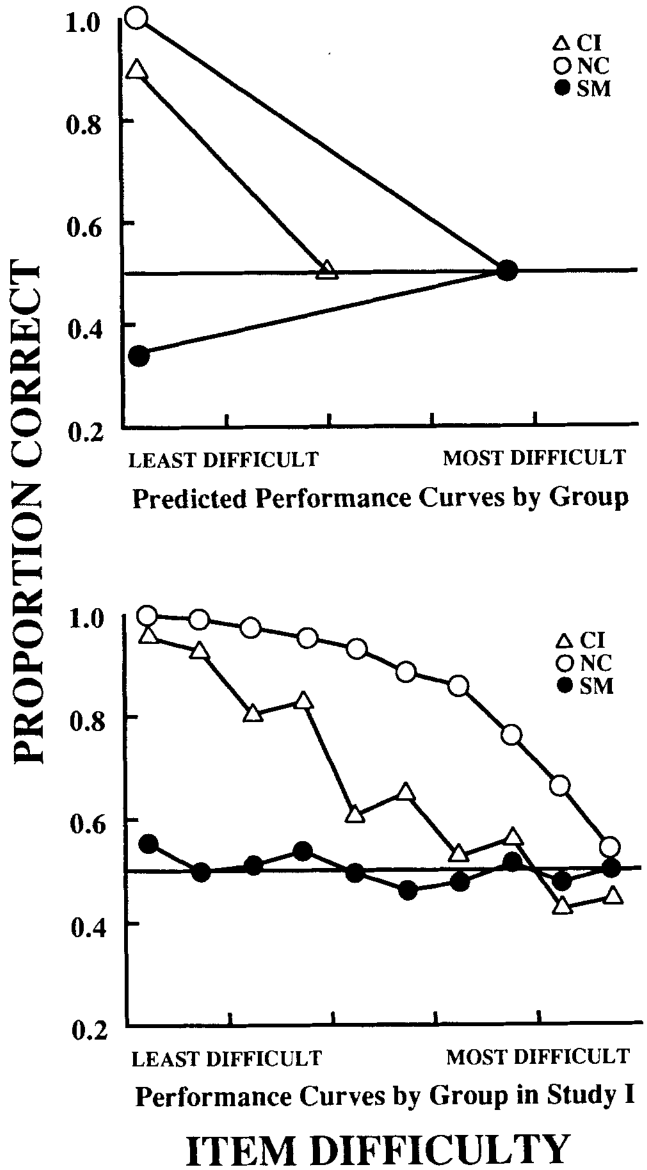

Figure 1. Predicted and actual performance curves for normal controls (NC), cognitively impaired forensic patients (CI), and simulating malingerers (SM) in Study 1.

Study 1. Developing the Instrument and Testing Hypotheses

Method

Making a two-alternative forced-choice cognitive test. The Test of Nonverbal Intelligence (TONI; Brown, Sherbenou, & Johnsen, 1982) is

Table 1
Examples of Response Pattern Extremes and Consistency Ratio Computation for Scores of 50

Equivalent item pair	Inconsistent Form A	Inconsistent Form B	Random Form A	Random Form B	Consistent Form A	Consistent Form B
1	0	1	0	0	1	1
2	1	0	0	1	1	1
3	0	1	1	0	1	1
4	1	0	1	1	1	1
—						
47	0	1	0	0	0	0
48	1	0	0	1	0	0
49	0	1	1	0	0	0
50	1	0	1	1	0	0
Column totals	25	25	25	25	25	25
Test score	50		50		50	
Item pairs[a]	0		12[b]		25	
Test score/2	25		25		25	
Consistency ratio	0.0		0.5		1.0	

Note. 0 = incorrect answer, 1 = correct answer.
[a] Number of equivalent item pairs in which both items are answered correctly. [b] Rounded from 12.5.

a language-free measure of cognitive ability that requires examinees to solve problems by identifying relationships among abstract figures. The TONI comprises two equivalent forms, each containing 50 items. The two forms, A and B, have been demonstrated to be approximately equal in difficulty (Brown et al., 1982), but because of our modification to the standard presentation, equivalent item pairs were generated by a post hoc item reordering of test items on the basis of the percentage answered correctly by NC and CI subjects. Although reordering item difficulty on an individual basis would certainly elevate individual CRs, a reordering on the basis of the collective responses of the control groups was not expected to unduly affect group means for CR.

Each TONI test item presents an incomplete picture puzzle, with four or six possible completing choices underneath. For this study, the choices were restricted to two possibilities: the correct answer and the item judged to be the most effective distractor. Distractor effectiveness was defined by the results of a pilot study in which five clinical psychologists (including both of us) served as judges. The alternative that the majority chose as the most effective distractor was retained. When a tie occurred between two alternatives, one of these was chosen randomly. The level of interrater agreement was not significant ($\kappa = 0.13$, $z = 0.25$, $p > .10$). If raters made reasonable judgments, this suggests that the TONI contains several effective distractors for most test items and that the effect of choosing a different alternative would have been minimal.

The alteration of test presentation was completed by a modified random arrangement of the test items so that order by item difficulty was random, but items from Form A consistently alternated with items from Form B.

In calculating the slope, the items are first reordered by difficulty, as opposed to their random presentation during testing. Once reordered, the easiest Form A item is followed by the easiest Form B item, which is followed by the second easiest Form A item. For each subject, numeric responses (0 = *incorrect*, 1 = *correct*) are also reordered to match corresponding items. Then, running means are computed (we used the SERIES module of SYSTAT; Wilkinson, 1981) for each 10 consecutive test responses (i.e., for Items 1–10, 2–11, . . . 91–100). *Slope* is then the slope of the line generated by the 91 consecutive means for each subject. We used running means, as opposed to nonoverlapping means, to reduce the potential effects produced by difference in the "true" order of item difficulty and our arrangement.

Subjects. Because of the scarcity of known malingerers, a simulation study was planned. One hundred seventy-one college students (128 female, 43 male; 154 White, 12 Black, 4 Hispanic, 1 Asian) made up the NC group ($n = 86$) and the SM group ($n = 84$). One college student was excluded for failure to comply with instructions. All received only extra credit for their participation.

The CI group comprised individuals who were predicted to perform at a below-average level on the experimental test. Using the criterion for brain damage used by Robinson, Heaton, Lehman, and Stilson (1980; cf. Heaton, 1981; Heaton, Baade, & Johnson, 1978), subjects were selected for inclusion in the CI group if they produced more than 19 perseverative responses on the Wisconsin Card Sorting Test (WCST; Milner, 1963). Fourteen male forensic psychiatric patients (mean age = 31.5 years, $SD = 8.7$, range = 23–60 years; 9 White, 5 Black) were selected for participation (mean perseverative responses = 57.8, $SD = 31.5$, range = 21–124). None were pending adjudication, all were on treatment units, and all but 1 had been administered the WCST in the past year for purposes other than this study. They did not receive any remuneration for their effort.

Procedure. NC and CI subjects completed the test after being given instructions to do their very best. The SM students were instructed to "fake brain damage," pretending that successfully doing so would result in a large court settlement. They were warned not to be obvious, as doing so "could result in a loss of settlement and severe court penalty." They were not given any advice on how to appear brain damaged.

Predictions about test score, slope, and consistency ratio were implemented as decision rules (see Table 3). Individuals who received variant values were classified as malingerers.

Results and Discussion

Figure 1 shows the performance curves for each group. Table 2 presents means for each group. CI subjects scored significantly below the group average for NCs, $t(98) = -7.33$, $p < .001$. All CIs scored below the group mean for NCs, which suggests that CIs constitute a reasonable comparison group for the SM subjects. Table 3 shows results of decision rule application. Fifty-one simulators were identified by the decision rules regarding the slope and the consistency ratio. Only 14 of the 84 simulators scored below 42%, and 13 of those were included in the group of 51 just mentioned. It should be noted that consistency ratios for NC and CI subjects may have been elevated by the post hoc reordering of item difficulty.

Post hoc discriminant function analysis revealed that the interaction between slope and consistency ratio was also an effective discriminator, $F(2,168) = 32.31$, $p < .001$. Therefore, we created a decision rule based on the products of the slope and CRs for malingerers. To develop the most sensitive decision rule, products for obvious malingerers (those with positively sloped performance curves) were excluded from the calculation. We used only the product values for the more sophisticated malingerers (those with a negatively sloped performance curve; $n = 49$, mean product = $-.0015$, $SD = .0013$). The criterion was established as the product value two standard deviations below the mean product for sophisticated malingerers (criterion = $-.0041$). The new decision rule states that a responder who obtained a more positive product value than the criterion would be classified as a malingerer.

Table 2
Group Means and Standard Deviations by Study

Group	*n*	Score *M*	Score *SD*	Slope *M*	Slope *SD*	CR *M*	CR *SD*	Slope*CR *M*	Slope*CR *SD*
Study 1									
Normal control	86	85.2	7.9	−.005	.002	.884	.006	−.004	.001
Cognitively impaired	14	67.6	10.9	−.008	.002	.738	.10	−.006	.001
Simulating malingerer	84	49.7	13.1	−.000	.003	.511	.13	−.000	.002
Study 2									
Normal control	32	85.9	7.5	−.005	.003	.898	.05	−.004	.002
Simulating malingerer	30	56.3	9.7	−.003	.003	.579	.12	−.002	.002
Study 3									
Normal control	57	86.1	6.4	−.004	.002	.896	.06	−.004	.002
Naive simulating malingerer	58	51.8	11.4	−.002	.004	.546	.13	−.001	.002
Informed simulating malingerer	56	60.4	8.5	−.005	.002	.643	.11	−.003	.002

Note. CR = consistency ratio.

This resulted in the classification of 47 NC, 0 CI, and 81 SM as malingerers, which reflects excellent sensitivity but mediocre specificity. An analysis of the responses of control subjects who were incorrectly labeled revealed that they were predominantly individuals who had scored very well on the test. Because high-scoring subjects are unlikely to be confused with persons attempting to feign brain damage, the decision rule was modified to exclude any subject scoring above the test mean as computed for all 103 control subjects ($M = 82.2$). This adjustment resulted in 4 NC (4.7%), 0 CI, and 81 SM (96.4%) being classified as malingerers. However, because the number of all controls who scored below average was 34, the accurate false positive rate was 4/34 or 11.8%.

Figure 2 demonstrates the potential of the new decision rule. CI subjects and the more sophisticated malingerers, which are the two groups most difficult to discriminate clinically, occupy distinctly different distributions. Instead, the overlap in product values occurs between high-scoring control subjects and sophisticated malingerers, two groups not likely to be confused with each other in the clinical setting. Despite this apparent success, a decision rule created post hoc is problematic because it may apply only to the current sample. Therefore, a replication study was conducted.

Study 2. Testing the New Decision Rule

Method

Subjects and procedures. Sixty-three college students (43 female, 20 male; 62 White, 1 Black) participated in the replication study. One subject was excluded for failure to comply with instructions. Students received the same instructions for compliance (NC, $n = 32$) or decep-

Table 3
Number of Subjects Classified as Malingerers by Decision Rules

Decision rule	Study 1 NC ($n = 86$)	Study 1 CI ($n = 14$)	Study 1 SM ($n = 84$)	Study 2 NC ($n = 32$)	Study 2 SM ($n = 30$)	Study 3 NC ($n = 57$)	Study 3 NSM ($n = 58$)	Study 3 ISM ($n = 56$)
Score < 42	0	0	14	0	3	0	8	1
Slope > 0	0	0	35	0	5	0	16	1
CR < .50	0	0	35	0	6	0	20	4
Slope > 0 *or* CR < .50	0	0	51	0	9	0	28	5
Slope*CR > −.0041	47	0	81	19	28	29	53	41
Slope*CR > −.0041 *and* Score < 83[a]	4	0	81	0	28	1	53	41

Note. NC = Normal control; CI = Cognitively impaired; SM = Simulating malingerer; NSM = Naive simulating malingerer; ISM = Informed simulating malingerer; CR = consistency ratio.
[a] See text.

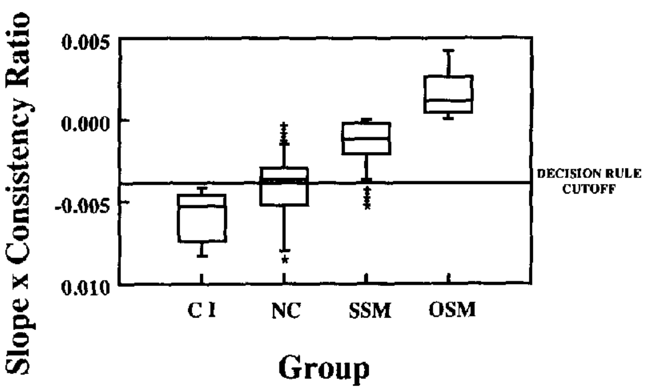

Figure 2. Distributions of slope–consistency ratio products by group in Study 1. (CI = Cognitively impaired; NC = Normal control; SSM = Sophisticated simulating malingerer; OSM = Obvious simulating malingerer)

tion (SM, $n = 30$) as used previously. They received only extra credit for their participation.

Results and Discussion

Group means for NC and SM are presented in Table 2. Use of the slope–consistency ratio product in the decision rules remained highly effective. No NC were misidentified, and 93.3% of the SM were detected (see Table 3). Mean consistency ratios for NCs in Study 2 (.898) were compared with those for NCs in Study 1 (.884) and revealed no significant difference, $t(116) = 1.21$, $p > .10$. This suggests that the high consistency ratios for NCs in Study 1 were not due to post hoc reordering of item difficulty.

Study 3. Testing the Robustness of the Forced-Choice Technique

A common assumption for tests that include validity scales is that even if examinees are aware that they are being checked for instruction compliance, they have little or no comprehension as to how that checking is accomplished (Power & O'Donovan, 1969). In this experiment, that assumption was violated for one group of simulating malingerers.

Method

Subjects. One hundred seventy-one college students (112 female, 59 male; 147 White, 15 Black, 7 Hispanic, 2 Asian) participated; 114 simulated malingering, 57 were normal control subjects. They received only extra credit for their participation.

Procedure. Within the group of malingerers, two categories were established, "naive," and "informed." Naive malingerers ($n = 58$) received instructions as before, but the informed group ($n = 56$) also read information on ways to successfully malinger. Specifically, they were warned to get at least one half of the answers correct, to correctly answer easy items, and to miss only more difficult problems.

Results and Discussion

Group means for NC subjects and for naive and informed malingerers are presented in Table 2. As before, NCs and naive malingerers were correctly classified most of the time (see Table 3). Furthermore, 73.2% of the malingerers who had been "tipped off" did not avoid detection. But most of the 15 informed malingerers who did escape detection produced realistic impairment: 14 scored at least one standard deviation below the mean score for control subjects across all three studies (mean test score = 84.0, $SD = 9.3$); 11 scored at least two standard deviations below the mean.

General Discussion

These studies have demonstrated measures, generated by a forced-choice technique, which effectively identified response patterns of individuals who simulated malingering on a cognitive test. A test incorporating these measures could be used as an adjunct to standard neuropsychological and intellectual batteries, providing an indicator of compliance, or it may be possible to develop procedures that yield estimates of cognitive abilities even if malingering is suspected (i.e., creating a correction factor). We have used a modified version of the TONI only to test hypotheses. A final test would not necessarily have to use a nonverbal format: The basic requirement is that the examinee has to choose between two answer choices among items representing a continuum of difficulty.

Several limitations are noted. Measures of slope, score, and consistency appear sensitive only to those malingerers who are uninformed or obvious. Nonetheless, these measures are specific only to malingerers. On the other hand, the slope–consistency ratio product is highly sensitive to all types of malingerers but lacks specificity with regard to high-scoring compliers. Whether this compromise in specificity is important depends on the base rate for bright individuals who deceptively present as "only average" on ability tests. The rate should certainly be quite low, considering the inverse relationship of demonstrated ability and secondary gain in the forensic arena. However, the slope–consistency ratio product decision rule has reasonable specificity with regard to low-scoring compliers. Of all the 56 control subjects who scored below average, 6 (10.7%) were labeled malingerers. It should be noted that the decision rule that incorporates the slope-consistency ratio product is highly test-based, because the magnitude of the slope is dependent on the overall difficulty of the items on the test. Perhaps an increase in test difficulty (raising the ceiling) would establish a more appropriate balance between sensitivity and specificity for high scorers.

Whether the lack of sensitivity manifested by the decision rule for test score has direct application for SVT is debatable. SVT uses 100 stimuli of equivalent intensity; we used a test with a hierarchy of difficulty. Furthermore, SVT typically incorporates immediate feedback for performance on each item, with the assumption that doing so tends to create a sense of performing "too well" in the malingerer (Binder & Pankratz, 1987). In this study, no subjects received individual feedback. Previous research on the SVT may also have relied on samples that are too small. Most of the published research on SVT is of only a few subjects (Binder, 1990; Binder & Pankratz, 1987; Hiscock & Hiscock, 1989; Pankratz, 1979, 1983; Pankratz, et al., 1975). In studies using SVT with larger samples, Haughton et al. (1979) could correctly identify only 11 out of 20 simulators in a hearing test, and Binder and Willis (1991) detected only 2 out of 13 simulators in a memory task. Therefore, although SVT certainly provides convincing evidence of malingering when subjects perform below chance, it may not be sensitive enough to be used in isolation, especially with perceptive examinees.

Simulation studies that use low-level inducements to validate measures of malingering may not be comparable to studies that use realistic incentives or actual malingerers (Gillis, Rogers, & Bagby, 1991; Rogers, 1988). But other studies have found no significant differences between groups of simulators on the basis of level of inducement to malinger (Bernard, 1990; Binder & Willis, 1991). Furthermore, limited cross-validation of the techniques described in this article on active, identified malingerers in a forensic hospital has been promising (Frederick, 1991). Nevertheless, a more expedient attempt at cross-validation might be to provide significant monetary motivators for successful malingering in another simulation study. One possibility would be to provide all subjects with a small remuneration simply for participating and to give larger bonus cash awards for malingered performances that result in classification as "compliant, but impaired" (cf. Schretlen & Arkowitz, 1990).

Cross-validation must also occur among a broader range and greater number of cognitively impaired individuals. A successful cross-validation will demonstrate differences in measures of response patterns (e.g., slope and consistency ratio) between cognitively impaired and malingering (simulated or actual) groups, despite a lack of significant differences between test scores (i.e., percentage correct) for the two groups.

Finally, the two-alternative forced-choice technique may yield more measures of discriminability than the few we have used. For example, the forced-choice technique might allow for the application of signal detection theory. Signal detection values such as d-prime or other receiver operating characteristics may prove useful in the detection of malingered performance.

References

Atkinson, L. (1991). On WAIS-R difference scores in the standardization sample. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3,* 292–294.

Bash, I. Y., & Alpert, M. (1980). The determination of malingering. *Annals of the New York Academy of Science, 347,* 86–99.

Bernard, L. C. (1990). Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. *Journal of Clinical and Experimental Neuropsychology, 12,* 715–728.

Bickart, W. T., Meyer, R. G., & Connell, D. K. (1990, March). *The symptom validity test as a measure of feigned short-term memory deficit.* Paper presented at the meeting of the American Psychology–Law Society, Williamsburg, VA.

Binder, L. M. (1990). Malingering following minor head trauma. *The Clinical Neuropsychologist, 4,* 25–36.

Binder, L. M., & Pankratz, L. (1987). Neuropsychological evidence of a factitious memory complaint. *Journal of Clinical and Experimental Neuropsychology, 9,* 167–171.

Binder, L. M., & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3,* 175–181.

Brady, P. B., & Lind, D. L. (1961). Experimental analysis of hysterical blindness. *Archives of General Psychiatry, 4,* 331–339.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1982). *Test of Nonverbal Intelligence: A language-free measure of cognitive ability.* Austin, TX: Pro-Ed.

Faust, D., Hart, K., & Guilmette, T. J. (1988a). Pediatric malingering: Capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 56,* 578–582.

Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988b). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice, 19,* 508–515.

Frederick, R. I. (1991). [Cross-validation of forced-choice measures of malingering in a forensic population]. Unpublished raw data.

Gillis, J. R., Rogers, R., & Bagby, R. M. (1991). Validity of the M Test: Simulation-design and natural-group approaches. *Journal of Personality Assessment, 57,* 130–140.

Goebel, R. A. (1983). Detection of faking on the Halstead-Reitan Neuropsychological Test Battery. *Journal of Clinical Psychology, 39,* 731–742.

Goldstein, H. (1945). A malingering key for mental tests. *Psychological Bulletin, 42,* 104–118.

Grosz, H. J., & Zimmerman, J. (1965). Experimental analysis of hysterical blindness: A follow-up report and new experimental data. *Archives of General Psychiatry, 13,* 255–260.

Haughton, P. M., Lewsley, A., Wilson, M., & Williams, R. G. (1979). A forced-choice procedure to detect feigned or exaggerated hearing loss. *British Journal of Audiology, 13,* 135–138.

Heaton, R. K. (1981). *The Wisconsin Card Sorting Test Manual.* Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., Baade, L. E., & Johnson, K. L. (1978). Neuropsychological test results associated with psychiatric disorders in adults. *Psychological Bulletin, 85,* 141–162.

Heaton, R. K., Smith, H. H., Lehman, R. A. W., & Vogt, A. T. (1978). Prospects of faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 46,* 892–900.

Hiscock, M., & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology, 11,* 967–974.

Hunt, W. A. (1946). The detection of malingering: A further study. *Naval Medical Bulletin, 46,* 249–254.

Hunt, W. A., & Older, H. J. (1943). Detection of malingering through psychometric tests. *Naval Medical Bulletin, 41,* 1318–1323.

Matarazzo, J. D. (1990). Psychological assessment versus psychological testing: Validation from Binet to the school, clinic, and courtroom. *American Psychologist, 45,* 999–1017.

Matarazzo, J. D., & Prifitera, A. (1989). Subtest scatter and premorbid intelligence: Lessons from the WAIS-R standardization sample. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 186–191.

Mensch, A. J., & Woods, D. J. (1986). Patterns of feigning brain damage on the LNNB. *International Journal of Clinical Neuropsychology, 8,* 59–63.

Milner, B. (1963). Effects of brain lesions on card sorting. *Archives of Neurology, 9,* 90–100.

Pankratz, L. (1979). Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. *Journal of Consulting and Clinical Psychology, 47,* 409–410.

Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills, 57,* 367–372.

Pankratz, L., Fausti, S. A., & Peed, S. (1975). A forced-choice technique to evaluate deafness in the hysterical or malingering patient. *Journal of Consulting and Clinical Psychology, 43,* 421–422.

Power, R. P., & O'Donovan, D. (1969). Detection of simulation on the MPI by subjects given the rationale of the lie scale. *British Journal of Psychology, 60,* 535–541.

Robinson, A. L., Heaton, R. K., Lehman, R. A. W., & Stilson, D. W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. *Journal of Consulting and Clinical Psychology, 48,* 605–614.

Rogers, R. (1984). Towards an empirical model of malingering and deception. *Behavioral Sciences and the Law, 2,* 93–111.

Rogers, R. (1988). Researching dissimulation. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 309–327). New York: Guilford Press.

Schretlen, D. J. (1988). The use of psychological tests to identify malingered symptoms of mental disorder. *Clinical Psychology Review, 8,* 451–476.

Schretlen, D. J., & Arkowitz, H. (1990). A psychological test battery to detect prison inmates who fake insanity or mental retardation. *Behavioral Sciences and the Law, 8,* 75–84.

Theodor, L. H., & Mandelcorn, M. S. (1973). Hysterical blindness: A case report and study using a modern psychophysical technique. *Journal of Abnormal Psychology, 82,* 552–553.

Wilkinson, L. (1981). *SYSTAT: The system for statistics* [Computer program]. Evanston, IL: SYSTAT.

Received September 13, 1990
Revision received November 29, 1990
Accepted December 17, 1990 ■