Detection of neurocognitive feigning: development of a multi-strategy assessment

Scott D. Bender a,*, Richard Rogers b

aDepartment of Neurology, University of Virginia Health Sciences Center, 500 Ray C. Hunt Drive, P.O. Box 801018, Charlottesville, VA 22908-1018, USA

bUniversity of North Texas, Denton, TX, USA

Accepted 12 August 2002

Abstract

Neuropsychological assessments can be completely invalidated by persons successfully feigning neurocognitive impairment. The current investigation examines via a research measure, the Test of Cognitive Abilities (TOCA), the usefulness of multiple detection strategies for the classification of neurocognitive feigning. Using a simulation design with a manipulation check and both positive and negative incentives, two groups of simulators (Cautioned and NonCautioned) were compared with brain-injured patients and nonimpaired controls. Among detection strategies, Magnitude of Error (hit rate = .94) was highly effective, while Floor Effect (hit rate = .80) and Reaction Time (hit rate = .85) were moderately effective. When presented with complex strategies, the cautioning of simulators did not improve their performances.

© 2002 National Academy of Neuropsychology. Published by Elsevier Ltd. All rights reserved.

Keywords: Malingering; Neuropsychology; Strategies; Cognitive


The problem of malingered performance in neuropsychological assessment has received considerable attention and researchers have proposed numerous tests, techniques, and strategies to detect feigned neurocognitive deficits (e.g., Amin & Prigatano, 1993; Binder, 1993; Guilmette, Hart, & Giuliano, 1993; Hiscock & Hiscock, 1989; Lees–Haley, Smith, Williams, & Dunn, 1996; Nies & Sweet, 1994; Rogers, 1997; Van Gorp, Humphrey, & Kalechstein, 1999).
These methods include the examination of atypical performance both within and between neuropsychological tests, documentation of unusual or incongruent behavioral presentation, and the use of specific tests of feigning (see Sweet, 1999). Many tests, designed specifically to detect feigning, evaluate performance on easy items and are limited to one domain (e.g., memory). More recent tests evaluate performance curves (Frederick & Foster, 1997), the Magnitude of Errors (Martin, Franzen, & Orey, 1998), and improbable response times (Rose, Hall, & Szalda–Petree, 1995). Despite a steady increase in the number of studies, data do not consistently support the use of a single strategy over another. In light of the current data, the use of multiple measures and strategies is strongly recommended (Rogers, Harrell, & Liff, 1993; Sweet, 1999).

1.1. Strategy-based assessment of feigning

Most research on feigning makes little or no reference to theory-driven strategies. However, the few available studies have produced highly effective measures for detecting feigned neurocognitive performance (e.g., Frederick, 2000; Martin, Franzen, & Orey, 1998). Rogers et al. (1993) proposed six strategies for detecting feigned performance on neuropsychological assessments: (a) Floor Effect, (b) Magnitude of Error, (c) Performance Curve, (d) Symptom Validity Testing (SVT), (e) Atypical Presentation, and (f) Psychological Sequelae. In the current study, the first four strategies plus Response Time were incorporated into a new test of feigning, the Test of Cognitive Abilities (TOCA; Rogers, 1996). First, the Floor Effect strategy is the most widely studied strategy and relies on malingerers’ unawareness of which items are too simple to be legitimately failed. The Floor Effect has been applied both to existing measures (e.g., “Reliable Digits” on the Wechsler Memory Scale-Revised (WMS-R; Meyers & Volbrecht, 1998; Wechsler, 1987) and to tests designed specifically for feigning (e.g., Test of Malingered Memory (TOMM); Tombaugh, 1997). The majority of data supports its use; however, a major drawback is its transparency as a detection strategy. Second, the Magnitude of Error strategy assumes that feigners will not be especially concerned about which incorrect responses they select. Magnitude of Error has been employed successfully on the WMS-R (Martin et al., 1998). Third, the Performance Curve strategy examines patterns across items based on their difficulty; malingerers are not likely to consider these patterns when feigning impairment. Performance Curve has shown promise on the Dot Counting Test (Binks, Gouvier, & Waters, 1997), the Validity Indicator Profile (VIP; Frederick & Foster, 1997), Raven’s Progressive Matrices (Gudjonsson & Shackleton, 1986), and the Category Test (Tenhula & Sweet, 1996). Fourth, Symptom Validity Testing examines whether the failure rate drops below-chance levels. SVT has been effectively used with Portland Digit Recognition Test (PDRT; Binder, 1993) and the VIP (Frederick & Foster, 1997).

Response Time has also been tested as a detection strategy for feigned cognitive impairment. Response Time differs from the formal detection strategies in its de-emphasis of a theoretical framework. This method has produced mixed results on an experimental letter recognition task (Boone et al., 2000), on the Memory Assessment Scales (MAS; Beattar & Williams, 1995), and on a computerized version of the PDRT (Rose et al., 1995).

Rogers et al. (1993) and Frederick and Foster (1991) have suggested that research should attempt to improve the accuracy of classification through the combination of strategies.
Beyond accuracy, corroboration by multiple strategies may increase clinicians’ confidence in their conclusions. Moreover, even sophisticated feigners are likely to be challenged by the complexity of the task when attempting to foil multiple strategies.

1.2. Should tests measure both feigning and ability?

An important consideration is whether feigning measures should serve dual purposes in measuring attempts to feign and also assessing neuropsychological abilities. If the assessment of feigning is confounded by abilities, then the measure is likely to be invalid. Therefore, feigning measures often do not assess abilities per se. For example, the TOMM relies chiefly on Floor Effect. With both impaired and unimpaired patients scoring at or above the 90th percentile on its second trial, the TOMM only addresses feigning and not abilities. In contrast, specialized indexes developed from standard assessment measures (e.g., Raven’s Standard Progressive Matrices) may provide information about abilities as well as potential feigning. However, these dual-purpose instruments are often less sensitive than focused feigning measures.

1.3. The effects of warning

Researchers have also discussed the effect of warning clients that feigning detection strategies are in place. Warning is distinguished from coaching in that the latter involves providing information to the client regarding the procedures of testing and/or purposes of specific tests in order to improve chances of feigning successfully. In contrast, a warning simply cautions the client that methods of feigning detection will be employed. Youngjohn, Lees–Haley, and Binder (1999) argue that warning malingerers will increase their sophistication rather than reduce attempts at malingering; thus, they conclude that warning clients is ill-advised. In contrast, Johnson and Lesniak–Karpiak (1997) argue that warning helps reduce the potential for feigning by informing the clients that they probably will be caught if they feign. They also contend that using detection techniques without informing the patient may be viewed as deceptive and therefore ethically questionable.

Beyond conceptual issues, Rogers (1997) suggests that informing the client about multiple detection strategies may actually improve a test’s clinical utility. Given this knowledge, many malingerers may feel challenged to foil each of the multiple strategies. This effort to thwart the test could easily backfire. For example, a would-be malingerer may be unsuccessful at simultaneously attempting to (a) account for item difficulty (Performance Curve), (b) make plausible errors (Magnitude of Error), and (c) put forth credible effort as measured indirectly by latency of response (Response Time). With its important clinical implications, this hypothesis deserves empirical testing.

1.4. Objectives of the current study

The Test of Cognitive Abilities (Rogers, 1996) was designed as a computer-administered multi-strategy measure of both feigning and cognitive ability. The current study, an initial validation of the TOCA, has several objectives. First, it examines the clinical utility of scales based
on the detection strategies. Second, it examines whether single or multiple strategies achieve higher classification rates of genuine and feigned presentations. Third, the study explores the effect of a warning on the performance of simulators.

2. Method

The study is a simulation design with the quasi-random assignment of participants to experimental (simulation) and control (honest) conditions. To improve its clinical relevance, the design is augmented with a clinical comparison sample. This critical provision tests to see if differences between simulators and controls will also hold for brain-injured populations.

2.1. Participants

The total sample was composed of 123 participants (68 women). Sixty nonclinical participants were randomly assigned to one of two groups of simulators: Cautioned (CS) and NonCautioned (NCS). These groups were compared to (a) a group of normal controls (NC, \( n = 22 \)) and (b) a clinical group of neurocognitively compromised patients (CL, \( n = 42 \)). All participants gave written informed consent.

Neurocognitively impaired participants were composed of inpatients and outpatients with a history of traumatic brain injury (TBI, \( n = 23 \)) or cerebrovascular accident (CVA, \( n = 18 \)). About 60% of the study patients were judged (by the treatment team) to have sustained mild neurologic injuries, with the remainder having severe injuries. Very few participants (two, or 4.9%) had pending litigation or any other claim related to their neurocognitive impairment. They were eligible for the experiment if they were free from significant visual and hearing deficits, familiar with the use of a computer keyboard, and agreed to be tested for approximately one hour. Nonimpaired volunteers were recruited from the University of North Texas and had no history of neurologic or psychiatric disorders.

2.2. Instrument

2.2.1. Test of Cognitive Abilities

The TOCA (Rogers, 1996) is a computerized measure of response patterns that employs multiple detection strategies. It uses a multiple-choice format in which the respondent chooses twice from four possible responses (i.e., each item consists of two trials). Each participant must complete a sequence by choosing the correct beginning and end responses (e.g., 5, 4, 3, _) from four possible choices (e.g., 7, 2, 6, 4). The TOCA is composed of 112 questions, each requiring a beginning and an end response for a total of 224 responses. The test is organized into three sections: The Sequencing Section (60 questions) assesses the ability to recognize alpha-numerical sequences (as above). The Designs Section (22 questions) involves recognition of patterns based on four design parameters: shape, color, size, and shading (e.g.,__). The Sentences Section (30 questions) assesses verbal comprehension of incomplete sentences (e.g., ___ ⊖ □ ⊖ __): releasing information about scale composition would violate test security and could potentially lead to the coaching of actual malingerers (see Sweet
et al., 2000), sensitive information about scale composition is not included, but is available to qualified researchers upon request.

The TOCA’s instructions can be configured to employ a warning at the beginning of the test that cautions against faking and includes descriptions of detection strategies. By presenting simulators with this information, they must perform multiple cognitive decisions (e.g., estimating item difficulty and deciding whether or not to miss the item, while maintaining a plausible Reaction Time) in an effort to avoid detection. The cautionary statement appears on the computer screen following the general instructions.

2.3. Procedure

Participants were informed about the general goals of the study, namely to conduct research on a cognitive measure under different instructional sets. With written informed consent, all participants were introduced to the TOCA and given sample items. Each unimpaired participant was also provided with an unmarked envelope from a randomly-sorted stack that included their experimental instructions. Because unimpaired participants were recruited from undergraduate classes, they were offered research credit for their involvement in addition to the incentives described below.

Unimpaired participants were tested in a group setting (i.e., computer labs). This procedure along with the use of unmarked envelopes masked the researcher to experimental conditions. Participants were allowed to proceed at their own pace.

Participants in the simulator groups (NCS and CS) were asked to imagine themselves in a particular scenario about a motor vehicle accident (modeled after Rose et al., 1998). To parallel real-world contingencies, the scenario provided a positive incentive (i.e., financial settlement) for successful feigning and a negative incentive (i.e., criminal sanctions for perjury) for unsuccessful feigning. Given a brief description of brain injury, they were asked to feign plausible impairment:

Your goal is to take this test in such a way that will convince the examiner that you have suffered permanent brain damage from the accident. You are 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. In addition, you may be fined or jailed for lying.

Simulators proceeded to take the TOCA when they felt sufficiently prepared. They were provided with positive and negative incentives such that the three “most successful malingerers” would be given $20.00 in addition to research credit, and the “three poorest malingerers” would have their names and results posted. To avoid ethical problems, the positive incentives were rewarded randomly and only fictitious names were posted. Cautioned simulators received additional instructions that the TOCA had safeguards against persons attempting to foil the measure. The key portion of these instructions is reproduced:

WARNING. Every now and then, someone tries to “fake” the test by doing a bad job. Please don’t do this. The test has many safeguards to stop persons from faking it. These safeguards identify people who may be faking. Some of the safeguards are listed below:
The test checks to make sure you miss more of the difficult items than easy items.
The test checks to see how many seconds you take on each item and compares it to the
difficulty of the item.
The test checks to see if you are trying by comparing your ability on similar items.
The test checks to see if you made careless errors, particularly on very easy items.
The test checks to see if you get the same wrong answers as most people do.
The test checks to see if you make more mistakes than expected by chance (probability)
alone.
Don’t worry about these safeguards, just put forth your best effort.

Participants in the NC group were given instructions to perform their best on the TOCA. Like
the simulators, they were provided with positive and negative incentives. The only difference
was the specification of “highest scorers” and “lowest scorers” on the TOCA.
The CL group received the same instructions as the NC group but two other procedures
were different. First, the TOCA was administered individually. Second, the CL group was
judged to be highly motivated to improve their cognitive status; therefore, no advantage was
seen in offering positive incentives, and negative incentives (even with fictitious names) might
unnecessarily induce stress in a vulnerable population.

2.4. Manipulation check

Rogers (1997) strongly recommended a manipulation check be used to ensure that patients
in simulation conditions complied with instructional sets. Nine of the 60 simulators in this
study were removed from subsequent analyses for either not correctly recalling the instructions
or reporting negligible effort at feigning.

2.5. Scale composition and cut scores

To minimize the number of exploratory analyses, we proposed several a priori rules for the
scale composition of each detection strategy. For instance, we tested the relative effectiveness
of Floor Effect scales when the threshold criterion for inclusion was 90% versus 95% of the
CL group having the correct response. This approach attempted to balance the need for (a)
scientific rigor with the use of a priori rules and (b) better classifications with the testing of
several alternatives.¹

3. Results

The first research issue, involving the relative efficacy of specific detection strategies, was
examined by effect sizes and utility estimates. As reported in Table 1, TOCA scales based
on detection strategies evidence group differences in the predicted direction. The individual
strategies vary substantially by effect sizes from moderate to large. Magnitude of Error

¹For interested researchers, expanded tables delineating each alternative to scale composition are available from
Dr. Bender along with utility estimates.
Table 1
Differences and effect sizes (Cohen’s d) for honest responders and simulators on the TOCA detection strategies

<table>
<thead>
<tr>
<th>Scale</th>
<th>Honest</th>
<th></th>
<th>Simulators</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CL</td>
<td>NC</td>
<td>CS</td>
<td>NCS</td>
<td>F</td>
<td>Cohen’s d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floor Effect</td>
<td>21.29 (1.56) a</td>
<td>21.68 (0.72) a</td>
<td>18.09 (4.36) b</td>
<td>17.48 (5.39) b</td>
<td>10.73</td>
<td>1.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnitude of Error&lt;sup&gt;c&lt;/sup&gt;</td>
<td>35.42 (17.88) a</td>
<td>31.45 (11.84) a</td>
<td>90.39 (14.00) b</td>
<td>86.28 (20.83) b</td>
<td>95.22</td>
<td>2.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance Curve</td>
<td>1.14 (3.16) a</td>
<td>1.55 (3.65) a</td>
<td>-1.64 (4.56) b</td>
<td>-.90 (3.19) b</td>
<td>4.88</td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response Time</td>
<td>79.18 (27.50) a</td>
<td>51.85 (9.32) b</td>
<td>48.64 (14.06) b</td>
<td>47.65 (12.15) b</td>
<td>8.80</td>
<td>1.33</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TOCA: Test of Cognitive Abilities. For groups, CL: clinical, CS: Cautioned simulators, NCS: NonCautioned simulators, and NC: normal controls. Group means with different letters (a, b) are significantly different by Tukey comparison, P < .01. Given the lack of difference between CS and NCS scores, the simulator groups were combined. The effect size shown reflects the magnitude of difference between mean CL score and combined simulators.

<sup>a</sup> Symptom Validity Testing (SVT) is not included in the table because it does not consist of specific items but rather below-chance responding.

<sup>b</sup> For F ratios, P < .01.

<sup>c</sup> To account for significant heterogeneity in variance, Magnitude of Error mean scores were transformed into ranked data before further statistical analysis (Conover, 1999).
Table 2
Utility estimates for classifying simulators versus brain-injured patients on specific TOCA scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPP</th>
<th>NPP</th>
<th>HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor Effect</td>
<td>.80</td>
<td>.81</td>
<td>.84</td>
<td>.77</td>
<td>.80</td>
</tr>
<tr>
<td>Magnitude of Error</td>
<td>.94</td>
<td>.93</td>
<td>.94</td>
<td>.93</td>
<td>.94</td>
</tr>
<tr>
<td>Performance Curve</td>
<td>.65</td>
<td>.73</td>
<td>.75</td>
<td>.63</td>
<td>.72</td>
</tr>
<tr>
<td>Reaction Time</td>
<td>.88</td>
<td>.78</td>
<td>.83</td>
<td>.84</td>
<td>.85</td>
</tr>
<tr>
<td>Symptom Validity Testinga</td>
<td>.03</td>
<td>1.00</td>
<td>1.00</td>
<td>.42</td>
<td>.43</td>
</tr>
</tbody>
</table>

TOCA: Test of Cognitive Abilities; PPP: positive predictive power; NPP: negative predictive power; HR: hit rate.

a Two participants scored significantly below-chance on the truncated version of SVT; these estimates are based on this (our most successful) approach to SVT.

(Sentence Completion Section) appears to be the most effective at differentiating simulators from patients with a very robust effect size of 2.99.

Although simulators educated through coaching and cautions are likely to be less detectable, an issue posed by the current study is whether this finding would hold true when faced with multiple strategies. As observed in Table 1, extensive cautioning did not have a significant effect on simulators’ performance. Indeed, some detection strategies (Floor Effect and Response Time) produced virtually identical results. For Magnitude of Error, coached simulators did slightly worse than noncoached simulators (Cohen’s $d$ for CS vs. NCS = .20).

Beyond effect sizes, utility estimates provide a useful and practical method of assessing the accuracy of specific detection strategies. As reported in Table 2, three strategies appeared moderately effective (i.e., hit rates $\geq .80$): Floor Effect, Magnitude of Error, and Response Time. Of these, Magnitude of Error provided outstanding estimates (> .90) and appears to be equally effective at ruling-out (i.e., NPP = .93) and ruling-in (i.e., PPP = .94) neurocognitive feigning. Of the strategies, SVT was the least successful because only two participants scored significantly below-chance. Because both participants were simulators and no patients were incorrectly classified as feigning, the PPP and the specificity were both 1.0. However, the very poor sensitivity (.03) revealed that the overall utility of SVT was severely limited by its restrictively high threshold for detecting feigners.

An additional research question posed by the current study was whether multiple strategies would increase the classification rates. In light of the stellar performance by Magnitude of Error with its hit rate of .94, the obvious ceiling effect limits the testing of this hypothesis. Nevertheless, we performed a stepwise discriminant analysis (Table 3) with the use of the four detection strategies.

Table 3
Stepwise discriminant analysis of the TOCA for classifying simulators and brain-injured patients

<table>
<thead>
<tr>
<th>Predicted membership</th>
<th>Actual membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulators ($n = 51$)</td>
<td>96.1% (49)</td>
</tr>
<tr>
<td>Brain-injured patients ($n = 42$)</td>
<td>4.9% (2)</td>
</tr>
</tbody>
</table>

Sensitivity = .96; specificity = .95; PPP = .96; NPP = .95; hit rate = .96. For the discriminant function, the canonical coefficients are .77 (Magnitude of Error) and −.66.
scales as predictor variables. Only two scales (Magnitude of Error and Response Time) were significant and contributed to the discriminant function. The discriminant function is highly significant with a Wilk’s $\lambda = .42$, $P < .0001$. Its overall classification of .96 represented a very slight improvement over Magnitude of Error alone. Given that single-stage discriminant analyses capitalize on chance variability, some shrinkage is likely on cross-validation.

As exploratory analyses, we conducted several post hoc analyses to test the effects of effort and perceived success on simulators’ performance. Based on the debriefing, we grouped simulators by their reported effort: 30 with high effort and 22 with moderate effort. We tested with individual ANOVAs the effects of effort on specific detection scales. We found no significant differences. We also divided the simulators into “perceived successful” (i.e., 21 participants rated themselves as “quite” to “very” successful at “feigning”) and “perceived unsuccessful” (i.e., 32 participants rated themselves from “not” to “a little” successful). Individual ANOVAs on four detection strategies consistently yielded nonsignificant results.

4. Discussion

4.1. Detection strategies

An important and unexpected finding was the usefulness of Magnitude of Error as a detection strategy for neurocognitive feigning. First described by Rogers et al. (1993), Magnitude of Error has only recently been researched. Martin et al. (1998) used Visual Reproduction and Logical Memory subtests of the WMS-R with a multiple-choice format. Consistent with the current study, they found Magnitude of Error to be highly effective, reporting very high classification rates for feigners (86.0% for simulators, 100% for suspected malingerers) and moderately high rates for brain-injured patients (80.0%). Taken together, Magnitude of Error shows great promise as a newly-tested detection strategy. As observed by Rogers and Bender (in press), malingerers are likely to focus more on what items to fail rather than how to fail these items. If correct, Magnitude of Error may lack the transparency found with some other detection strategies, such as Floor Effect.

The relative failure of SVT as a detection strategy was unexpected. While past research had found SVT to have poor PPPs, its unmatched NPPs tended to compensate for this limitation. Additionally, Rogers and Shuman (2000) hypothesized that low PPPs might reflect researchers’ insistence on simplistic two-choice formats. By extrapolation, more complex versions could be expected to produce more positive results. In line with this reasoning, SVT was examined for several options, including (a) the likelihood of a correct initial choice (.25) and (b) the likelihood that one of the two choices would be correct (.41). A truncated version of the .41 alternative produced just two below-chance performances, while the other levels of chance were unproductive.² At least with the current study, increased complexity partially satisfied the intended goal (i.e., to improve PPP) but at the expense of extremely poor sensitivity and

² As a post hoc analysis, we eliminated very easy items that might appear too obvious and very difficult items that might encourage guessing. Two simulators (and zero patients) scored significantly below-chance using this truncated approach.
lower than expected NPP. The probabilities we used appear to have been too restrictive for significantly below-chance responding, and likely contributed to the poor sensitivity. Rather than requiring “significantly below-chance performance,” cutoff scores may prove more clinically useful. Another limitation of the TOCA, as amply demonstrated by Magnitude of Error, is that its responses lack equiprobability. Therefore, we recommend further testing of more complex SVT models with equiprobable alternatives.

An important objective of the study was to test whether combined strategies resulted in incremental validity. The strong effect size and outstanding utility estimates for Magnitude of Error constrained our ability to evaluate fully the incremental validity of detection strategies for neurocognitive feigning. The key issue will be cross-validation. With an expected attenuation in classification, the relative usefulness of Magnitude of Error or combined strategies (i.e., Magnitude of Error plus Response Time) can be more fully tested in cross-validation research.

The effect of cautions on the detectability of neurocognitive feigning has not been systematically evaluated. With multiple strategies, Rogers (1997) suggested that cautions are not likely to reduce the effectiveness of the strategies and even have a paradoxical effect in improving detection. The current data suggest that cautions do not harm detection strategies. In the case of Magnitude of Error, a slight trend (Cohen’s $d = .20$) was observed suggesting the possibility that a few simulators may even perform worse when cautioned.

Data from the current study suggest that persons feigning neurocognitive impairment are unlikely to have an accurate appraisal of their success. When comparing participants on their perceptions (i.e., successful vs. unsuccessful), we found no significant differences in their performances. In light of the current data, neither self-appraisal of success nor effort appear to affect performance on detection strategies.

4.2. Methodological considerations

The strengths of the current study involved its attempts to reduce threats to external validity. To parallel real-world applications, both positive and negative incentives were considered. Positive incentives, contingent on successful feigning (see Bernard, 1990) minimize extreme presentations as noted by Franzen and Martin (1996). Although often overlooked in simulation research, Rogers and Cruise (1998) found negative incentives to be more influential than positive incentives on feigning performances. These incentives were addressed both in the scenarios (i.e., financial settlement vs. criminal sanctions) and in participants’ success (i.e., the payment of “best” simulators vs. a public posting of “worst” simulators).

Simulation research of neurocognitive feigning often neglects basic methodological issues. In a review of 19 studies, Nies and Sweet (1994) found that only two studies used manipulation checks and that none provided highly specific instructions to simulators. Moreover, none of the reviewed studies used both positive and negative incentives. Clearly, future research needs to address these basic methodological concerns.

Researchers may wish to consider at least four directions for future research. First, a critical issue to be addressed is whether specific detection strategies are measure-specific. For example, does the effectiveness of the Magnitude of Error strategy hold only for the TOCA and modified scales of the WMS-R? Second, the generalizability of findings is essential. For instance, do the current findings hold for (a) specific conditions (e.g., dementia) and (b) comorbidity (e.g.,
major depression secondary to TBI? Third, researchers may wish to consider refinements in the current detection strategies. We might consider the relationship between item difficulty (Performance Curve) and Response Time rather than treat these scales independently (see Wogar, Van den Broek, Bradshaw, & Szabaldi, 1998). Fourth, the adequacy of incentives for simulation research remains a perennial problem. Despite ethical concerns and budgetary constraints, better approximations of positive and negative incentives are needed.

In closing, the systematic assessment of neurocognitive feigning has clearly advanced during the last several decades. The current findings deserve replication, including (a) the superiority of Magnitude of Error, (b) the lack of advantage for cautioned simulators, and (c) the usefulness of complex detection strategies as found in the TOCA. Continued efforts are essential to the practice of clinical neuropsychology in refining detection strategies and improving their classification rates.

References


