Serial Malingering on Verbal and Nonverbal Fluency and Memory Measures: An Analog Investigation

George J. Demakis

Elmhurst College

Analysis of response consistency on neuropsychological test performance, both within and across testing sessions, can be an important method of detecting malingering. Little systematic research, however, has examined how suspected malingerers perform across repeat evaluations, a common forensic occurrence. To address this issue, we examined performance across a 3-week interval in an analogue malingering design on the California Verbal Learning Test (CVLT), the Rey Complex Figure, the Controlled Oral Word Association Test, and the Ruff Figural Fluency Test. Malingering simulators (n = 21) performed more poorly on all measures than the controls (n = 21) and demonstrated practice effects on the nonverbal, but not the verbal, tests. Controls demonstrated practice effects on all measures across time. Contrary to hypotheses, malingering simulators demonstrated high and similar levels of between and within time consistency as controls when assessed via a series of correlations. Despite this consistency, when qualitative performance patterns were assessed on the CVLT, simulators were less likely to consistently recall the same word across successive learning trials. The following issues are discussed: (a) the differential pattern of practice effects on verbal and nonverbal tasks, (b) qualitative and quantitative differences in assessment of consistency, and (c) how future research should study consistency/inconsistency. © 1999 National Academy of Neuropsychology. Published by Elsevier Science Ltd

Several methods of detecting malingering have been devised, including the use of specific malingering tests (e.g., forced choice tests), analysis of performance level and pattern on commonly used neuropsychological tests, and examination of nonsensical test patterns. While considerable research has investigated these approaches (for reviews, see Nies & Sweet, 1994 and Rogers, Harrell, & Liff, 1993), less systematic research has examined performance consistency, whether between or within test sessions, in suspected malingerers. This gap is surprising, given the frequency with which suspected malingerers are evaluated in the forensic context and suggestions that malingerers may...
present with clinical and/or neuropsychological test performance inconsistencies (Cul-
lum, Heaton, & Grant, 1991; Rawling, 1993; Reitan & Wolfson, 1998; Rogers, 1984; Zie-
linski, 1994). These inconsistencies can be an important way of detecting malingering
and can manifest in several ways, with poor or unexplainable relationships between the
following: injury severity and cognitive function, premorbid abilities and abilities mea-
sured postinjury, tests tapping the same construct (e.g., verbal memory), and repeat ad-
ministrations of the same test or test items (Rawling, 1993). Presumably, these inconsis-
tencies become apparent during the neuropsychological evaluation, as malingerers do
not consistently and fully exert maximal effort. Overall, despite the relative paucity of
research in this area, many clinicians appear to agree with Cullum et al. (1991) that as-
essment of response consistency is a powerful means of detecting malingering.

A common approach used by investigators to demonstrate the importance of re-
sponse consistency is case study analysis of neuropsychological performance. For in-
stance, Cullum et al. (1991) presented three case studies of serial performance by sus-
ppected malingerers in the forensic setting. Compared to controls or neurologically stable
patients, they demonstrated highly variable neuropsychological performance over two
or more evaluations that did not reflect a consistent pattern of improvement or deterio-
ration. For example, Patient 2 demonstrated a marked deterioration in performance on
Trails B (+129 seconds), which, according to the authors, was suspicious of malingering.
In a quite different, but relevant, case study by Sweet and Kuhlman (1993), consistency
across four repeat neuropsychological examinations was viewed as critical to estab-
lishing the veracity of a litigant’s complaints. Putnam, Adams, and Schneider (1992) also
demonstrated fairly consistent performance, with some expected practice effects, in a
patient who underwent forensic neuropsychological evaluation on two successive days.
Although not concerned with malingering per se, the similar performance on both days
could be used to bolster the conclusion that this patient exerted maximal effort on test-
ing and responded honestly. In all, these case studies illustrate how clinicians tend to assess
and use response consistency (or inconsistency) to make clinically meaningful decisions.

In addition to these case studies, a few systematic efforts have assessed response con-
sistency and test-retest performance. Paul, Franzen, Cohen, and Fremouw (1992) exam-
ined serial malingering on 2-week retesting on Rey’s Dot Counting and Memorization of
16 Items tests. Relative to controls, malingering simulators performed more consistently
on four of the seven measures assessed on these tests. Lezak (1995), however, notes that
the relatively low correlations for the controls likely reflected ceiling effects, as most
performed with near-perfect accuracy statistically exaggerating the effect of slight
changes in performance between trials. In a more clinically relevant study, Trueblood
(1994) examined neuropsychological performances of a small group of documented ma-
ingerers and head-injury controls. Though the malingerers scored significantly more
poorly than controls on several memory and intelligence measures, they responded as
consistently as controls on several verbal memory tasks, including the California Verbal
Learning Test (CVLT) and Logical Memory and Paired Associates subtests of the
Wechsler Memory Scales. Consistency was operationalized as the difference between
each person’s highest- and lowest-ranked memory score; each score was ranked based
on its standing with the combined group of control and malingering subjects. Malinger-
ers also performed as consistently across the five learning trials of the CVLT. Two stud-
ies by Reitan and Wolfson (1996, 1997) used archival data to compare response consist-
tency and test-retest performance in head-injured subjects in litigation and those not in
litigation. In their 1996 study, nonlitigating head-injured subjects performed more con-
sistently across two evaluations than litigating subjects on the same items from several
Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) subtests. In their
1997 study, nonlitigants were significantly more likely to demonstrate practice effects on several WAIS-R subtests than the litigants. A Dissimilation Index was developed that combined findings from both the consistency and test-retest data that perfectly (100%) classified litigants and nonlitigants. Despite such robust findings, these two studies are inherently limited because only one of the subjects was a documented malingeringer. A study by Sewick and Sobota (1996) was also unable to replicate these findings. Overall, although direct comparison between the above studies is limited given significant differences in rationale, sample, and methodology, they provide some of the only systematic information on response consistency and serial evaluation in malingers and/or litigants.

The current study was designed to systematically examine test-retest effects and response consistency in malingering simulators. As our dependent measures, we choose the following tests: the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987), the Rey Complex Figure Test (RCF; Meyers & Meyers, 1995), the Controlled Oral Word Association Test (COWAT; Ruff, Light, Parker, & Levin, 1996), and the Ruff Figural Fluency Test (RFFT; Ruff, 1988). The CVLT was selected for several reasons: It is frequently used in neuropsychology, has multiple quantitative and qualitative measures, and has been previously used to detect malingering (Millis, Putnam, Adams, & Ricker, 1995). Another consideration for its use was that it contains a consistency variable, in which the ability to recall the same word across the five successive learning trials is evaluated. The other measures were chosen because they are also frequently used in forensic neuropsychology, but little research has assessed malingered performance on them. These measures were administered to participants in an analogue malingering design in which half were instructed to malinger on two testing sessions separated by approximately 3 weeks. Controls were instructed to perform their best on two testing sessions also separated by about 3 weeks. For the two memory measures, we a priori selected quantitative variables (i.e., variables that assess number of words learned or recalled) that provide clinically relevant information, that have some degree of redundancy, and that have relatively normal and nonskewed distributions. This last point is critical, as correlations between skewed measures, with ceiling or floor effects, often cannot be meaningfully interpreted. Based on these criteria, we selected CVLT variables of total words learned across the five learning trials (i.e., list A trials 1–5 total), short-delay free recall, and long-delay free recall, and RCF variables of immediate and delayed recall. For the CVLT, we additionally selected one qualitative variable—percent recall consistency across the five learning trials—because it assesses a performance pattern relevant for our investigation. Percent recall consistency is calculated from the formula \[\frac{\text{Number of times a correct item recalled on List A Trials 1–4 was recalled on the next trial}}{\text{Total number of correct items recalled on List A Trials 1–4}} \times 100\].

Compared to malingering simulators, we made the following predictions: (a) Controls would perform better at both testing sessions. (b) Controls would improve across the testing sessions. (c) Controls would perform more consistently on each variable across the testing sessions. (d) Controls would perform more consistently within each testing session on related measures (e.g., short- and long-delay recall). (e) Controls would perform more consistently (i.e., tend to repeatedly recall the same word) across the five CVLT learning trials.

**METHOD**

**Participants**

Forty-three students from undergraduate psychology courses at a small midwestern liberal arts college participated. One subject was excluded from the malingering condi-
tion because she did not follow instructions, as was evident from responses on the de-
briefing questionnaire. Overall, 21 participants were assessed at both time periods in
each condition. Controls and malingering simulators were of similar age (controls = 22.5
years, $SD = 7.99$; simulators = 24.1 years, $SD = 7.49$), education (controls = 13.6 years,
$SD = 1.46$; simulators = 13.8 years, $SD = 1.5$), and gender composition (controls = 67%
female, simulators = 71% female).

Procedures

After completion of demographic questionnaires, participants were handed an enve-
lope with instructions on how to perform. In the malingering group, participants were in-
structed to imagine that they had sustained a head injury in a minor car accident that was
not their fault. They were then told to assume that although they initially had memory
and concentration problems, as well as headaches, these resolved a short time after the
accident. Despite resolution of these symptoms, they were instructed to imagine that
they had decided to pursue a lawsuit against the individual who hit them. They were
then told to fake common symptoms of head injury in the most realistic fashion to pre-
vent detection by the examiner. Participants were informed about the most common
symptoms of head injury and common performance patterns of head-injured patients
(e.g., free recall worse than recognition, and have memory difficulties, but do not forget
everything). To also approximate real-life conditions, they were told that if they avoided
detection by the experimenter, their monetary compensation would increase from $5 to $10.

Control group participants were told that they were in a car accident, but that they
had not suffered any injuries. They were then simply instructed to perform to the best of
their ability on the various tests. After reading their respective instructions, participants
were administered the two memory and fluency measures according to directions pro-
vided in respective test manuals. After testing, participants completed a postexperimen-
tal questionnaire that asked them to briefly paraphrase the instructions provided at the
beginning of the experiment, to indicate whether they followed these, and to indicate
their strategy. Participants were randomly assigned to each condition and the experi-
menter was blind to each participant’s condition.

Participants in both conditions were tested approximately 3 weeks after the initial
testing session and were provided the same instructions. After the second testing, partic-
ipants again completed the postexperimental questionnaire and were debriefed. All par-
ticipants were paid $10 regardless of how well they complied with instructions.

RESULTS

Test-Retest Analyses

To address the hypotheses that controls would perform better than malingering simu-
lators and that they would demonstrate improvement across testing sessions, we per-
formed two mixed-model multivariate analyses of variance (MANOVAs) with one be-
tween-subjects factor, condition (control and malingering), and one within-subjects
factor, time (time 1 and time 2). For one MANOVA, dependent variables were the ver-
bal measures, including CVLT variables that assess quantitative aspects of performance
(total words learned across the five trials, short-delay recall, and long-delay recall) and
the COWAT. Because the consistency variable assesses a performance pattern, it was
evaluated separately. A second MANOVA had the same between and within-subjects
factors, but with dependent measures of the nonverbal variables—immediate and de-
layed recall on the RCF and the RFFT. For each MANOVA, if the multivariate omni-
bus test was significant, univariate analyses of variance (ANOVAs) followed by Tukey’s
post-hoc HSD tests with an alpha level of .05 were performed. See Table 1 for means
and standard deviations of all variables.

For the verbal MANOVA, the multivariate effect of condition was significant, Wilks’
lambda, $F(4, 37) = 18.16, p < .0001$, with significant univariate $F$s ($p < .0001$) for each
dependent variable. In each analysis, the controls performed significantly better than the
malingering simulators. The multivariate effect of time was significant, Wilks’ lambda,
$F(4, 37) = 5.86, p < .001$, with the following significant univariate effects: total words
learned ($p < .0001$), short-delay recall ($p < .006$), long-delay recall ($p < .04$), and
COWAT ($p < .001$). For each test, performance was significantly better at time 2. The
condition $\times$ time interaction was significant, Wilks’ lambda $F(4, 37) = 4.86, p < .003$.
Univariate $F$s for the total words learned ($p < .001$), short-delay recall ($p < .026$), long-
delay recall ($p < .008$), and the COWAT ($p < .008$) were significant. Tukey’s post-hoc
testing indicated that these interaction effects were significant due to the improvement
of the controls, but not simulators, across time. For the nonverbal MANOVA, the multi-
variate effect of condition was significant, Wilks’ lambda, $F(3, 38) = 12.46, p < .0001$,
with significant univariate $F$s ($p < .008$) for each dependent variable. In each analysis,
the controls performed significantly better than the malingering simulators. The multi-
variate effect of time was significant, Wilks’ lambda, $F(3, 38) = 20.85, p < .0001$, with
significant univariate effects $F$s ($p < .0001$) for each dependent variable indicating signif-
icantly better performance at time 2 versus time 1. Finally, the condition $\times$ time interac-
tion was not significant, Wilks’ lambda $F(3, 38) = 1.46, p < .24$. In summary, these anal-
yses indicate that (a) controls performed significantly better than simulators on verbal
and nonverbal measures at both times, (b) both groups improved on nonverbal mea-
sures, and (c) only the controls improved on the verbal measures.

We next conducted classification analyses based on the change scores (i.e., practice
effects) across the two testing sessions. Due to the small participant-to-variable ratio we

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Means (M) and Standard Deviations (SD) for Malingering Simulators and Controls at Times 1 and 2</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Simulators</th>
<th></th>
<th>Controls</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
<td>SD</td>
<td>Time 2</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Verbal measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>35.9</td>
<td>14.8</td>
<td>36.0</td>
<td>16.1</td>
</tr>
<tr>
<td>SDFR</td>
<td>5.7</td>
<td>3.9</td>
<td>5.9</td>
<td>4.5</td>
</tr>
<tr>
<td>LDFR</td>
<td>5.3</td>
<td>4.1</td>
<td>5.0</td>
<td>3.9</td>
</tr>
<tr>
<td>COWAT</td>
<td>29.5</td>
<td>8.0</td>
<td>30.2</td>
<td>7.3</td>
</tr>
<tr>
<td><strong>Nonverbal measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMRECAL</td>
<td>13.9</td>
<td>6.9</td>
<td>17.3</td>
<td>7.2</td>
</tr>
<tr>
<td>DLRECAL</td>
<td>13.0</td>
<td>6.7</td>
<td>16.4</td>
<td>7.3</td>
</tr>
<tr>
<td>RFFT</td>
<td>63.2</td>
<td>17.8</td>
<td>71.5</td>
<td>22.1</td>
</tr>
</tbody>
</table>

COWAT = Controlled Oral Word Association Test; DLRECAL = Rey Complex Figure delayed recall; IMRECAL = Rey Complex Figure immediate recall; LDFR = California Verbal Learning Test (CVLT) long-delay free recall; RFFT = Ruff Figural Fluency Test; SDFR = CVLT short-delay free recall; Total = CVLT total number of words re-
called across the 5 learning trails.
did not use multivariate classification techniques (Stevens, 1996), but rather determined
cutoffs from the distribution of change scores for each variable individually and com-
puted sensitivity and specificity. Sensitivity was the percentage of malingering simulators
correctly classified as malingering; specificity was the percentage of controls correctly
classified as not malingering. Cutoffs based on the change scores for the verbal variables
achieved the following: for total words learned (CVLT trials 1–5), sensitivity was 71% and
specificity was 71%; for change in CVLT short-delay recall, sensitivity was 67% and
specificity was 71%; for change in CVLT long-delay recall, sensitivity was 81% and spec-
ificity was 62%; for the COWAT, sensitivity was 76% and specificity was 57%. Cutoffs
based on change scores for the nonverbal variables yielded the following: for RCF im-
mediate recall, sensitivity was 57% and specificity was 52%; for RCF long-delay recall,
sensitivity was 57% and specificity was 43%; for the RFFT, sensitivity was 67% and
specificity was 62%. Overall, the classification rates for verbal material are significantly
better than nonverbal material and above chance levels of classification.

Consistency Analyses

To address hypotheses that controls would demonstrate greater consistency than the
simulators within and across testing sessions, we conducted a series of correlational anal-
yses. We compared these correlations with Fisher’s $r$-to-$z$ transformations with a Bon-
ferroni correction to control for Type I error rate. In our first analysis, test-retest corre-
lations were computed for the verbal and nonverbal variables between times 1 and 2 for
both groups. As is evident in Table 2, malingering simulators and controls displayed high
and roughly similar levels of consistency. Inferential analyses of these correlations based
on Fisher’s $r$-to-$z$ transformations with a Bonferroni correction confirmed the lack of
statistically significant differences. Next, we computed correlations between redundant

<table>
<thead>
<tr>
<th>Variable</th>
<th>Malingering</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.73</td>
<td>.77</td>
</tr>
<tr>
<td>SDFR</td>
<td>.73</td>
<td>.62</td>
</tr>
<tr>
<td>LDFR</td>
<td>.76</td>
<td>.75</td>
</tr>
<tr>
<td>COWAT</td>
<td>.82</td>
<td>.83</td>
</tr>
<tr>
<td>Nonverbal measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMRECAL</td>
<td>.79</td>
<td>.76</td>
</tr>
<tr>
<td>DLRECAL</td>
<td>.74</td>
<td>.72</td>
</tr>
<tr>
<td>RFFT</td>
<td>.76</td>
<td>.88</td>
</tr>
</tbody>
</table>

COWAT = Controlled Oral Word Association Test; DLRECAL = Rey Complex Figure (RCF) delayed recall; IMRECAL = RCF immediate recall; LDFR = California Verbal Learning Test (CVLT) long-delay free recall; RFFT = Ruff Figural Fluency Test; SDFR = CVLT short-delay free recall; Total = CVLT total number of words recalled across the five learning trials.

All correlations are significant at $p < .003$.

None of the correlations between malingering and control conditions were significant. See text for statistical methodology.
variables (e.g., CVLT short- and long-delay) within times 1 and 2, respectively. As is evident in Table 3, malingering simulators and controls displayed roughly similar levels of consistency at both times on these measures. Fisher’s r-to-z transformation with Bonferroni correction demonstrated only one significant difference—the correlation between short-delay free recall and total words recalled at time 2. Contrary to predictions, these correlational analyses indicate that when quantitative levels of performance are assessed, malingering simulators and controls perform with similar levels of consistency between and within sessions, for the most part.

Our final hypothesis was that controls would perform more consistently (i.e., recalling the same word on successive trials) across the five CVLT learning trials than malingering simulators. As predicted, controls responded significantly more consistently ($M = 87.7\%, SD = 9.4$) than the simulators ($M = 74.6\%, SD = 15.5$), $t(40) = 3.31, p = .002$, at time 1. Similarly, controls were also more consistent at time 2 ($M = 92.4\%, SD = 7.2$, simulators $= 79.0\%, SD = 18.8$), $t(40) = 3.06, p = .004$. This finding indicates that when a qualitative analysis of consistency is made, controls perform more consistently than malingering simulators.

**DISCUSSION**

Our results indicate that malingering simulators perform more poorly than controls, but as consistently, when consistency is assessed via quantitative levels of performance. However, when assessed qualitatively, malingering simulators perform more inconsistently than controls. The pattern of improvement for both groups also differed, as controls improved across time on both verbal and nonverbal measures, but malingering simulators improved on only the nonverbal measures.

This differential pattern of improvement on verbal and nonverbal tasks for controls and malingering simulators was unexpected. This difference may be due to the fact that the nonverbal tasks differ in the following ways from the verbal tasks: They have a motor component, are relatively novel, particularly the RFFT, and tap more implicit types of information processing (see Tombaugh, Faulkner, & Hubley, 1992). Because they require less explicit information processing than the verbal tasks, they are probably more
difficult to verbalize, thus making development of a sophisticated malingering strategy more difficult. While this does not translate to more difficulty malingering these tests at each time, it may be why malingering simulators demonstrate practice effects on them at retest. Specifically, simulators may assume that neurologically damaged patients do not improve with repeat testing and thus may attempt to perform as poorly on the second testing as they did on initial testing. If so, this may be easier with explicit verbal tasks. Initial support of this position is evident from open-ended responses to queries about response strategy. While we did not obtain enough responses to compare malingering strategies for verbal and nonverbal tests, participants were more likely to at least mention the verbal tests. At time 1, five participants mentioned malingering strategies on the CVLT and four on the RCF, whereas at time 2, eight participants mentioned the CVLT and only two the RCF. The COWAT and RUFF were each mentioned by only one participant. That participants were more likely to indicate a malingering strategy for verbal tests, particularly at time 2, suggests that these tests tend to be more explicit and easier to verbalize than the nonverbal tests. It is our contention that this property of the verbal tests resulted in simulators inhibiting the tendency to improve over a relatively short retest interval. Overall, while past research has obtained mixed results using implicit tests to detect malingering (see Davis, King, Klebe, Bajszar, Bloodworth, & Wallick, 1997; Wiggins & Brandt, 1988), our data suggest that test-retest effects in malingering may vary based on the task’s material specificity (verbal vs. nonverbal) and the type of processing necessary to compete it. This finding, of course, needs replication with further participants and measures.

We also found high and equivalent levels of consistency for the malingering simulators and controls within and across time on quantitative variables. The generally high consistency within each time of the short-delay, long-delay, and total words across the five trials is similar to findings by Trueblood (1994) who found that malingerers performed consistently on several verbal memory measures within a testing session. Paul et al. (1992) also found high levels of test-retest consistency in malingerers on specific malingering tests. Taken together, these findings suggest that when quantitative aspects of performance are assessed, malingerers tend to perform more poorly than controls, but as consistently. In other words, they do not generate variable and/or nonsensical quantitative patterns of performance (e.g., poor short-delay recall, but good long-delay recall). Despite this robust finding, qualitative or performance pattern analysis of data indicate that malingering simulators demonstrate a less consistent recall pattern than controls across the five CVLT learning trials. This finding may simply reflect a tendency to randomly or indiscriminately recall items and to not “build on” previously recalled items. Overall, our findings suggest that (a) different results can be obtained when quantitative or qualitative approaches are used to assess consistency and (b) fine-grained analysis of consistency, perhaps even at the level of test item, may prove most useful when examining response consistency within or across sessions.

One caveat must be made about these consistency findings. This study only addressed the consistency of neuropsychological test data and not other data points typically thought to be important for assessing consistency. Factors such as premorbid intelligence, severity and location of brain injury, patient history, effect of apparent brain injury on daily life, etc. were not addressed, nor could they be, in such an analog experiment. For instance, inconsistency and potential malingering is suggested when an individual has a minor, uncomplicated head injury, yet demonstrates significant and debilitating behavioral and cognitive deficits (see first case presented in Reitan & Wolfson, 1998, pp. 177–183). Clinically, many more extra-test variables are considered than presented in this study, as relationships between these different data points, in conjunction
with test data, is typically used to determine the validity of a test protocol. Given these multiple data points, Williams (1998) notes the substantial empirical challenges in moving malingering detection from a primary clinical to an actuarial or regression-based model. If this is to be achieved, considerably more research is required to develop a model in which, for example, expected cognitive outcomes of a variety of brain injuries can be determined. Presumably, at that point, this information can be used to empirically detect malingered performance. While the ability to do this with any degree of certainty is far off and probably not of immediate clinical benefit, it does suggest that future simulation research should include more variables than just test performance to approximate the real-life conditions in which clinicians think through malingering issues.

There are two limitations of this study. First, the use of undergraduates raises the question of generalizability to clinical samples of suspected malingerers. While an analogue study cannot obviously replicate the complex real-world situation in which suspected malingerers are evaluated, some research suggests that findings with undergraduates may be externally valid. For instance, Iverson and Franzen (1994) found that undergraduates and prison inmates performed similarly when instructed to mangle on several neuropsychological tests. Mittenberg, Theroux-Fichera, Zielinski and Heilbronner (1995) accurately applied decision rules obtained with undergraduates on the WAIS-R to published data on genuine malingerers. Analog studies with undergraduates may thus generalize to real malingerers, though future comprehension of malingering is best advanced by a convergence of experimental and clinical research efforts (Rogers et al., 1993). Second, analysis of response consistency is probably best done with malingerers (whether real or simulating) and head-injured patients, given that fluctuating attention and concentration is a common cognitive symptom in patients who have suffered such trauma. Their performance, even when they exert maximal effort and clearly do not mangle, is thus likely to be less consistent than individuals without brain damage. Rather than non-brain–damaged controls, future research on response consistency should use actual brain-damaged patients. More broadly, given the frequency of repeat evaluation in forensic neuropsychology, further examination of test-retest and performance reliability issues on a wide range of neuropsychological measures should be areas of future research.

**REFERENCES**


