Improvements in the M Test as a Screening Measure for Malingering

Richard Rogers, PhD; R. Michael Bagby, PhD; and J. Roy Gillis, PhD

Recent studies of the M have failed to confirm its effectiveness as a screening measure for malingering. The present study involved a further analysis of 99 cases reported by Gillis et al. and the construction of Rule-Out and Rule-In scales. Both scales evidence excellent internal reliability, and their combined use was effective in accurately screening >80 percent of potential malingerers.

Recent curtailments in social and mental health services have increased the likelihood that relatively unimpaired individuals will feign mental illness in order to access psychiatric hospitalization and the benefits accrued from patient status. Although important advances have been made in the clinical assessment of malingering, the emphasis has been on the development of comprehensive assessment procedures for those suspected of malingering. What is conspicuously absent are rapid and effective screening measures suitable for general psychiatric settings.

Beaber and his associates constructed the M test, a 33 item true-false scale as a screening measure for the detection of feigned schizophrenia. The M test is comprised of three scales: the Confusion (C) scale to measure highly atypical attitudes not associated with mental illness; the Schizophrenia (S) scale to assess common symptoms of schizophrenic disorders; and Malingering (M) scale to examine bizarre and unusual symptoms that, by definition, are unlikely to be genuine. Through a series of chi-squared analyses, Beaber and his colleagues were able to classify 78.2 percent of male students feigning mental illness and 87.2 percent of VA inpatients. Important limitations of the study were the use of males subjects only, faking a single diagnosis (only schizophrenia), and the lack of actual malingerers (only student feigners).

Smith investigated the discriminability of the M test on 23 malingerers and 62 inmates referred for a forensic evaluation. In attempting to cross-validate Beaber's findings, he was able to detect only 69.6 percent of the malingerers and
66.1 percent of psychiatrically disordered inmates. These modest results led Smith et al. to question the practical usefulness of the M test as a screening instrument.

Gillis, Rogers, and Bagby mounted an ambitious study of the M test with male and female subjects from a variety of settings (inpatient, outpatient, correctional, community, and university) with a total of 169 subjects under honest instructions, 124 simulators (i.e., subjects feigning mental illness), and 25 suspected malingers. Malingers were classified on the basis of a thorough inpatient evaluation by an clinician independent of the research; they are labeled as “suspected” to reflect the lack of explicit criteria in DSM-III-R. Factor analysis of the M test items (principal components, varimax rotation) produced a three factor solution with close correspondence to the C, S, and M scales. In addition, all three scales had good internal reliabilities with KR-20s ranging from .87 to .93. However, the performance of the Beaber et al. criteria in discriminating between malingers and patient controls was disappointing. Although 92.3 percent of subjects under honest instructions and 79.8 percent of those feigning mental illness were accurately classified, only a minority (40.0%) of malingers were correctly identified. Attempts to improve the M test’s accuracy through a stepwise discriminant analysis proved unsuccessful. In light of these disappointing classification rates, Gillis et al. concluded the the M test, at least in its current form, had limited usefulness as a screen for malingering.

The purpose of the present study was to provide a further analysis of the Gillis et al. data and the construction of new scales in order to improve the effectiveness of the M test as a screening measure. Toward this end, we were particularly interested in minimizing missed classifications of potential malingers because their exclusion on the basis of the M test would likely preclude further investigations.

**Method**

The psychiatric sample combined outpatient (N = 33) and inpatient (N = 39) groups drawn from the Gillis et al. study. These patients were a diagnostically heterogeneous group (e.g., primarily schizophrenic, delusional, mood, organic, and personality disorders). Patients were predominantly male (45 or 62.5%) with a mean age 35.8 ± 10.2 years and 14.1 ± 3.6 years of education. The suspected malingers (N = 25) were almost entirely male (24 or 96.0%), younger (28.3 ± 10.3; t(98) = 3.26, p < .002), and less educated (9.2 ± 2.6; t(98) = 6.21, p < .001), than the psychiatric patient sample.

All subjects were administered the M test under honest instructions. Inpatients and suspected malingers were invited to participate in the research project; outpatients were remunerated for their time and transportation ($20 per subject) for completion of the M test and other measures (i.e., SIRS and MMPI).

Initial inspection of M test items suggested that most did not discriminate between malingers and patients. We
computed the positive predictive power (PPP) and the negative predictive power (NPP) for each M test item. Since very few items were efficient on both PPP and NPP, we developed two separate scales to be used in a two-step procedure: a Rule-Out scale (i.e., eliminate patients) comprised of 10 items with the highest NPP values and a Rule-In scale (i.e., identify potential malingerers) comprised of 10 items with the highest PPP values. The estimates of internal reliability (KR-20 coefficients) were excellent for both Rule-Out ($r = .85$) and Rule-In ($r = .87$) scales.

**Results and Discussion**

Optimum cutting scores were derived from a cross-tabulation of Rule-Out and Rule-In scale scores for malingerers and patients (see Table 1). Clinicians who wish to employ the M test are provided with two options: Option A may be more useful in general clinical practice because it eliminates most bona fide patients (83.8%) from further assessment but retains four-fifths of potential malingerers for further evaluation. Option B may be more applicable in forensic settings since it retains nearly all potential malingerers (95.2%) and screens out somewhat fewer (approximately 2/3) of bona fide patients. In either case, the M test can be administered in 10 to 15 minutes and scored on the Rule-Out and Rule-In scales in 1 to 2 minutes as a time-efficient screen for malingering.

Results from recent studies underscore the importance of validating screening measures under “real-world” conditions. Simulation studies that ask unimpaired subjects to feign mental illness simply do not capture the interpersonal and situational variables confronting a would-be malingerer. Results of this study would suggest that a modified scoring (Rule-Out and Rule-In scales) of the M test can provide a rapid and effective method of screening out most patients for whom there is very little likelihood of malingering. For the remaining patients, comprehensive assessments of possible malingering should employ both psychometric measures, such as the MMPI, as well as structured interviews, such as the SIRS.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Effectiveness of Rule-Out and Rule-In Scales for Identifying Potential Malingerers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scales</td>
<td>KR-20</td>
</tr>
<tr>
<td>Rule-Out</td>
<td>.85</td>
</tr>
<tr>
<td>Rule-In</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Screening Out Patients**

Option A: To screen out patients, a Rule-Out score of <4 (i.e., actual patients typically endorse less than four of these symptoms) followed by Rule-In score on Option A (<2) eliminated from further consideration 83.8% (57 of 68) of the patients and retained 81.0% (17 of 21) of the potential malingerers for further assessment.

Option B: To screen out patients, the same Rule-Out criterion (<4) followed by Rule-In on Option B (0) eliminated 70.6% (48 of 68) of the patients and retained 95.2% (20 of 21) of the malingerers.

1 Because of missing data on six subjects, the final patient sample was 68.
2 Because of missing data on four subjects, the final suspected malingering group was 21.
SIRS in particular would suggest that it has a high level of discriminability with both simulators and suspected malingers.11

References