Psychological Assessment
2010, Vol. 22, No. 1, 50–56

1040-3590/10/$12.00 DOI: 10.1037/a0016585

Detection of Malingered Mental Retardation

Anne L. Shandera, David T. R. Berry, Jessica A. Clark, Lindsey J. Schipper,
Lili O. Graue, and Jordan P. Harp
University of Kentucky

In a cross-validation of results from L. O. Graue et al. (2007), standard psychological assessment instruments, as well as tests of neurocognitive and psychiatric feigning, were administered under standard instructions to 24 participants diagnosed with mild mental retardation (MR) and 10 demographically matched community volunteers (CVH). A 2nd group of 25 community volunteers was instructed to malinger MR (CVM) during testing. CVM participants obtained Wechsler Adult Intelligence Scale (3rd ed.; D. Wechsler, 1997) Full Scale Intelligence Quotient scores that were significantly lower than the demographically similar CVH group but comparable to the MR group, suggesting that CVM subjects feigned cognitive impairment. On the basis of standard cutting scores from test manuals or published articles, of the 11 feigning measures administered, only the Test of Memory Malingering (TOMM; T. N. Tombaugh, 1996) retention trial had a specificity rate >.90 in the MR group. However, the 2nd learning trial of the TOMM, as well as a short form of the Digit Memory Test (T. J. Guilmette, K. J. Hart, A. J. Guiliano, & B. E. Leininger, 1994), approached this level of specificity, with both at .88. These results raise concerns about the specificity rates at recommended cutting scores of commonly used feigning tests in defendants with MR.

Keywords: malingering, mental retardation, neuropsychological assessment, feigned neurocognitive impairment, intellectual disability

Psychologists conducting evaluations in forensic settings must address the possibility of malingered symptoms using objective procedures (Otto, 2008). Complicating this task, research over the past two decades has suggested that feigning may take many forms, including false psychiatric, neuropsychological, and somatic complaints (Granacher & Berry, 2008). Fortunately there has been intense research aimed at validating malingering tests in each of these domains (Larrabee, 1998; Rogers, Bagby, & Dickens, 1992; Vickery, Berry, Inman, Harris, & Orey, 2001). More recently, a Supreme Court ruling that executing mentally retarded capital case defendants is unconstitutional (*Atkins v. Virginia*, 2002) has sparked interest in detecting feigning of this disorder (Victor & Boone, 2007).

Because there is a rich body of work on identifying feigned cognitive symptoms following traumatic brain injury (Boone, 2007; Larrabee, 2007), one approach to detecting malingered mental retardation (MR) has been to generalize procedures from the traumatic brain injury literature. An important strand of this effort has involved examining specificity rates (percentage of honest test takers correctly classified by the test) for existing malingering-detection instruments in populations with low IQ scores resulting from diverse etiologies. Dean, Victor, Boone, and Arnold (2008) reported compromised specificity rates for most malingering measures in noncompensation-seeking neuropsychological evaluees with low IQs, and performance on most feigning indicators declined steadily with IQ. Dean, Victor, Boone, Philpott, and Hess (in press) examined the specificity of multiple feigning indicators in a large sample of noncompensation-seeking dementia patients and also found high false positive rates for most of these procedures. Although these studies are valuable in documenting compromised performance of malingering procedures in severely brain-damaged subjects with low IQs, few of the participants had been diagnosed with MR, raising possible questions about the generalizability of findings to the latter disorder.

Prior to the *Atkins v. Virginia* decision, relatively little published work focused on the specificity of malingering measures in mentally retarded populations. Brockhaus and Merten (2004) described good performance by mentally retarded individuals on the Word Memory Test (WMT; Green, 2005). However, Marshall and Happe (2007; see below) reported a personal communication from Brockhaus indicating that the mentally retarded patients in the WMT study were selected to be able to understand and perform on cognitive tests and to be sufficiently emotionally stable to complete testing. Marshall and Happe therefore raised questions about the generalizability of the Brockhaus and Merten (2004) results. In a meta-analytic review of available literature on the Rey 15-Item

Anne L. Shandera, David T. R. Berry, Jessica A. Clark, Lindsey J. Schipper, Lili O. Graue, and Jordan P. Harp, Department of Psychology, University of Kentucky.

This article is based on a thesis submitted by Anne L. Shandera in partial fulfillment of the requirements for a Master's degree in Psychology at the University of Kentucky. Support for this study by a research grant from the College of Arts & Sciences, University of Kentucky, is gratefully acknowledged. We thank the Bluegrass Regional Mental Health/Mental Retardation Board (BGRMH/MRB) for assistance in this study. Opinions expressed here do not represent the position of the BGRMH/MRB. David T. R. Berry holds the copyright to the Letter Memory Test. All proceeds from the Letter Memory Test are donated to the Harris Psychological Services Clinic.

Correspondence concerning this article should be addressed to David T. R. Berry, Department of Psychology, University of Kentucky, Lexington, KY 40506-0044. E-mail: dtrb85@gmail.com

Test, Reznek (2005) showed that this procedure had poor specificity in mentally retarded samples. Hurley and Deal (2006) administered tests sensitive to feigning of psychiatric as well as cognitive deficits to mild and moderately mentally retarded individuals, and reported high false-positive rates on most procedures. In contrast, Simon (2007) administered the Test of Memory Malingering (TOMM; Tombaugh, 1996) to mentally retarded forensic evaluees and found high specificity rates. Marshall and Happe (2007) reported data on malingering measures collected from mild to moderately mentally retarded individuals evaluated in noncompensation-seeking neuropsychological evaluations. They found poor specificity rates for most of the indices evaluated. Thus, the published data on malingering test performances from mentally retarded individuals are somewhat mixed, although several reports raise concerns about elevated false-positive rates in this population, and some recommend downward adjustment of cutting scores to mitigate this problem. One limitation of studies on the specificity rates of malingering-detection procedures in mentally retarded individuals is that it is not possible, on this basis alone, to ascertain the impact on sensitivity (percentage of malingering individuals correctly classified by the test) of adjusting cutting scores downward.

A few published studies have included both honestly responding patients with MR and individuals known or suspected to be malingering intellectual deficits. Schretlen and Arkowitz (1990) found that a modified form of the Wechsler Adult Intelligence Scale Information subtest, as well as certain error types on the Bender-Gestalt, contributed to a discriminant function that correctly classified 19 out of 20 severely mentally retarded participants (*M* Full Scale Intelligence Quotient [FSIQ] = 47) but only 1 out of 20 prison inmates faking the condition. Hayes, Hale, and Gouvier (1997) reported that the Rey 15-Item Test, Rey Dot Counting Test, and the M-Test were ineffective at distinguishing non-malingering mentally retarded forensic patients from similar patients thought to be feigning. Hayes, Hale, and Gouvier (1998), apparently using a subset of patients from their previous study, found that the Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992) discriminated mentally retarded forensic evaluees found not guilty by reason of insanity from mentally retarded patients thought to be malingering with 95% accuracy. However, it is not clear that the malingering participants were feigning cognitive deficits exclusively, as they were said to have endorsed "significant psychiatric complaints; one endorsed comorbid complaints of memory dysfunction" (p. 34), and the SIRS has not been validated to detect false cognitive deficits (Rogers et al., 1992). Chafetz, Abrahams, and Kohlmaier (2007) reported a validation study of feigning predictors internal to the Psychological Consultive Examination used for disability determinations in mentally retarded individuals. On the basis of the TOMM, WMT, and the Psychological Consultive Examination indicators, they found that approximately 50%–60% of compensation-seeking adults appeared to be feigning deficits. The Chafetz et al. (2007) study is of particular interest because, at present, it provides the only published, objectively based estimate of base rates of feigned MR in compensation-seeking circumstances. Graue et al. (2007) administered a large battery of tests to individuals with mild MR as well as demographically matched community volunteers instructed to feign MR or answer honestly. In addition to multiple malingering measures, Graue et al. also gave intellectual and neuropsychological tests to approximate clinical practice in which feigning procedures are typically embedded in a large battery. Consistent with much of the previous literature, Graue et al. found significantly compromised specificity rates for most malingering measures. Alternative cutting scores that maintained specificity at 95% or greater in the sample of mentally retarded individuals were identified, although Graue et al. cautioned that these needed to be cross-validated in new samples before being applied in clinical settings.

The present study expands on the Graue et al. (2007) work using similar methodology but adding malingering measures and new samples. The goals of the present work were to cross-validate previous findings and investigate the potential utility of alternative measures, including both neurocognitive and psychiatric feigning tests.

Method

Participants

Participants included 24 adults with previous diagnoses of mild MR and 35 demographically matched community volunteers. None of these subjects had been included in the Graue et al. (2007) study. The MR group was recruited through the Bluegrass Regional Mental Health/Mental Retardation Board (BGRMH/MRB), which provides a variety of services for developmentally disabled adults. MR participants were required to have a *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM–IV–TR*; American Psychiatric Association, 2000) diagnosis of mild MR and to have no comorbid substance abuse, psychotic or neurologic diagnoses. Given the high rate of comorbid psychiatric disorders found in individuals with MR (American Psychiatric Association, 2000; p. 45), other coexisting psychological conditions, such as anxiety, depressive disorders, and adjustment disorders, were not a basis for exclusion. Community volunteer (CV) participants were recruited through flyers posted at local unemployment offices and clinic waiting rooms. They were screened to have completed no more than 11 years of education and to have no substance abuse, neurologic, or psychotic diagnoses. CV participants were randomly assigned to the honest responding (CVH; n = 10) or malingered MR (CVM; n = 25) groups. Because the CVH group was included primarily as an experimental manipulation check, only a small number of CV participants were assigned to this condition.

Materials

Nontest materials included Institutional Review Board-approved informed consent documents, a demographics questionnaire, a "fact sheet" on MR, a quiz on the MR fact sheet, a debriefing questionnaire and group-specific instruction sets. All of these nontest materials were previously used in the Graue et al. (2007) study. MR and CVH instructions indicated that participants were to try their hardest on all tests and that they would receive payment at the end of the testing (MR \$20; CVH \$75; CVH participants took longer to test and provided their own transportation to the study site). Both CVH and CVM groups were warned not to reveal their instructions to the examiner. CVM participants were asked to imagine that they were being evaluated for MR in a death penalty case and to perform as if they had this condition without being detected. They were offered a bonus of $20 if they successfully feigned without being detected by the tests or examiner, for a total payment of $95. CVM participants also read the fact sheet on MR and completed a quiz on the symptoms. All CV participants also answered a debriefing form at the end of the study to check for understanding of, and compliance with, their instructions.

Standard psychological tests included the Wechsler Test of Adult Reading (WTAR; Psychological Corporation, 2001) as well as the Wechsler Adult Intelligence Scale (3rd ed.; WAIS–III; Wechsler, 1997). In addition to feigning indices derived from the WAIS–III (see below), neurocognitive malingering tests included the TOMM (Tombaugh, 1996), the Letter Memory Test (LMT; Inman et al., 1998), and a 36-item short version of the Digit Memory Test (DMT; Guilmette, Hart, Guiliano, & Leininger, 1994), all of which were used by Graue et al. (2007). We added the b Test (Boone et al., 2000) and an oral presentation form of the WMT (Green, 2005) to expand coverage of this domain. Finally, feigning of psychiatric symptoms was assessed using the SIRS, chosen because of the promising report of Hayes et al. (1998) noted earlier. Several other tests were given to all groups as well, but because of space considerations, their results are not reported here.

Procedure

With the exception of changes in some of the tests administered, procedures followed in the present study were very similar to those described by Graue et al. (2007). After obtaining appropriate Institutional Review Board approvals, we approached BGRMH/MRB case managers for individuals with MR with study details and recruitment criteria. Case managers explained the study in broad terms to eligible clients and notified study personnel about those who were interested in participating. A research assistant (RA) met with identified clients at a mutually convenient BGRMH/MRB facility, explained the study, obtained consent and permission to review medical records, read aloud items from the demographic questionnaire, and recorded responses. After reviewing the instructions, the RA administered the test battery in counterbalanced order using standard procedures. Breaks were allowed as needed, and at the conclusion of testing, clients were thanked, paid, and escorted back to their case manager.

Community volunteers responded to recruitment flyers at various locations in an urban area and contacted Anne L. Shandera, who screened for study criteria described above. Those who met criteria were scheduled for a meeting at the Clinical Psychology Doctoral Training Program's clinic. Two RAs worked with each participant. RA1 met with each participant, obtained informed consent, supervised completion of questionnaires and the WTAR given under standard instructions, and provided the participant with an envelope including instructions for either the honest (CVH) or malingering (CVM) group. For CVH participants, RA1 answered any questions the participant had, cautioned him or her not to reveal instructions to RA2, and then escorted the participant to the testing room where RA2 waited. For CVM participants, RA1 also provided the MR fact sheet, allowed time for review and questions, and then administered the MR quiz. Next, CVM participants were cautioned not to reveal their instructions to RA2 and were then escorted to the testing room where RA2 waited. RA2, blind to participant instructions, then administered the test battery. Breaks were allowed as needed. At the completion of testing, RA2 escorted the participant back to RA1, who administered the debriefing form and paid and thanked participants.

Results

All CV participants were able to reiterate their instructions on the debriefing form and indicated that they had followed them as they completed testing. For present purposes, two group comparisons were of most interest: CVH versus CVM, which captures the extent of malingered performance in the CVM group, and MR versus CVM, which contrasts genuine and feigned MR. Table 1

Table 1
Demographic and Background Characteristics of Groups

Variable	CVH (n = 10)	CVM (n = 25)	MR (n = 24)	CVH vs. CVM (t/χ^2)	MR vs. CVM (t/χ^2)
Age (years)					
M	33.80	33.84	36.04	0.01[a]	0.75[b]
SD	14.34	10.43	9.67		
Education (years)					
M	9.90	10.20	11.22	0.80[a]	2.96**[b]
SD	1.10	0.96	1.51		
Developmental disability (%)	0	0	100	—	49.00***[c]
Male (%)	40	56	58	0.73[c]	0.03[c]
Caucasian (%)	50	40	71	0.97[d]	5.35[d]
Right-handed (%)	90	80	79	0.50[c]	1.12[d]
Misdemeanor history (%)	50	20	8	3.15[c]	1.36[c]
Felony history (%)	10	24	8	0.88[c]	2.20[c]
WTAR predicted IQ					
M	86.20	83.24	—	0.81[a]	—
SD	6.93	10.60	—		

Note. CVH = community volunteer honest; CVM = community volunteer malingering; MR = mentally retarded; WTAR = Wechsler Test of Adult Reading (Psychological Corporation, 2001).
[a] $df = 33$. [b] $df = 47$. [c] $df = 1$. [d] $df = 2$.
** $p < .01$. *** $p < .001$.

provides background data on the groups. There were no significant demographic differences between the groups being compared, with the exception of more years of education completed by the MR group (all of it special education) and the presence of developmental disability, which was true for all of the MR and none of the CV participants. Table 1 also indicates nontrivial rates of misdemeanor and felony convictions in the CV groups. Finally, the mean WTAR-estimated WAIS–III FSIQs were comparable in the two CV groups who took this test under standard instructions, suggesting that they were similar in intellectual abilities.

Table 2 provides WAIS–III IQ results from the three groups. Here it can be seen that the obtained FSIQ for the CVH group was very close to that estimated using the WTAR procedure (84.8 vs. 86.2), suggesting that the latter test provided a reasonable estimate of IQ in the present samples. It is also clear that the CVM group significantly depressed its obtained WAIS–III FSIQ relative to the estimate generated by the WTAR taken under standard instructions (67.8 vs. 83.2). Finally, the CVM WAIS–III FSIQ was also significantly lower than that of the CVH group (67.8 vs. 84.8). Taken together, these results suggest that the experimental manipulation was effective and that the CVM group malingered lower cognitive abilities on the WAIS–III.

Table 2 also shows that the CVM group significantly depressed all three IQ scores relative to the CVH group, with Cohen's *d* scores, calculated as (CVH − CVM)/pooled *SD*, ranging from 1.48 to 1.91. However, the CVM group approached, but did not quite equal, the intellectual deficits found in the MR group on the three IQ scores, with *d* scores of −.21 to −.82, calculated as (MR − CVM)/pooled *SD*. Nevertheless, all three mean IQ scores for the CVM group fell in a range consistent with mild MR.

Table 3 presents results from indices and procedures intended to detect feigned impairment, along with the standard cutting scores recommended for each. Four indices may be derived from the WAIS–III to detect feigning, including the Age Corrected Digit Span Subtest Scale Score (Babikian, Boone, Lu, & Arnold, 2006), Vocabulary–Digit Span (Iverson & Tulsky, 2003), the Mittenberg Discriminant Function (Mittenberg et al., 2001), and Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994). Only the Digit Span Subtest Score and Reliable Digit Span generated significant differences between the CVM and CVH groups, and only RDS produced a statistically significant difference in the CVM versus MR contrast. However, examination of mean scores shows that the latter effect is in an unexpected direction: The MR group performed worse than the CVM group. On the basis of standard cutting scores, specificity rates for the WAIS–III feigning indices in the MR group were very low for three of these indices. The Vocabulary−Digit Span specificity rate was excellent, although its sensitivity was essentially nil.

Table 3 also presents results from several procedures dedicated to detecting feigned neurocognitive impairment. The contrast of CVH and CVM consistently produced statistically significant differences suggestive of feigning in the latter group, indicating that these procedures reflect malingered deficits in individuals with low average to borderline FSIQs. However, only a subset of these procedures produced statistically significant differences in the crucial CVM versus MR contrast, including the Digit Memory Test, the Test of Memory Malingering second learning and retention trials, and the b Test E score. When we applied a commonly used "rule of thumb" minimum desired specificity rate of .90 in MR patients, only the TOMM retention trial met this threshold. However, the DMT and TOMM second learning trial closely approached this level, with specificity rates of .88 for both. Specificity rates for the remaining procedures were clearly unacceptable, ranging from a low of .25 for WMT Consistency to a high of .69 for the b Test error score. Finally, the SIRS, a measure of psychiatric feigning, fared poorly as well.

Graue et al. (2007) identified modified (lower) cutting scores for three tests that produced .95 or higher specificity rates in their sample of MR patients, and these values appear in Table 4. This table also presents the classification accuracies for these modified cutting scores applied to the present samples. Clearly, the modified cutting score for the LMT failed to cross-validate successfully here. However, in three cases (DMT, TOMM Trial 2, and TOMM retention trial), these modified cutting scores produced comparable specificity rates in the present sample. At first blush, it may seem surprising that only modest changes in specificity accompany the large drop in cutting scores applied here relative to those presented in Table 3 (original cutting scores were <93% for the LMT, and <90% for the two TOMM trials and the DMT). This occurred because low-scoring outliers in TOMM and DMT results from the

Table 2
WAIS–III IQ Scores From Community Volunteers and MR Participants

Variable	CVH	CVM	MR	*F*(2, 56)	CVH vs. CVM (Cohen's *d*)	CVM vs. MR (Cohen's *d*)
WAIS FSIQ						
M	84.80	67.76	63.21	25.55***	1.91***	−.59
SD	9.37	8.78	6.66			
WAIS VIQ						
M	86.10	69.88	64.13	26.15***	1.90***	−.82*
SD	12.10	7.18	6.92			
WAIS PIQ						
M	85.90	70.80	68.96	13.12***	1.48***	−.21
SD	10.55	10.10	7.16			

Note. WAIS–III = Wechsler Adult Intelligence Scale (3rd ed.; Wechsler, 1997); CVH = community volunteer honest; CVM = community volunteer malingering; MR = mentally retarded; FSIQ = Full Scale Intelligence Quotient; VIQ = Verbal Intelligence Quotient; PIQ = Performance Intelligence Quotient.
* $p < .05$. *** $p < .001$.

Table 3
Feigning Measure Results

Variable	Cut score	CVH *M*	CVH *SD*	CVM *M*	CVM *SD*	MR *M*	MR *SD*	*F*(2, 56)	CVH vs. CVM (Cohen's *d*)	CVM vs. MR (Cohen's *d*)	SENS	SPEC	HR
WAIS digit span SS	<5	7.70	1.25	5.68	2.10	4.88	1.65	8.66**	1.08**	−.43	.44	.38	.41
WAIS voc—DS	>4	0.30	3.53	−1.08	2.48	−1.63	1.58	2.31	.50	.27	.04	1.00	.51
WAIS MDF	>0.21	−0.15	0.61	−0.13	0.84	0.07	0.59	.60	−.03	.28	.28	.58	.43
WAIS RDS	<7	8.20	1.03	6.68	1.93	5.67	1.27	9.51***	.90**	−.63*	.48	.33	.41
DMT % correct	<90%	99.42	1.22	81.42	19.69	95.19	6.14	9.27***	1.23***	1.05**	.48	.88	.67
LMT % correct	<93%	99.55	0.95	77.46	24.80	87.87	12.47	5.69**	1.21***	.55	.60	.61	.60
TOMM Trial 2 % correct	<90%	99.00	1.70	81.28	23.56	94.50	12.20	5.33**	1.01**	.73*	.40	.88	.64
TOMM retention trial % correct	<90%	99.80	0.63	78.80	25.85	97.08	6.27	8.76***	1.11***	1.12**	.44	.92	.68
b Test E score	>120	96.97	123.75	494.65	607.08	212.81	229.93	4.21*	−.84**	−.67*	.56	.69	.63
WMT IR % correct	<82.5%	90.75	6.73	69.92	20.88	78.40	13.92	5.83**	1.22***	.49	.68	.42	.51
WMT DR % correct	<82.5%	94.65	6.03	71.34	24.53	80.19	13.81	5.77**	1.20***	.46	.60	.42	.51
WMT consistency %	<82.5%	88.25	9.34	57.02	27.82	67.98	18.05	7.32**	1.37***	.48	.80	.25	.53
SIRS total score	>76	48.00	16.89	82.20	46.22	66.70	39.77	2.73	−.89**	−.36	.52	.65	.58

Note. CVH = community volunteer honest; CVM = community volunteer malingering; MR = mentally retarded; SENS = sensitivity; SPEC = specificity; HR = hit rate; WAIS = Wechsler Adult Intelligence Scale (3rd ed.; Wechsler, 1997); voc = vocabulary; DS = digit span; MDF = Mittenberg discriminant function; RDS = reliable digit span; DMT = Digit Memory Test (Guilmette, Hart, Guiliano, & Leininger, 1994); LMT = Letter Memory Test (Inman et al., 1998); TOMM = Test of Memory Malingering (Tombaugh, 1996); WMT = Word Memory Test (Green, 2005); IR = immediate recognition; DR = delayed recognition; SIRS = Structured Interview of Reported Symptoms (Rogers, Bagby, & Dickens, 1992).
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4
Neurocognitive Malingering Test Operating Characteristics for Discriminating Community Volunteer Malingerers and Mentally Retarded Participants Using Modified Cutting Scores From Graue et al. (2007)

Test	Cut score	SN	SP	HR	PPP	NPP
DMT % correct	<80%	.40	.96	.67	.91	.61
LMT % correct	<70%	.36	.83	.59	.69	.55
TOMM Trial 2 % correct	<60%	.24	.96	.57	.86	.55
TOMM retention trial % correct	<60%	.20	1.00	.59	1.00	.55

Note. SN = sensitivity; SP = specificity; HR = hit rate (percentage correct classification); PPP = positive predictive power; NPP = negative predictive power; DMT = Digit Memory Test (Guilmette, Hart, Guiliano, & Leininger, 1994); LMT = Letter Memory Test (Inman et al., 1998); TOMM = Test of Memory Malingering (Tombaugh, 1996).

original and present MR samples necessitated a large adjustment in cutting scores to obtain the target specificity rates. Sensitivity rates for the modified cutting scores were rather weak.

Table 4 also presents positive predictive power (PPP) and negative predictive power (NPP) values using the prevailing base rate in this study (51%). As previously noted, the only available published data on objectively determined feigning rates in compensation-seeking individuals with MR suggest comparable base rates of 50%–60% (Chafetz et al., 2007). Thus, the malingering prevalence in the present samples appears reasonable to apply here. Of course, in settings with different base rates of feigning, estimated predictive values will need to be recalculated to reflect the varying prevalence rate.

Finally, there has been increased interest in evaluating the combined accuracy rates of groups of neurocognitive feigning tests (Larrabee, 2008; Victor, Boone, Serpa, Buehler, & Ziegler, in press). Table 5 presents results from the application of increasingly stringent criteria for predicting feigning using only the instruments whose modified cutting scores cross-validated successfully here (e.g., TOMM and DMT). Table 5 indicates that, using the TOMM Trial 2 and retention trial, as well as the DMT, and at the approximately 51% base rate of feigning in these samples, predicting feigning when at least one of the three procedures was failed had

Table 5
Test Operating Characteristics for Using Increasingly Stringent Criteria for Predicting Feigned MR Using the DMT, TOMM Trial 2, and TOMM Retention With Graue et al. (2007) Revised Cutting Scores (LMT Excluded)

Test failures	SN	SP	HR	PPP	NPP
≥1	.44	.92	.68	.85	.61
≥2	.24	1.00	.61	1.00	.56

Note. MR = mental retardation; DMT = Digit Memory Test (Guilmette et al., 1994); TOMM = Test of Memory Malingering (Tombaugh, 1996); LMT = Letter Memory Test (Inman et al., 1998); failures = number of neurocognitive malingering tests falling below recommended; SN = sensitivity; SP = specificity; HR = hit rate (percentage correct classification); PPP = positive predictive power; NPP = negative predictive power.

a specificity of .92 and a PPP of .85 (PPP = probability of malingering given positive test sign), whereas requiring failure of at least two of the three indices had a specificity rate of 1.00 and a PPP of 1.00. However, NPP (probability of honest responding given a negative test sign) was modest for both these thresholds.

Discussion

The present study attempted to cross-validate the findings of Graue et al. (2007) in new samples as well as to explore the possible efficacy of additional predictors of feigned MR. Results were mixed. Overall, the present findings confirmed reduced specificity rates using widely employed cognitive malingering procedures at standard cutting scores, particularly for three of the WAIS–III feigning indicators, as well as for the LMT. The one WAIS–III feigning indicator that produced strong specificity (Vocabulary − Digit Span) had a dismal sensitivity rate of .04 for feigned intellectual impairment. One unexpected finding in the present samples was, on the basis of standard cutting scores, an increase in the specificity rates for the TOMM and DMT, approaching .90 for two of the indicators and exceeding this level for the TOMM retention trial. None of the feigning indicators new to the present study fared well, with the WMT, b Test, and SIRS all showing low specificity for MR. The SIRS result may not be entirely surprising, as it was not intended to detect feigned cognitive impairment (Rogers et al., 1992) and was used here only because of the positive results reported by Hayes et al. (1998).

One finding here that was generally consistent with Graue et al. (2007) was the increased specificity rates for the TOMM and DMT using the modified cutting scores suggested because they maintained specificity rates at or above .95 in the previous sample. Thus, the revised cutting scores were successfully cross-validated in a new sample of individuals with mild MR. Combining the three indices resulted in strong PPP for failure of two or more indices. However, predictive values are, of course, partly dependent on base rates, suggesting that PPP will fall in settings with lower base rates of feigning. Nevertheless, if these revised cutting scores are confirmed in additional samples of noncompensation-seeking patients with MR, they may contribute to assessment of feigning of this condition. However, one issue that will need to be addressed is the relatively low sensitivity of the combination of the TOMM and DMT using the revised cutting scores (.24).

One major factor to consider in the present results is the base rate of malingered MR in the current samples (51%). Base rates are not thought to affect sensitivity and specificity parameters, but they may have an important impact on PPP and NPP. Although a feigning prevalence of 51% might initially appear rather high, as noted earlier in the article, the only objective, published study estimating base rates of feigning in compensation-seeking patients undergoing evaluation for MR reported slightly higher prevalences of malingering, at 50%–60% (Chafetz et al., 2007). In addition, Ardolf, Denney, and Houston (2007) reported that 53.3% of criminal defendants referred for forensic neuropsychological examinations failed three or more objective symptom validity tests, whereas 54.3% were classified as probable or definite feigners of neurocognitive dysfunction. Clinicians evaluating defendants claiming MR may choose to calculate positive and negative predictive powers using lower base rates of feigning, hopefully on the basis objective indicators of malingering prevalences in their settings.

There are a number of limitations to the present study that should be carefully considered. These results are based on a simulation design, and although several measures for boosting the generalizability of the study were taken, it is unlikely that a simulation design can approximate all the factors in play during real-world, high-stakes evaluations. Another issue is the modest sample sizes that included only 25 malingering community volunteers and 24 persons with mild MR. These small sample sizes put the stability of extrapolated sensitivity and specificity values into question. However, the fact that the modified cutting scores for the TOMM and DMT generally replicated previous specificity values raises some hope for their generalizability. The uncertain comparability of the orally presented and computer-presented versions of the WMT is a further question. Another concern is that RAs testing the patients with MR were not blind to their diagnoses, as these evaluations all occurred at BGRMH/MRB facilities. In addition, the present samples were all drawn from an urban area in central Kentucky, again possibly limiting generalizability. Finally, clinical evaluations for diagnosing MR require documentation of developmental onset, as well as deficits in adaptive functioning, which were not addressed in the present malingering group. It is important to consider these two factors, as well as results from symptom validity tests, in examinations for MR in adults.

Overall, although there were some helpful findings in the present study, broadly speaking, neurocognitive feigning measures derived primarily on the basis of traumatic brain injury samples do not seem to work well in patients with MR. Given the fairly consistent reports to this effect in the published literature, it may well be that a new approach is needed to develop feigning-detection instruments that are both sensitive and specific for the evaluation of possible malingered MR. Research assessing a broad array of possible approaches to this issue is clearly needed by the forensic clinical community.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.) Washington, DC: Author.

Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *Clinical Neuropsychologist, 21,* 899–916.

Atkins v. Virginia, 536 U.S. 304 (2002).

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *Clinical Neuropsychologist, 20,* 145–159.

Boone, K. B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: Guilford Press.

Boone, K. B., Lu, P., Sherman, D., Palmer, B., Back, C., Shamieh, E., et al. (2000). Validation of a new technique to detect malingering of cognitive symptoms: The b Test. *Archives of Clinical Neuropsychology, 15,* 227–241.

Brockhaus, R., & Merten, T. (2004). Neuropsychological assessment of suboptimal performance: The Word Memory Test. *Nervenarzt, 75,* 882–887.

Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the Social Security Disability Consultative Exam: A new rating scale. *Archives of Clinical Neuropsychology, 22,* 1–14.

Dean, A. C., Victor, T. L., Boone, K. B., & Arnold, G. (2008). The relationship of IQ to effort test performance. *Clinical Neuropsychologist, 22,* 705–722.

Dean, A. C., Victor, T. L., Boone, K. B., Philpott, L. M., & Hess, R. A. (in press). Dementia and effort test performance. *Clinical Neuropsychologist.*

Granacher, R. P., & Berry, D. T. R. (2008). Feigned medical presentations. In R. Rogers (Ed.), *Clinical assessment of malinger and deception* (3rd ed., pp. 145–156). New York: Guilford Press.

Graue, L. O., Berry, D. T. R., Clark, J. A., Sollman, M. J., Cardi, M., Hopkins, J., & Werline, D. (2007). Identification of feigned mental retardation using the new generation of malingering detection instruments: Preliminary findings. *Clinical Neuropsychologist, 21,* 921–942.

Green, P. (2005). *Green's Word Memory Test for Windows user's manual.* Seattle, WA: Green's Publishing.

Greiffenstein, M. F., Baker, J. W., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6,* 218–224.

Guilmette, T. J., Hart, K. J., Guiliano, A. J., & Leininger, B. E. (1994). Detecting simulated memory impairment: Comparison of the Rey Fifteen-Item Test and the Hiscock Forced-Choice Procedure. *Clinical Neuropsychologist, 8,* 283–294.

Hayes, J. S., Hale, D. B., & Gouvier, W. D. (1997). Do tests predict malingering in defendants with mental retardation? *Journal of Psychology, 131,* 575–576.

Hayes, J. S., Hale, D. B., & Gouvier, W. D. (1998). Malingering detection in a mentally retarded forensic population. *Applied Neuropsychology, 5,* 33–36.

Hurley, K. E., & Deal, W. P. (2006). Assessment instruments measuring malingering used with individuals who have mental retardation: Potential problems and issues. *Mental Retardation, 44,* 112–119.

Inman, T. H., Vickery, C. D., Berry, D. T. R., Lamb, D., Edwards, C., & Smith, G. T. (1998). Development and initial validation of a new procedure for evaluating adequacy of effort during neuropsychological testing: The Letter Memory Test. *Psychological Assessment, 10,* 128–139.

Iverson, G. L., & Tulsky, D. S. (2003). Detecting malingering on the WAIS–III: Unusual digit span performance patterns in the normal population and in clinical groups. *Archives of Clinical Neuropsychology, 17,* 1–9.

Larrabee, G. J. (1998). Somatic malingering on the MMPI and MMPI–2 in litigating subjects. *Clinical Neuropsychologist, 12,* 179–188.

Larrabee, G. J. (2007). *Assessment of malingered neuropsychological deficits.* New York: Oxford University Press.

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *Clinical Neuropsychologist, 22,* 666–679.

Marshall, P., & Happe, M. (2007). The performance of individuals with mental retardation on cognitive tests assessing effort and motivation. *Clinical Neuropsychologist, 21,* 826–840.

Mittenberg, W., Theroux, S., Aguila-Puentas, G., Bianchini, K. J., Greve, K. W., & Rayls, K. (2001). Identification of malingered head injury on the Wechsler Adult Intelligence Scale (3rd ed.). *Clinical Neuropsychologist, 15,* 440–445.

Otto, R. K. (2008). Challenges and advances in the assessment of response styles in forensic examination contexts. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 365–375). New York: Guilford Press.

Psychological Corporation. (2001). *Wechsler Test of Adult Reading manual.* San Antonio, TX: Psychological Corporation.

Reznek, L. (2005). The Rey 15-item memory test for malingering: A meta-analysis. *Brain Injury, 19,* 539–543.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms (SIRS) professional manual.* Odessa, FL: Psychological Assessment Resources.

Schretlen, D., & 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.

Simon, M. J. (2007). Performance of mentally retarded forensic patients on the Test of Memory Malingering. *Journal of Clinical Psychology, 63,* 339–344.

Tombaugh, T. N. (1996). *Test of Memory Malingering.* Toronto: Multi-Health Systems.

Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology, 16,* 45–73.

Victor, T. L., & Boone, K. B. (2007). Identification of feigned mental retardation. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 310–345). New York: Guilford Press.

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *Clinical Neuropsychologist, 23,* 297–313.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio, TX: Psychological Corporation.

Received February 23, 2009
Revision received April 20, 2009
Accepted April 21, 2009 ■