

Efficacy of Test of Memory Malinger Trial 1, Trial 2, the Retention Trial, and the Albany Consistency Index in a Criterion Group Forensic Neuropsychological Sample

R W Schroeder^{*,a}, W H Buddin^a, Jr D D Hargrave^a, E J VonDran^a, E B Campbell^a, C J Brockman^a, R J Heinrichs^a, L E Baade^a

^a Department of Psychiatry and Behavioral Sciences, University of Kansas School of Medicine, Wichita, KS, USA.

* Corresponding author at: Department of Psychiatry and Behavioral Sciences, University of Kansas School of Medicine, Wichita, KS 67206, USA. Tel.: 316-293-3850; fax: 316-683-6733. E-mail address: ryan.w.schroeder.psyd@hotmail.com (R. Schroeder).

Abstract The Test of Memory Malinger is one of the most popular and heavily researched validity tests available for use in neuropsychological evaluations. Recent research has suggested, however, that the original indices and cutoffs may require modifications to increase sensitivity rates. Some of these modifications lack cross-validation and no study has examined all indices in a single sample. This study compares Trial 1, Trial 2, the Retention Trial, and the newly created Albany Consistency Index in a criterion group forensic neuropsychological sample. Findings lend support for the newly created indices and cutoff scores. Implications and cautionary statements are provided and discussed.

Keywords: Test of Memory Malinger, TOMM, Albany consistency index, Forensic neuropsychology, Validity.

Introduction

According to a survey of neuropsychologists' beliefs and practices, the Test of Memory Malinger (TOMM; Tombaugh, 1996) is the most frequently used performance validity test (PVT; Sharland & Gfeller, 2007). This is not surprising given that the measure is heavily researched and multiple studies have found patients' scores to be unaffected by age, education, pain, psychiatric conditions, and all but the most severe neurocognitive conditions (Ashendorf, Constantinou, & McCaffrey, 2004; Gunner, Miele, Lynch, & McCaffrey, 2012; Iverson, Le Page, Koehler, Shojania, & Badii, 2007; Tombaugh, 1996, 1997, 2003). Despite clinicians' favorable attitudes toward the measure and the abundance of research supporting its use, it has recently been suggested that the TOMM cutoffs and indices

may require modifications to maximize sensitivity rates (Greve, Binder, & Binachini, 2009; Greve, Ord, Curtis, Bianchini, & Brennan, 2008; Gunner et al., 2012).

To increase sensitivity, some authors have altered the TOMM Trial 2 and Retention Trial cutoff scores. When maintaining specificity at 90%—the desired minimum level of specificity for validity testing (Boone, 2007)—it was determined that a cutoff of ≤ 48 for both Trial 2 and the retention trial could be applied to some clinical samples (Greve, Bianchini, & Doane, 2006). Specifically, when this new cutoff was used in place of the traditional TOMM cutoff in a mild traumatic brain injury (TBI) sample grouped by Malingered Neurocognitive Dysfunction (MND) criteria (Slick, Sherman, & Iverson, 1999), sensitivity rates increased from 40% to 70% on Trial 2 and from 57% to 60% on the Retention

Trial (Greve, Bianchini, & Doane, 2006). Additionally, when the new cutoff replaced the traditional cutoff in a toxic exposure sample grouped by MND criteria (Greve, Bianchini, Black, et al., 2006), sensitivity rates increased from 55% to 61% on Trial 2 and from 52% to 68% on the Retention Trial, with specificity remaining above 90% on both trials.

Although the ≤ 48 cutoff showed promise in the mild TBI and toxic exposure samples, specificity suffered when the cutoff was applied to a moderate-to-severe TBI sample differentiated by MND criteria (Greve, Bianchini, & Doane, 2006). In the credibly performing moderate-to-severe TBI sample, a cutoff of ≤ 46 was required to maintain adequate specificity rates on Trial 2 and the Retention Trial (91% specificity was observed on both trials). When this cutoff was compared with the traditional TOMM cutoffs in the non-

credible moderate-to-severe TBI sample, sensitivity increased from 46% to 55% on Trial 2 and from 46% to 64% on the Retention Trial.

In addition to modifying the traditional TOMM cutoff scores, some authors have attempted to increase sensitivity by utilizing Trial 1 as a validity measure. Using MND criteria to derive groups, Greve, Bianchini, and Doane (2006) found that a Trial 1 cutoff score of ≤ 43 resulted in a sensitivity rate of 73% and a specificity rate of 91% in a mild TBI sample. The authors indicated, however, that this Trial 1 cutoff score would not be appropriate for their moderate-to-severe TBI sample if 90% specificity was desired. For the moderate-to-severe TBI sample, a cutoff score of ≤ 38 produced the best sensitivity (46%) while maintaining adequate specificity (91%). Overall, the authors concluded that Trial 1 can be an accurate indicator of negative response bias.

Others have found similarly promising results when utilizing Trial 1 in mixed clinical samples. For example, O'Bryant, Engel, Kleiner, Vasterling, and Black (2007) evaluated Trial 1 cutoffs in a mixed neuropsychological outpatient sample divided by definite MND and non-MND criteria. Using a cutoff of ≤ 40 , the authors found sensitivity and specificity rates of 79% and 90%, respectively. These rates are strikingly similar to rates reported in a study that reviewed and combined multiple TOMM Trial 1 findings (Denning, 2012). When 18 independent studies utilizing diverse clinical and forensic groups were pooled using weighted averages, an average cutoff of ≤ 40 yielded mean sensitivity and specificity rates of 77% and 92%, respectively.

Finally, in the most recent attempt to increase sensitivity rates, Gunner and colleagues (2012) developed a consistency index for the TOMM,

called the Albany Consistency Index (ACI). For a complete description of the computation of the ACI, the reader is referred to the original article. In brief, the index consists of summing the number of items that are inconsistently responded to across Trial 1, Trial 2, and the Retention Trial. For example, an item that is correctly answered on two TOMM trials (e.g., Trial 1 and Trial 2) but incorrectly answered on a third trial (e.g., the Retention Trial) is classified as an inconsistent item response. When comparing groups of patients classified as providing optimal or suboptimal effort, derived from Word Memory Test (WMT; Green, 2003) performances, the traditional TOMM Trial 2 cutoff score resulted in sensitivity and specificity rates of 33% and 96%, respectively. The ACI, however, yielded sensitivity and specificity rates of 71% and 100%, respectively, when using a cutoff of ≥ 10 inconsistent responses.

As can be seen, studies have shown that both adjustments of traditional TOMM cutoff scores and the addition of new indices may increase the measure's sensitivity to neurocognitive malingering. However, this body of literature is relatively small and it is lacking in studies that examine all TOMM indices in a single forensic sample. The purpose of the present study was to examine the utility of TOMM Trial 1, Trial 2, the Retention Trial, and the ACI in an outpatient forensic neuropsychological sample grouped by MND criteria.

Method

Participants

This is an archival study of 69 consecutive forensic cases (i.e., compensation seeking, litigation, or disability), some of which were utilized in a previous study (Schroeder, Baade, et al., 2012). All

patients were referred to a university medical center neuropsychology clinic, directed by a board-certified neuropsychologist, for forensic evaluations. The majority of patients presented with complaints related to TBIs. Specifically, 34 patients had histories consistent with mild TBIs, as defined by the American Congress of Rehabilitation Medicine's Mild Traumatic Brain Injury Committee (Committee on Mild Traumatic Brain Injury, 1993). Of these patients, 26 had uncomplicated mild TBIs (i.e., lack of acute intracranial pathology on neuroimaging), whereas 8 had complicated mild TBIs (i.e., positive findings of acute intracranial pathology of neuroimaging). In addition to patients with mild TBIs, patients with moderate-to-severe TBIs were included in this study ($n = 7$). The remaining patient diagnoses were major depressive disorder ($n = 5$), frontotemporal dementia ($n = 5$), cerebrovascular accident ($n = 3$), hypoxic brain injury ($n = 3$), posttraumatic stress disorder ($n = 3$), mild cognitive impairment ($n = 2$), psychotic disorder (not actively psychotic at the time of testing; $n = 2$), mental retardation ($n = 2$), Huntington's disease ($n = 1$), non-epileptic seizures ($n = 1$), and chronic pain ($n = 1$). Because patients with mental retardation and dementia have neurocognitive impairments that can potentially result in false-positive errors on some validity measures, patients with these diagnoses were excluded from the final analyses. As a result, the final study sample was comprised of 62 patients.

All of the 62 patients included for final analyses were differentiated by MND criteria, as described in the "Procedures" section. Overall, 36 patients (58%) did not meet criteria for any degree of MND, 24 patients (39%) were categorized as meeting criteria for probable MND, and two patients (3%) were categorized as

meeting criteria for definite MND. Thus, 42% of the forensic cases, which are primarily TBI-related, met criteria for neurocognitive malingering: a rate that is similar to base rates reported in the literature (e.g., Larrabee, 2003; Mittenberg, Patton, Canyock, & Condit, 2002). Table 1 shows demographic information for the groups “passing” and “failing” MND criteria.

Procedures

All patients included in this study underwent comprehensive forensic neuropsychological evaluations consisting of record reviews, a clinical diagnostic interview, neurocognitive testing, psychological/personality testing, and validity testing. Although there were slight variations in the tests administered across the neuropsychological batteries, as dictated by clinical need, each patient received a similar core set of tests. All tests were administered according to standardized instructions by neuropsychology post-doctoral fellows, neuropsychology pre-doctoral interns, or trained neuropsychology technicians working under the supervision of a board-certified neuropsychologist.

As outlined in the MND criteria (Slick et al., 1999), patients were differentiated into appropriate criterion groups using both behavioral criteria of negative response bias and the results of validity testing. There were three behavioral criteria of negative response bias utilized in this study. The first criterion was a pattern or severity of neuropsychological dysfunction not consistent with the neuropsychological condition. The second criterion was having markedly inconsistent performances across neuropsychological testing. The third criterion was having implausible self-reported symptoms on clinical interview. All of these behavioral criteria of negative response bias contributed to MND classification, however, for this study, at least one validity measure also had to be failed in order to meet MND criteria.

The validity measures and cutoffs used for the classification of MND criteria are detailed in Table 2. It should be noted that not all patients were administered the exact same validity measures. Specifically, the Validity Indicator Profile (Frederick, 1997) was only given to a select number of patients based

on clinical necessity. Additionally, clinic policy dictated transition to the newer versions of the Wechsler Adult Intelligence Scale (WAIS) and Wechsler Memory Scale (WMS) upon their releases. Because this study utilizes data from clinical forensic patients, some of the included patients were administered the WAIS-Third Edition (WAIS-III; Wechsler, 1997a), whereas others were administered the WAIS-Fourth Edition (WAIS-IV; Wechsler, 2008) (Wechsler, 2008). Similarly, some patients were administered the WMS-Third Edition (WMS-III; Wechsler, 1997b) while others were administered the WMS-Fourth Edition (WMS-IV; Wechsler, 2009). Thus, depending on the test edition utilized, the appropriate WAIS and WMS embedded validity measures were employed.

It should also be noted that for this study, the WMT was examined for possible genuine memory impairment profile (GMIP; Green 2003) when one or more of the initial three WMT trials were failed. Although Green (2003) has noted that the initial three WMT trials are insensitive to all but the most extreme forms of cognitive dysfunction, Greve, Ord, Curtis,

Table 1. Demographic information by the classification group

Group	Age	Education	Gender (% male)	Race (% Caucasian)
Pass MND Criteria	40.83 (14.70)	12.89 (2.35)	56	92
Fail MND Criteria	44.08 (11.26)	12.68 (1.93)	65	85

Note: MND Criteria ¼ Malingered Neuropsychological Dysfunction Criteria.

Table 2. Validity measures and cutoff scores

Test	Cutoff score	Study
1. WAIS-III Processing Speed Index	St score ≤65	Curtis, Greve, and Bianchini (2009)
2. WAIS-III/WAIS-IV Reliable Digit Span	Reliable Digit Span score ≤6	Schroeder, Twumasi-Ankrah, Baade, and Marshall (2012)
3. Finger Tapping average dominant finger	Men ≤35, Women ≤28	Arnold and colleagues (2005)
4. WMS-III Auditory Immediate Index	St score ≤80	Ord, Greve, and Bianchini (2008)
5. WMS-IV Verbal Paired Associates-II Recognition	Raw score ≤27	Pearson (2009)
6. WMS-IV VR-II Recognition	Raw score ≤3	Pearson (2009)
7. Minnesota Multiphasic Personality Inventory (MMPI)-2 F or MMPI-2 Fp	T-score ≥80	Greve and colleagues (2008)
8. MMPI-2 Symptom Validity Scale	Raw score .27	Greve and colleagues (2008)
9. Word Memory Test	≤82.5%; No GMIP	Green (2003)
10. Validity Indicator Profile	Failure of either subtest	Frederick (1997)

Note: WAIS¼Wechsler Adult Intelligence Scale; WMS ¼ Wechsler Memory Scale; VR ¼ Visual Reproduction; GMIP ¼ genuine memory impairment profile. Some patients were administered the WAIS-III, whereas others were administered the WAIS-IV. Similarly, some patients were administered the WMS-III, whereas others were administered the WMS-IV. No patient received both versions of the WAIS or WMS. Thus, depending on the test edition utilized, the appropriate WAIS and WMS embedded validity measures were employed.

Bianchini, and Brennan (2008) have indicated that the initial three trials can result in relatively high false-positive error rates when applied to a TBI sample differentiated by MND criteria. Because the current study sample includes multiple patients with TBIs and it utilizes MND criteria, a more conservative approach of evaluating initial WMT failures in the context of a GMIP was utilized for this study.

Because multiple, diverse validity measures were used in this study, it is not surprising that sensitivity rates vary between many of the measures. Although it is exceedingly important to use validity measures that have high sensitivity rates, those with lower sensitivity rates may still have value when combined with the highly sensitive measures. For example, some patients feign global cognitive deficits, but others feign deficits in specific cognitive domains—typically the domains in which they report having cognitive difficulties (Boone, 2007). Thus, if a validity measure that generally has low sensitivity rates appears to be testing the cognitive domain that is being feigned, it might yield a more accurate outcome than a validity measure that has higher sensitivity rates but appears to be testing a cognitive domain that is not being feigned. An additional value of having multiple diverse validity measures is that a patient's effort/response bias can greatly fluctuate over the course of a neuropsychological evaluation

(Boone, 2009; Heilbronner et al., 2009; Schroeder & Marshall, 2011). A patient might start the evaluation by providing good and credible effort (and passing validity measures) but later lose motivation toward testing (and fail validity measures). Again, although one validity measure might be more sensitive than another, having multiple diverse validity measures could increase the overall true-positive hit rate (Larrabee, 2008). Indeed, this is a primary reason that all of the aforementioned validity measures were included in the current study.

Once patients were classified as passing or failing MND criteria, statistical analyses were performed. Mean scores and ranks for the TOMM indices were computed for groups passing and failing MND criteria. Statistics comparing and contrasting sensitivity, specificity, and overall hit rates for each of the TOMM indices were also calculated. Finally, correlations within TOMM indices and between TOMM scores and visual memory test scores were conducted.

Results

Table 3 shows mean scores and ranks for each TOMM index by the groups passing and failing MND criteria. As can be seen, the group passing MND criteria produced significantly better scores on all TOMM indices, $p < 0.01$.

Next, a receiver operating characteristic (ROC) curve was

generated for each index. As can be seen in Fig. 1, all TOMM indices provided good to excellent discriminative ability. The ACI achieved the highest area under the curve value ($AUC = 0.926$, 95% CI = 0.865–0.987), followed by Trial 1 ($AUC = 0.900$, 95% CI = 0.827–0.972), the Retention Trial ($AUC = 0.879$, 95% CI = 0.779–0.978), then Trial 2 ($AUC = 0.869$, 95% CI = 0.765–0.972). These results indicate that the ACI has the greatest classification ability when considering the combined effects of sensitivity and specificity for each measure.

Table 4 shows sensitivity and specificity rates for various cutoff scores on TOMM Trial 1, Trial 2, the Retention Trial, and the ACI when the sample is differentiated by MND criteria. Please note that Gunner and colleagues (2012) score the ACI as the number of inconsistent responses obtained (e.g., 10 inconsistent responses). To improve the readability of Table 4, the ACI was scored in the opposite direction (i.e., number of consistent responses attained). Thus, higher scores represent better performances on all four of the TOMM indices. As can be seen by examining the table, when specificity is set at 89% or greater, the ACI yielded the highest sensitivity rate (81%) of any index. When specificity is set at 90% or greater, various Trial 2 and Retention Trial cutoffs yielded the highest sensitivity rates (77%). When specificity is set at 95% or greater, a cutoff score of ≤ 47 on the

Table 3. Group performances on TOMM indices

Index	Group	Mean score (SD)	Mean rank	Mann-Whitney <i>U</i>	<i>p</i> -value
TOMM Trial 1	Pass MND	47.17 (3.86)	41.89	94.00	< .01
	Fail MND	35.92 (9.47)	17.12		
TOMM Trial 2	Pass MND	49.86 (0.68)	41.08	123.00	< .01
	Fail MND	41.96 (8.88)	18.23		
TOMM Retention	Pass MND	49.69 (0.95)	41.35	113.50	< .01
	Fail MND	39.88 (10.99)	17.87		
ACI	Pass MND	46.89 (4.48)	42.57	69.50	< .01
	Fail MND	30.15 (11.90)	16.17		

Notes: TOMM ¼ Test of Memory Malinger; MND ¼ Malingered Neuropsychological Dysfunction Criteria; ACI ¼ Albany Consistency Index. The TOMM Trial 1, Trial 2, and Retention mean scores are the mean number of items correct. The ACI mean score is the mean number of consistent responses.

Retention Trial yielded the highest sensitivity rate (73%). Thus, although the ACI has the greatest classification ability when considering the average effects of sensitivity and specificity, scores on other TOMM indices may be more accurate than the ACI at specific cutoff points.

Kappa statistics were computed to determine reliability between the MND criteria and TOMM Trial 1, Trial 2, the Retention Trial, and the ACI cutoffs. The cutoff for Trial 1 was set at 40 or fewer correct items (Denning, 2012), the cutoff for the ACI was set at 40 or fewer consistent item responses (Gunner et al., 2012), and the cutoffs for Trial 2 and the Retention Trial were the traditional cutoffs provided in the TOMM manual (Tombaugh, 1996). Results of these analyses indicate that the levels of concordance (Landis & Koch, 1977) between the TOMM Trial 1, Trial 2, and the Retention Trial and our MND groups were moderate with an absolute value of 0.41, 0.41, and 0.45, respectively. The level of concordance with the ACI was substantial at an absolute value of 0.62.

Next, Pearson's product-moment correlational analyses were performed to investigate possible relationships between TOMM scores and scores on true memory tests. As previously noted, some examinees were

administered the WMS-III, whereas the remaining consecutively referred forensic examinees were administered the WMS-IV. As a result of this test transition, the use of either battery alone for correlational analysis would

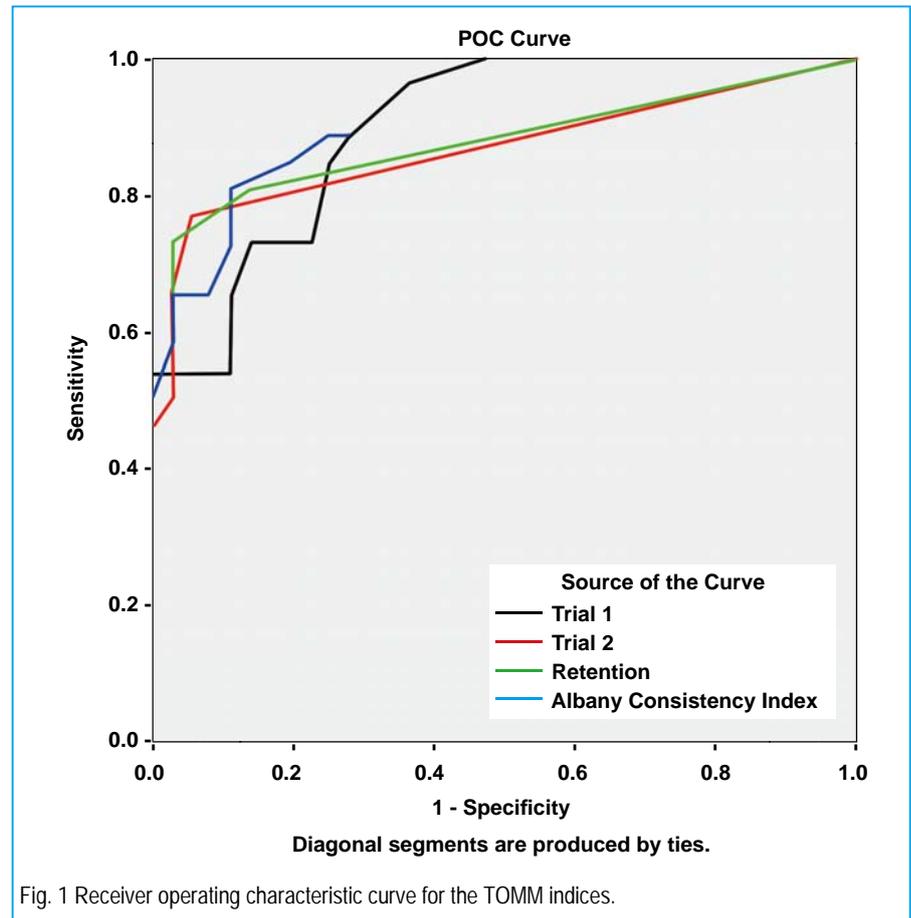


Fig. 1 Receiver operating characteristic curve for the TOMM indices.

Table 3. Group performances on TOMM indices

Cutoff	Trial 1		Trial 2		Retention		ACI	
	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
< 20	< 8	100	0	100	8	100	23	100
20	8	100	0	100	12	100	23	100
25	12	100	8	100	19	100	35	100
30	39	100	15	100	23	100	50	100
35	54	100	27	100	27	100	65	97
36	54	97	27	100	27	100	65	97
37	54	97	27	100	27	100	65	92
38	54	92	31	100	27	100	65	92
39	54	92	31	100	39	100	73	89
40	54	89	35	100	39	100	77	89
41	54	89	35	100	42	100	81	89
42	65	89	42	100	46	100	81	89
43	73	86	46	100	46	100	85	81
44	73	78	46	100	50	100	89	75
45	85	75	50	97	58	97	89	75
46	89	72	50	97	58	97	89	72
47	96	64	62	97	73	97	96	64
48	100	53	65	97	77	92	100	53
49	100	42	77	94	81	86	100	42
50	100	0	100	0	100	0	100	0

Notes: Sens. = Sensitivity; Spec. = Specificity; ACI = Albany Consistency Index. The Trial 1, Trial 2, and Retention Trial scores are the number of correct items. The ACI score is the number of consistent responses.

have resulted in an extremely small sample size. Consequently, the WMS-III and WMS-IV Visual Reproduction 2 (VR 2) and VR 2 scaled scores were coded independently and then combined, which allowed for a larger sample size to offset the aforementioned limitation (this is further described in the "Discussion" section). Results indicated that among the group passing MND criteria, none of the TOMM index scores correlated with either VR 1 or VR 2. However, in the group identified as failing MND criteria, TOMM Retention was found to correlate significantly with VR 1 ($r = 0.72$, $p < 0.01$).

Finally, correlational analyses demonstrated significant relationships between each of the four TOMM indices ($p < 0.001$), with values of 0.84 or higher. The authors had initially intended to use binomial logistic regression for further examination of TOMM indices; however, the strong relationship between these variables and the likelihood of multicollinearity among predictors precluded this analysis.

Discussion

The TOMM is one of the most popular and heavily researched PVTs available for use in neuropsychological evaluations. Nonetheless, recent research has suggested that modifications of the original indices and cutoff scores could increase sensitivity rates (Greve et al., 2008, 2009). Consequently, new cutoff scores for Trial 2 and the Retention Trial have been suggested, the utilization of Trial 1 as a PVT has been proposed, and a consistency index has been created. Some of these modifications lack cross-validation; however, no study has examined all indices in the same forensic sample. This study was undertaken to examine the efficacy of Trial 1, Trial 2, the Retention Trial, and the ACI in a

sample of forensic neuropsychological patients differentiated by MND criteria.

All indices included in this study significantly differentiated the groups of patients passing and failing MND criteria. The index achieving the greatest average classification ability was the ACI (AUC = 0.926). This was followed by Trial 1 (AUC = 0.900), the Retention Trial (AUC = 0.879), and then Trial 2 (AUC = 0.869). Although Trial 1 yielded the second largest AUC, the authors suggest caution in its clinical application. This is because individuals obtaining low scores on Trial 1 are likely to obtain low scores on all other indices as well (thus, there is a high sensitivity rate), but false-positive errors may occur among those who demonstrate genuinely poor learning with adequate performance on Trial 2 and the Retention Trial. This is demonstrated by the lower specificity of Trial 1, which does not approach the levels of specificity obtained by Trial 2 or the Retention Trial until 11 items are missed, at which point the sensitivity drops to 54% (compared with 77% on Trial 2 and the Retention Trial when at least 90% specificity is maintained).

Kappa statistics were computed in order to determine the agreement between each TOMM index and MND criteria when controlling for chance. This was deemed important as chance identification may result in either false-negative or false-positive errors, and because the other analyses do not take this confound into consideration. Classification of groups determined via TOMM Trial 1, Trial 2, and the Retention Trial all achieved moderate overall agreement with classification of groups via MND criteria, while the ACI achieved substantial agreement. These findings provide support for the use of each TOMM index, especially the ACI, in differentiating individuals providing credible versus non-

credible performances.

Finally, correlational analyses of TOMM scores from the group passing MND criteria provided evidence of divergent validity between all of the TOMM indices and true visual memory tests (VR 1 or VR 2). This is a function of the TOMM's exceedingly low ceiling in terms of its measurement of true memory abilities; thus, the lack of a correlation is expected. Conversely, the Retention Trial significantly correlated with visual memory test performances among the group performing non-credibly. This was also expected, as it was thought that patients who suppressed their TOMM scores were likely to suppress their scores on true memory tests as well.

This article contributes to the literature by comparing, contrasting, and providing data on the new TOMM indices and cutoff scores. However, further cross-validation is recommended. Few studies have evaluated TOMM Trial 1 scores when differentiating patients by MND criteria. Across studies that have used MND-based criteria (Denning, 2012; Greve, Bianchini, & Doane, 2006; O'Bryant et al., 2007), when 90% specificity rates were derived, sensitivity rates ranged from 46% to 79% depending on the clinical sample—the current sensitivity rates fall within this range as well (54%). This is a large range, and continued research should assist in determining more precise cutoffs and sensitivity rates for specific clinical groups.

A similar suggestion is offered for findings related to the ACI. Both the present study and the study by Gunner and colleagues (2012) indicated that the ACI is superior to the other TOMM indices in its ability to discriminate between groups of patients providing credible and non-credible performances. These are the only published studies on the ACI, and different criteria for classification of

credible performances were employed (i.e., MND criteria vs. WMT scores). Thus, further crossvalidation is recommended for this index as well.

A potential limitation of the current study is that some patients received the WMS-III during their forensic evaluations, whereas others received the WMS-IV. When conducting the correlational analyses, the authors combined VR scaled scores from both WMS batteries. The authors fully realize that many changes were made between the third and fourth versions of the WMS, rendering their simultaneous use methodologically tenuous. However, it is also realized that these two batteries measure the same construct driven by the same theory of memory. In addition, the tests chosen for analyses were VR 1 and VR 2, which retain the same set of stimuli from the WMS-III to the WMS-IV. Although the two versions of this test employ different raw scoring criteria, both sets of scores are linearly transformed to normalized scaled scores, which was the metric used for our analyses. Thus, it was decided to evaluate the scores individually and when combined, as the combination offers greater insight into the convergent and divergent validity of each TOMM index.

Another potential limitation of the study is that failure of two or more validity measures was not necessarily required for classification of probable MND (one validity measure failure and the presence of behavioral negative response bias was considered adequate). It could be argued that requiring failure of two or more validity measures would result in a more conservative criterion group. The authors contend, however, that the use of behavioral criteria combined with a validity measure failure is methodologically similar to requiring two validity measure failures. This has been supported by

research showing that the probability of identifying negative response bias via a combination of behavioral criteria and a single validity measure failure was comparable to the probability of identifying negative response bias via two validity measure failures (Marshall et al., 2010). Nevertheless, the current authors reviewed the data of those patients characterized as meeting probable MND in this study. Of the 24 patients who met criteria for probable MND, the number of validity measure failures ranged from 1 to 7 (mean $\frac{1}{4}$ 3.54), and all but three patients failed at least two validity measures. Those three patients failed one validity measure (two failed the WMT; one failed the MMPI-2 measures) and met criteria for behavioral negative response bias. Overall, rates of probable MND would have changed only slightly if failure of two or more validity measures were required as the criterion. Given this information and the increase in generalizability to clinical decision-making (Bush et al., 2005) and to other studies that utilize MND criteria, the authors retained the original classification criteria in this study.

Additional limitations of the current study deserve mention. First, the vast majority of our sample was comprised of Caucasian patients (89%). Although this sample is representative of the patients seen in our Kansas-based practice, the extent to which these results will generalize to samples of different racial and cultural backgrounds is unknown. Another potential limitation is that our sample was largely comprised of patients with mild TBIs. Further cross-validation with additional clinical groups is therefore advised. Finally, future research should utilize even larger study samples to allow for increased confidence and power.

Conclusions

Notwithstanding the noted limitations, this is the first study to evaluate all of the new TOMM indices and cutoffs in a single criterion group neuropsychological sample. Evidence was provided for convergent and divergent validity for all TOMM indices, which increases confidence for the clinical utility of both the new and traditional indices. Although each index well differentiated patients passing and failing MND criteria, the ACI was found to be the superior index. Because research on the new TOMM indices is still limited, however, further cross-validation is recommended.

Reference

- [1] Arnold, G., Boone, K. B., Lu, P., Dean, A., Wen, J., Nitch, S., et al. (2005). Sensitivity and specificity of finger tapping scores for the detection of suspect effort. *The Clinical Neuropsychologist*, 19, 105–120.
- [2] Ashendorf, L., Constantinou, M., & McCaffrey, R. J. (2004). The effect of depression and anxiety on the TOMM in community-dwelling older adults. *Archives of Clinical Neuropsychology*, 19, 125–130.
- [3] Boone, K. B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: Guilford Press.
- [4] Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, 23, 729–741.
- [5] Bush, S. S., Ruff, R. M., Trostler, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology*, 20, 419–426.
- [6] Committee on Mild Traumatic Brain Injury, American Congress of Rehabilitation Medicine (ACRM). (1993). Definition of mild traumatic brain injury. *Journal of Head Trauma Rehabilitation*, 8 (3), 86–87.

- [7] Curtis, K. L., Greve, K. W., & Bianchini, K. J. (2009). The Wechsler Adult Intelligence Scale-III and malingering in traumatic brain injury: Classification accuracy in known groups. *Assessment*, 16, 401–414.
- [8] Denning, J. H. (2012). The efficiency and accuracy of the Test of Memory Malingering Trial 1, errors on the first 10 items of the Test of Memory Malingering, and five embedded measures in predicting invalid test performance. *Archives of Clinical Neuropsychology*, 27, 417–432.
- [9] Frederick, R. I. (1997). *Validity indicator profile manual*. Minneapolis: National Computer Systems.
- [10] Green, P. (2003). *Green's Word Memory Test*. Edmonton, CA: Green's Publishing.
- [11] Greve, K. W., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., Swift, D. A., et al. (2006). Classification accuracy of the test of memory malingering in persons reporting exposure to environmental and industrial toxins: Results of a known-groups analysis. *Archives of Clinical Neuropsychology*, 21, 439–448.
- [12] Greve, K. W., Bianchini, K. J., & Doane, B. M. (2006). Classification accuracy of the test of memory malingering in traumatic brain injury: Results of a known groups analysis. *Journal of Clinical and Experimental Neuropsychology*, 28, 1176–1190.
- [13] Greve, K. W., Binder, L. M., & Bianchini, K. J. (2009). Rates of below-chance performance in forced-choice symptom validity tests. *The Clinical Neuropsychologist*, 23, 534–544.
- [14] Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced-choice symptom validity tests. *The Clinical Neuropsychologist*, 22, 896–918.
- [15] Gunner, J. H., Miele, A. S., Lynch, J. K., & McCaffrey, R. J. (2012). The Albany Consistency Index for the Test of Memory Malingering. *Archives of Clinical Neuropsychology*, 27 (1), 1–9.
- [16] Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R., & Conference Participants. (2009). *American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering*. *The Clinical Neuropsychologist*, 23, 1093–1129.
- [17] Iverson, G. L., Le Page, J., Koehler, B. E., Shojania, K., & Badii, M. (2007). Test of Memory Malingering (TOMM) scores are not affected by chronic pain or depression in patients with fibromyalgia. *The Clinical Neuropsychologist*, 21, 532–546.
- [18] Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33 (1), 159–174.
- [19] Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17, 410–425.
- [20] Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, 22, 666–679.
- [21] Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., et al. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. *The Clinical Neuropsychologist*, 24, 1204–1237.
- [22] Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094–1102.
- [23] O'Bryant, S. E., Engel, L. R., Kleiner, J. S., Vasterling, J. J., & Black, F. W. (2007). Test of memory malingering (TOMM) trial 1 as a screening measure for insufficient effort. *The Clinical Neuropsychologist*, 21, 511–521.
- [24] Ord, J. S., Greve, K. W., & Bianchini, K. J. (2008). Using the Wechsler Memory Scale-III to detect malingering in mild traumatic brain injury. *The Clinical Neuropsychologist*, 22, 689–704.
- [25] Pearson. (2009). *Advanced clinical solutions for the WAIS-IV and WMS-IV*. San Antonio, TX: Pearson.
- [26] Schroeder, R. W., Baade, L. E., Peck, C. P., VonDran, E. J., Brockman, C. J., Webster, B. K., et al. (2012). Validation of MMPI-2-RF validity scales in criterion group neuropsychological samples. *The Clinical Neuropsychologist*, 26, 129–146.
- [27] Schroeder, R. W., & Marshall, P. S. (2011). Evaluation of the appropriateness of multiple symptom validity indices in psychotic and non-psychotic psychiatric populations. *The Clinical Neuropsychologist*, 25, 437–453.
- [28] Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable digit span: A systematic review and cross-validation study. *Assessment*, 19, 21–30.
- [29] Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologist's beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology*, 22, 213–223.
- [30] Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 4, 545–561.
- [31] Tombaugh, T. N. (1996). *Test of Memory Malingering*. North Tonawanda, NY: Multi-Health Systems.
- [32] Tombaugh, T. N. (1997). *The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals*. *Psychological Assessment*, 9, 260–268.
- [33] Tombaugh, T. N. (2003). *The Test of Memory Malingering (TOMM) in forensic psychology*. *Journal of Forensic Neuropsychology*, 2, 69–96.
- [34] Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale - Third edition manual*. San Antonio, TX: The Psychological Corporation.
- [35] Wechsler, D. (1997b). *Wechsler Memory Scale - Third edition manual*. San Antonio, TX: The Psychological Corporation.
- [36] Wechsler, D. (2008). *Wechsler Adult Intelligence Scale - Fourth edition manual*. San Antonio, TX: The Psychological Corporation.
- [37] Wechsler, D. (2009). *Wechsler Memory Scale - Fourth edition manual*. San Antonio, TX: The Psychological Corporation. ■