Psychological Trauma:
Theory, Research, Practice, and Policy

ISSN: 1942-9681

2025, Vol. 17, No. 8, 1699–1706
https://doi.org/10.1037/tra0001754

Distinguishing Genuine From Simulated Dissociative Identity Disorder Using the Morel Emotional Numbing Test

John-Christopher A. Finley[1], Michelle Barth[2], and Bethany Brand[2]
[1] Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine
[2] Department of Psychology, Towson University

Objective: Few studies have identified validity tests that are appropriate for use in individuals with dissociative identity disorder (DID). This study investigated whether the Morel Emotional Numbing Test (MENT) could distinguish between genuine and simulated DID. ***Method:*** Thirty-five participants with DID diagnosed via a semistructured interview were compared to 88 participants taking an abnormal psychology course who were instructed to simulate DID. Group comparison analyses were conducted to examine differences in the MENT errors. Univariate logistic regression and receiver operating characteristic curve analyses examined how well the MENT errors discriminated between groups and identified a cutoff optimizing sensitivity while maintaining high specificity. ***Results:*** Simulators had significantly more errors across all three MENT sets compared to genuine DID participants. The total number of MENT errors across sets most strongly and significantly predicted genuine versus simulated DID groups. Receiver operating characteristic curve analyses indicated that the MENT could discriminate between genuine and simulated DID with outstanding classification accuracy (area under the curve = .95). A cutoff of ≥10 total errors on the MENT yielded an optimal balance of sensitivity (.86) and specificity (.94). However, the previously defined cutoff of ≥8 total errors also demonstrated high sensitivity (.87) and specificity (.89). ***Conclusions:*** Findings provide preliminary evidence for the MENT as a validity test for DID populations. Although the commonly used cutoff of ≥8 errors on the MENT demonstrated excellent psychometric properties, we recommend using a cutoff of ≥10 errors for individuals with DID to minimize the risk of false positives while maintaining strong sensitivity.

Clinical Impact Statement
This study examined whether the Morel Emotional Numbing Test (MENT) could identify individuals with clinically diagnosed dissociative identity disorder (DID) from individuals instructed to simulate a diagnosis of DID. Findings indicated that a cutoff of ≥10 errors on the MENT could discriminate between genuine and simulated DID with outstanding classification accuracy. This is among the few extant validity tests that may accurately detect feigning and malingering in DID populations without resulting in a high false positive rate.

Keywords: dissociative identity disorder, symptom validity, malingering, dissociation, feigning

Dissociative identity disorder (DID) is a psychiatric disorder characterized by the presence of more than one personality state and recurrent gaps in memory that is associated with complex psychological trauma in childhood (American Psychiatric Association, 2022; Dalenberg et al., 2012). The severe and wide range of psychopathology and comorbidities associated with DID can make diagnosis challenging (Brand, 2023). Complicating this issue further, practitioners must also consider the possibility that individuals may feign symptoms of DID in both forensic and clinical evaluations. Many neuropsychiatric populations with

This article was published Online First July 25, 2024.
John-Christopher A. Finley https://orcid.org/0000-0002-1854-8360
The authors have no conflicts of interest to disclose. The data reported in this article were collected as part of a larger study investigating measures that can distinguish between clinical and simulated dissociative identity disorder. While none of the primary variables (Morel Emotional Numbing Test scores) reported in this article have been previously published, some of the data from the inclusion criteria variables (Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders, 4th edition* Dissociative Disorders–Revised and Dissociative Experiences Scale) have been used in ten published articles.

John-Christopher A. Finley played a lead role in formal analysis, validation, visualization, writing–original draft, and writing–review and editing and a supporting role in conceptualization and methodology. Michelle Barth played a lead role in conceptualization, investigation, and project administration, a supporting role in formal analysis, writing–original draft, and writing–review and editing, and an equal role in methodology. Bethany Brand played a lead role in data curation, resources, and supervision and a supporting role in conceptualization, investigation, methodology, project administration, and writing–review and editing.

Correspondence concerning this article should be addressed to John-Christopher A. Finley, Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, 425 East Ontario Street, Chicago, IL 60611, United States. Email: jfinley3045@gmail.com

legitimate concerns overestimate their severity of symptoms during routine psychological evaluations (Finley, Cladek, et al., 2024), some of whom deliberately or inadvertently exaggerate (i.e., feign) symptoms even when there is no apparent primary or secondary gain (Bass & Wade, 2019; Merten et al., 2020). For these reasons, practitioners should incorporate well-validated measures of symptom exaggeration, referred to as symptom validity tests (SVTs), in their evaluations, including for DID (Larrabee, 2012). SVTs assess the veracity of symptomatic complaints with cutoffs that are either embedded within clinical self-report measures or self-report measures solely dedicated to symptom exaggeration (Larrabee, 2012). Scores exceeding a cutoff do not infer motivation, but rather raise concern about the accuracy of symptom reports (Sherman et al., 2020). Elevated scores should be interpreted alongside other measures, behavioral observations, and historical information to determine whether the individual's data is deemed "invalid" and noninterpretable (Boone, 2021).

It can be particularly challenging to interpret the nature of elevated scores among individuals with disorders of poorly understood etiology and diagnostic criteria that rely on self-report (e.g., Sellbom et al., 2012). This is especially true among individuals with DID since the etiology of dissociative self-states remains a subject of debate (Reinders & Veltman, 2021). One study found that 10% of 112 consecutive admissions to a clinic for dissociative disorders had factitious disorder or malingered DID (Coons & Milstein, 1994), but the base rate of feigning or elevated scores on SVTs in DID populations remains largely unknown. It is imperative to first identify theoretically and psychometrically appropriate tests for DID populations that do not conflate feigned and genuine symptoms before making judgments about the extent of and underlying reason(s) for elevated validity scores in this population. It is particularly important to identify tests that can differentiate genuine from feigned symptoms with high specificity (i.e., accurately detecting genuine symptoms without misclassifying them as feigned symptoms; Bianchini et al., 2005). Yet, reduced specificity is a primary concern in DID research (Brand et al., 2006), which increases the chance of classifying individuals as feigning when they are not. There is a clear need to identify more tests that can accurately assess symptom validity in individuals who may have DID. However, few studies have investigated the utility of validity tests in populations with dissociative symptoms (Brand, 2023; Merten & Merckelbach, 2013; Palermo & Brand 2019), most of which have focused on criminal defendants (Brand et al., 2017). Fewer studies have exclusively focused on DID populations. Although various types of validity tests exist, researchers in these studies have predominately investigated content-based measures of symptom overreporting (e.g., Ambrose et al., 2023; Barth et al., 2023; Brand & Chasson, 2015; Brand et al., 2006, 2014, 2019, 2021).

Overreporting indices are cutoffs that typically determine the validity of responses based on items with rare, obvious, improbable, extremely or unusually severe symptoms, unlikely combination of symptoms, or erroneous stereotypes (Rogers, 2018). Cutoffs should demonstrate at least 90% specificity and 50% sensitivity, suggesting that no more than 10% of individuals with genuine symptoms are deemed invalid, while detecting at least half of individuals who are feigning (Boone, 2021). Because the accuracy of these cutoffs depends on the sample from which they were derived, they may not work as well when applied to significantly more impaired populations (Rogers, 2018). However, validity failures should not be blindly attributed to genuinely severe psychopathology (González et al., 2010), especially since the association between the severity of genuine psychopathology and elevated validity scores is often nonlinear (Finley, Brook, et al., 2023). Most SVTs function effectively in individuals with serious mental illnesses; however, some include item content that increases the risk of false positives in those with genuinely severe psychopathology (Rogers, 2018). Cutoffs based on items that are rare in normative samples, but not necessarily rare in clinical samples, are most influenced by the presence of genuine and severe psychopathology (Sharf et al., 2017). Cutoffs that index the endorsement of highly improbable symptoms may be more appropriate for populations with severe psychopathology, but can still yield false positives when applied to unique neuropsychiatric populations (Gonzalez et al., 2024). Many validity test symptoms that are recognized as "improbable," such as having dissociative states (e.g., asking someone if they feel like they have different personalities), depersonalization (e.g., asking someone if they ever do not recognize themselves in the mirror), derealization, auditory hallucinations, or transient amnesia (e.g., asking someone if they ever totally lose their memory), may actually be associated with trauma exposure and/or DID (Brand et al., 2009; Foote & Park, 2008; Merckelbach et al., 2015; Pilton et al., 2015).

Using SVTs embedded within clinical self-report questionnaires, rather than freestanding SVTs, may exacerbate these issues regarding specificity because they sometimes comprise items that are also used to assess legitimate psychopathology. Yet, many DID studies have focused on embedded SVTs (e.g., Brand & Chasson, 2015; Brand et al., 2014). To improve specificity in embedded measures, researchers will often propose population-specific cutoffs (e.g., Erdodi et al., 2014; Finley, Cerny, et al., 2024; Tse et al., 2023). However, few studies have proposed cutoffs for DID (Barth et al., 2023). Some cutoffs have been adjusted for individuals with posttraumatic stress disorder (PTSD), but thus far, these have not worked well for individuals with DID, presumably because they have different symptomology and comorbidities that may influence the severity of symptom endorsement (e.g., Brand et al., 2006; Rogers et al., 2009). Revising the cutoffs to be more conservative can increase specificity and mitigate the risk of false positives, but likely at the cost of reduced sensitivity. Different types of validity tests, like freestanding tests, should therefore be considered.

Freestanding SVTs are measures solely dedicated to assessing symptom validity. They are typically more accurate and resistant to false positives in individuals with severe psychopathology as they often use forced-choice paradigms to elicit the presence of highly improbable symptoms. The few studies examining freestanding validity tests in individuals with DID have shown promising results. For example, Barth et al. (2023) found that a revised cutoff from the Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2000), a freestanding SVT, demonstrated high specificity (.89) and sensitivity (.96) in individuals with DID. Another freestanding test, the Test of Memory Malingering (TOMM; Tombaugh, 1997), has also shown efficacy in individuals with dissociative disorders and symptoms (Brand et al., 2019; Simotas, 2000). Unlike the M-FAST, the TOMM assesses the validity of test-taking performance rather than symptom reporting. Even if neurocognitive testing is not part of the evaluation, including the TOMM may be useful as it appears to measure amnesia, a symptom commonly feigned in criminal forensic cases (Bourget & Whitehurst, 2007). Unless an individual with DID is actively experiencing severe dissociative symptoms during the testing

(Frewen & Lanius, 2015), there is little reason to suspect that they lack the basic cognitive abilities to pass the TOMM. The Structure Inventory of Malingered Symptomology (SIMS; Widows & Smith, 2005) is a freestanding SVT that has not worked well with DID populations (Brand et al., 2021). This may be because it includes items known to be common in DID as well as in PTSD, psychosis, and other related conditions. A recent review found that the SIMS does not reliably distinguish feigned psychopathology from genuinely severe psychiatric conditions (Shura et al., 2022). This is an important finding because many studies examining symptom feigning in dissociative disorders have relied on the SIMS (for review, see Merckelbach et al., 2017).

These findings underscore the importance of determining which validity tests capture behavior or performance that is improbable in DID populations. However, identification of such measures beyond the TOMM and M-FAST is limited. It would be helpful to expand the repertoire of validity tests that demonstrate acceptable sensitivity and specificity for individuals with DID so that multiple tests can eventually be used throughout an evaluation. Employing multiple validity tests throughout an evaluation is recommended since individuals may feign different types of symptoms at different times (Boone, 2009). Furthermore, test security for common freestanding validity tests, like the TOMM, is becoming increasingly compromised (Miele et al., 2012).

The Morel Emotional Numbing Test (MENT; Morel, 1998) is another forced-choice freestanding validity test designed to identify feigned symptoms of PTSD. Given that PTSD is a common comorbidity with DID, the MENT may prove useful for assessing symptom feigning among individuals presenting with possible DID. The MENT paradigm is performance-based but is unique from most performance validity tests like the TOMM, since it does not appear to assess cognitive ability. It requires examinees to look at photographs of faces and select between words (e.g., happy or sad) that depict which feeling the person is experiencing. The test comprises 60 photographs, divided into three sets of 20, which can be administered in 5–10 min. For test security purposes, we refer readers to the manual for a full description of the MENT paradigm (Morel & Shepherd, 2008). The total score is based upon the number of errors made throughout all three sets, with higher scores indicating a greater likelihood of feigning. A review by Morel and Shepherd (2008) demonstrated that this test works well for individuals with PTSD. They found that across five studies using various methodological approaches, including known groups of individuals malingering PTSD as well as individuals simulating PTSD, the pooled sensitivity and specificity of the MENT was .79 (95% CI [.66, .88]) and .96 (95% CI [.85, .99]), respectively. Based on their findings, a cutoff of ≥8 should sufficiently discriminate between genuine and feigned PTSD symptoms. Although the MENT appears to be a promising validity test for individuals with PTSD, it has not been examined in individuals with DID.

The Present Study

Based on the extant literature, there are many reasons why existing validity tests may be inappropriate for individuals with DID. These reasons, however, do not negate the possibility that individuals presenting with symptoms suggestive of DID may still deliberately or inadvertently exaggerate symptoms. Rather, they underscore the need for further research determining which validity tests are appropriate for cases in which DID is a diagnostic possibility. This study examined how well the MENT could discriminate between participants with genuine versus simulated DID. Subsequently, we identified which MENT cutoff optimized sensitivity while maintaining adequate (≥.90) specificity for practice settings.

Method

Procedure

This study utilized a between-groups simulation design, comprising 35 participants with reliably diagnosed DID and 88 participants with simulated DID. Data were collected as part of a larger study investigating DID that was approved by the institutional review board. All participants provided informed consent and were evaluated in-person. Evaluations for participants with diagnosed DID were conducted individually and lasted ~3–5 hr. These participants first were administered the Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders, 4th edition* Dissociative Disorders-Revised (SCID-D-R; Steinberg, 1994), a semistructured interview used to ensure DID diagnosis was accurate; this was conducted by a licensed clinical psychologist experienced in DID assessment or research assistants supervised by the psychologist. DID participants then completed several clinical and validity tests including the MENT. DID participants were compensated $20 before and after the evaluation.

Evaluation for participants with simulated DID were conducted in small groups and lasted ~90 min since they did not undergo the SCID-D-R. The evaluation began with administration of the Dissociative Experiences Scale (DES; Bernstein & Putnam, 1986) to ensure no simulators had clinically elevated dissociative symptoms, followed by a demographic questionnaire that they were instructed to complete truthfully. To prevent simulators from exaggerating DID symptoms and to increase the effort they put into simulation, they were given the following instructions:

> You are invited to participate in a study investigating how effectively people can fake having a psychological disorder called DID. After completing some brief forms about yourself, you will learn about dissociative disorders by reading information provided to you. Then you will take a short true–false test to show how well you understand the symptoms of DID. Next you will complete a number of personality tests, answering as though you actually have DID. A research assistant will later interview you using standardized questions. You should answer the interview questions as well as the personality test questions in the way you think someone with DID would answer. However, during the interview you do not have to actually act like you have DID. Rather, you can act like you do normally, but answer the questions the way you think someone with DID would answer. The researchers would like you to fake having this disorder to the very best of your ability. The three students who best match responses from actual patients with DID will win $50 each. However, only students who have earned a grade higher than 70% on the true/false test will be eligible for the $50 prizes.

Simulators learned about DID symptoms from an online brochure published by an organization providing education about trauma-related disorders and verified for accuracy by the senior author (available upon request) and then completed a 10-item quiz requiring them to identify symptoms of DID from other psychiatric disorders. Simulators were then debriefed and informed that the $50 prizes would be awarded randomly to ensure anonymity. Simulators also had the option to receive course credit for participation. As

indicated in a previous study examining data from this sample (Brand et al., 2016), a manipulation check indicated that this group of instructed simulators performed significantly better on the quiz about DID symptomology than a group of simulators (not included in this study) who did not receive the educational brochure and did not take an abnormal psychology course, which was an inclusion criteria for this study group.

Participants

Participants with clinical DID ($n = 39$) were recruited either from a psychiatric inpatient unit for trauma disorders ($n = 27$) or outpatient private practices ($n = 12$). Inpatient participants were recruited via unit announcements, while outpatient participants were recruited by therapists who learned about the study from email announcements and professional listservs. Participants were included if they had a diagnosis of DID, had at least an eighth-grade reading level, were ages 18 and older, and were involved in psychotherapy. Participants were excluded ($n = 4$) from the study if their DID diagnosis was not confirmed by the SCID-D-R. Participants with psychotic disorders or significant brain injury were also excluded from the study, but other comorbidities were permitted. A prior publication using data from this sample (Brand et al., 2016) indicated that these participants had severe internalizing psychopathology related to depression, anxiety, and PTSD. None of these participants were involved in litigation or evaluation for disability or similar legal entitlements. The final sample comprised 35 individuals with clinical DID after exclusion criterion were applied.

Participants simulating DID ($n = 91$) were undergraduates 18 years and older who passed a college course in abnormal psychology and were recruited from a Mid-Atlantic university. Participants were excluded from the study if they exhibited clinically significant dissociative symptoms, as evidenced by scoring >30 on the DES, or if they scored <70% on the preliminary exam about DID symptomology ($n = 3$). The final sample comprised 88 simulators after exclusion criterion were applied. See Table 1 for sample demographics.

Measures

Historical diagnosis of DID was verified using the SCID-D-R (Steinberg, 1994). This is a validated semistructured interview comprising 277 items that have been shown to reliably assess DID criteria (Mychailyszyn et al., 2021; Steinberg, 2000). Participants simulating DID were not assessed with the SCID-D-R, but were screened for dissociative symptoms with the DES (Bernstein & Putnam, 1986). This is a validated self-report measure comprising 28 items that assess various dissociative symptomology. A score of >30 is an appropriate and empirically supported exclusionary cutoff for this measure (Carlson & Putnam, 1993). DID participants did not complete the DES.

The MENT (Morel, 1998) is a 60-item measure with three sets of 20 items that are used to assess feigned symptoms associated with PTSD. Participants were administered all three sets, and their scores were based on the number of errors made, with more errors indicating a greater likelihood of feigning. Based on prior research, the ideal cutoff for PTSD populations is ≥8 total errors (Morel & Shepherd, 2008). The total number of errors and errors per set were included as independent variables for statistical analyses. Prior research has demonstrated adequate internal consistency equivalences of the MENT (ranging from .87 to .94; Morel, 1998), as well as adequate convergence with the widely used symptom ($r = .71$; Morel, 1998) and performance ($r = .83$; Morel, 2008) validity tests.

Statistical Analyses

Independent-samples t tests or chi-square tests were conducted to examine differences in demographics and the total number of MENT errors between groups. Multivariate analysis of variance was run to analyze differences in errors per set, with follow-up univariate analysis of variances examining differences between groups. Simple logistic regression analyses were conducted to assess whether the total and set errors were associated with group membership. Receiver operating characteristic curve analyses were conducted to derive a MENT cutoff that meaningfully discriminated simulated from genuine DID with optimal sensitivity while maintaining ≥.90 specificity. Sensitivity and specificity values were used to calculate the positive and negative predictive accuracy for identifying feigning at various base rates. Overall classification accuracy for the MENT set and total errors was based on the area under the curve values. Acceptable area under the curve values ranged from .70 to .79, excellent values ranged from .80 to .89, and outstanding values ranged from .90 to 1.00 (Hosmer et al., 2013).

Results

Group comparison analyses indicated that participants with simulated DID were significantly younger than participants with diagnosed DID $t(40.52) = 10.95$, $p < .001$, but there were no

Table 1
Sample Demographics

Demographic	Genuine DID group ($n = 35$)	Simulated DID group ($n = 88$)
Age (years)	$M = 44.14$ ($SD = 11.97$; range = 19–60)	$M = 21.33$ ($SD = 5.26$; range 18–58)
Female sex	31 (89%)	66 (75%)
Racial identity		
Non-Hispanic White	31 (89%)	61 (69%)
Non-Hispanic Black	3 (8%)	14 (16%)
Hispanic	1 (3%)	5 (6%)
Asian	0 (0%)	7 (8%)
Other	0 (0%)	1 (1%)

Note. $N = 123$; DID = dissociative identity disorder.

significant differences regarding gender, $\chi^2(1) = 2.01$, $p = .128$, or race, $\chi^2(1) = 1.10$, $p = .295$. Given these differences, we then assessed whether age significantly influenced MENT scores within groups, and the findings were nonsignificant ($p > .05$).

Independent-samples t test showed that simulators ($M = 25.42$, $SD = 13.61$, range = 0–59) had more total MENT errors than participants with DID ($M = 2.63$, $SD = 3.39$, range = 0–12); $t(109.41) = 14.61$, $p < .001$, $d = 1.95$, 95% CI [1.49, 2.41]. A multivariate analysis of variance revealed a main effect of group status on the number of MENT errors per set, Wilks' $\lambda = .51$, $F(3,119) = 38.89$, $p < .001$, $\eta_p^2 = .50$. Follow-up univariate analysis of variances indicated that simulators had significantly more errors on Set 1 ($M = 9.01$, $SD = 4.58$; range = 0–2), Set 2 ($M = 8.83$, $SD = 4.85$, range = 0–20), and Set 3 ($M = 7.58$, $SD = 5.17$, range = 0–20) than DID participants' errors on Set 1 ($M = 0.89$, $SD = 1.37$, range = 0–4); Set 2 ($M = 0.94$, $SD = 1.37$, range = 0–4); and Set 3 ($M = 0.80$, $SD = 1.11$, range = 0–4).

Simple logistic regression analyses revealed that the total MENT errors ($\chi^2 = 83.65$, $\beta = 1.30$, $p < .001$, 95% CI [1.90, 1.48]) significantly predicted genuine versus simulated DID (Nagelkerke's $R^2 = .71$). The errors on Set 1 ($\chi^2 = 81.10$, $\beta = 1.89$, $p < .001$, 95% CI [1.53, 2.51]; Nagelkerke's $R^2 = .69$); Set 2 ($\chi^2 = 80.09$, $\beta = 2.04$, $p < .001$, 95% CI [1.60, 2.84]; Nagelkerke's $R^2 = .69$); and Set 3 ($\chi^2 = 65.93$, $\beta = 2.07$, $p < .001$, 95% CI [1.60, 3.03]; Nagelkerke's $R^2 = .60$) also significantly predicted genuine versus simulated DID groups.

As shown in Table 2, receiver operating characteristic curve analyses indicated that the total MENT errors significantly discriminated between genuine and simulated DID, with outstanding classification accuracy (area under the curve = .95). When ensuring ≥.90 specificity, cutoffs of ≥9 and ≥10 yielded similar sensitivity (.86) as the previously established cutoff of ≥8 (sensitivity = .87). The cutoff of ≥10 yielded the optimal balance of sensitivity (.86) and specificity (.94). Similarly, the number of errors per MENT set significantly discriminated between individuals with genuine and simulated DID, with slightly worse classification accuracy statistics than the total MENT errors. Nonetheless, sensitivity and specificity were exceptionally high for each MENT set.

Discussion

This study used a simulation design with a carefully controlled sample and manipulation checks to investigate whether the MENT could effectively serve as a validity test for adults with DID. Findings overall indicated that the MENT meaningfully discriminated between genuine and simulated DID. However, the exact cutoffs and operating characteristics identified in this study require confirmation.

Participants with diagnosed DID had substantially fewer errors across all MENT sets as compared to simulators. However, the magnitude in difference between the groups for each set score ($d = .50$) was weaker than for the total score ($d = 1.95$). On average, participants with DID had approximately two total errors on the MENT, which is consistent with previous research examining individuals with PTSD. A review study by Morel (2008) found that PTSD populations, on average, produced 3.3 total errors (95% CI [2.6, 4.1]) on the MENT. This similarity is important because most individuals with DID have comorbid PTSD, and DID is considered a developmental trauma disorder (American Psychiatric Association, 2022). Classification accuracy statistics were also slightly higher for the total number of errors than the number of errors per set. These findings suggest that the MENT total score may elicit the greatest difference in responses between persons with genuine versus malingered DID. However, the operating characteristics also suggest that using the set scores in isolation may sufficiently detect feigning or malingering.

After exploring the overall classification accuracy of the MENT total and set scores, we identified cutoffs that best discriminated simulated from genuine DID. Prior research suggests that validity cutoffs for PTSD might not work well for individuals with DID (Brand et al., 2006; Rogers et al., 2009). Yet, this study found that a cutoff of ≥8 total errors, which has previously been described as an ideal cutoff for individuals with PTSD (Morel & Shepherd, 2008), classified group membership with excellent sensitivity (.87) and specificity (.89). Furthermore, classification values in this study were similar to those in a review of PTSD populations, reporting a pooled sensitivity of .76 and specificity of .96 (Morel & Shepherd, 2008).

Table 2

Classification Accuracy Statistics and Optimal MENT Cutoffs for Classifying Genuine Versus Simulated DID

					10% base rate		20% base rate		30% base rate	
MENT indicator	AUC (95% CI)	Cutoff	SN	SP	PPP	NPP	PPP	NPP	PPP	NPP
Total errors	.95* [.92, .98]	**≥10**	**.86**	**.94**	**.61**	**.98**	**.78**	**.96**	**.86**	**.94**
		≥9	.86	.91	.51	.98	.70	.96	.80	.94
		≥8	.87	.89	.47	.98	.66	.96	.77	.94
Set 1 errors	.93* [.89, .97]	≥5	.81	1.00	1.00	.98	1.00	.95	1.00	.92
		≥4	**.88**	**.89**	**.47**	**.99**	**.67**	**.97**	**.77**	**.95**
		≥3	.89	.83	.37	.99	.57	.97	.69	.95
Set 2 errors	.94* [.90, .98]	≥5	.81	1.00	1.00	.98	1.00	.95	1.00	.92
		≥4	**.84**	**.91**	**.51**	**.98**	**.70**	**.96**	**.80**	**.93**
		≥3	.86	.80	.32	.98	.52	.96	.65	.93
Set 3 errors	.90* [.85, .95]	≥4	.74	.94	.58	.97	.76	.94	.84	.89
		≥3	**.80**	**.94**	**.60**	**.98**	**.77**	**.95**	**.85**	**.92**
		≥2	.84	.77	.29	.98	.48	.95	.61	.92

Note. $N = 123$. Bolded values are associated with cutoffs that may yield an optimal balance between sensitivity and specificity for clinical use. MENT = Morel Emotional Numbing Test; DID = dissociative identity disorder; AUC = area under the curve; CI = confidence interval; SN = Sensitivity; SP = Specificity; PPP = positive predictive power; NPP = negative predictive power.

* $p < .001$.

The present study is among the first to find an existing validity cutoff with high sensitivity and an acceptably low false positive rate (≤10%) in a DID sample. One reason the MENT may work well for individuals with DID is that it does not include self-reported questions about dissociation, psychological trauma, or symptoms associated with DID. Instead, it utilizes a forced-choice paradigm to elicit obviously correct or incorrect responses. Performance-based validity tests like this may reduce the chance of false positives compared to self-report tests, perhaps because they more objectively assess behavior that is uncharacteristic of DID.

Although the MENT cutoff of ≥8 hovered around the ≥.90 specificity recommendation for use in practice settings (Boone, 2021) and had exceptionally high sensitivity, it is important to consider using cutoffs with even higher specificity for a few reasons. First, simulation designs often elicit methodological problems related to external validity and result in overly optimistic models (Rogers, 1997). Second, specificity should be prioritized because misclassifications of symptom feigning and invalid responding typically have greater consequences than missing the detection of an invalid presentation (Bianchini et al., 2005). This is especially true for forensic-legal settings when dissociation may be used as part of a defense strategy in pleading "not guilty by reason of insanity." Thus, practitioners may wish to use cutoffs with higher specificity on the MENT when making determinations regarding the validity of data that is obtained during a forensic evaluation. A cutoff of ≥10 yielded higher specificity, reducing the misclassification rate to 6%, while maintaining similar sensitivity (.86) when compared to a cutoff of ≥8. If this cutoff of ≥10 exhibits a similar degree of classification accuracy in future studies, that would mean it could potentially serve as an effective validity indicator in forensic practice settings. Practitioners should always interpret elevated MENT scores, regardless of the cutoff chosen, in relation to the individual's medical and clinical history, behavioral presentation, reason for evaluation, and other assessment data (Boone, 2021; Sherman et al., 2020).

Limitations and Future Directions

Although simulation designs are helpful in the inception of cross-validating validity tests for specific populations, they may not adequately represent real-life malingering. We attempted to address this external validity issue and simulate actual malingering by offering educational resources and monetary incentives, which were validated through posttest quizzes and group comparisons. However, the large variability in the simulators' MENT scores suggests that some participants took a less nuanced malingering approach than might be expected in real-life settings. This is a commonly reported issue in simulation studies (Ingram & Ternes, 2016) and may be a result of under incentivization. A $50 incentive is likely substantially less motivating than the high stakes of legal-criminal evaluations where successful malingering may reduce sentencing. It is also possible that including undergraduate students in the simulating group may have biased our findings and exacerbated issues regarding the variability in test-taking approaches. Thus, the high MENT sensitivity values found in this study should be interpreted with caution since they may be overinflated from the lack of skillful and subtle malingering. Future studies may wish to integrate more manipulation checks to ensure practical malingering strategies. For instance, it could be helpful to also examine indicators of careless or haphazard responding or response time to help identify those who were less incentivized by the $50 to effectively simulate DID. To reduce the possibility of DID participants exaggerating symptoms, this study only included participants who were undergoing treatment for DID, uninvolved in litigation, and had a diagnosis verified with an extensive diagnostic interview. Additionally, DID participants produced MENT scores similar to those with PTSD in previous research (Morel & Shepherd, 2008). However, the small sample size restricts the generalizability of these findings. Sensitivity and specificity values found in this study require additional examination using known groups or differential prevalence study designs. It would be informative to also examine how the MENT performs with other validity tests (e.g., TOMM) since multivariable models can optimize the detection of symptom and performance invalidity (Erdodi, 2023; Finley, Brooks, et al., 2023). Future researchers may also wish to explore newer versions of the MENT or compare newer and older versions of the MENT in DID samples. Finally, comparing the number of errors in the DID group with normative data could further our understanding of the MENT's psychometric properties.

Conclusion

These findings provide preliminary evidence that the MENT can be used to assess feigning in DID evaluations. The commonly used cutoff of ≥8 total errors demonstrated sufficient psychometric properties for practical use, but a cutoff of ≥10 yielded an ideal balance between sensitivity and specificity. While the MENT had outstanding classification accuracy, it should be acknowledged that this simulation design may not generalize well to settings where individuals employ more sophisticated malingering approaches. For this reason, practitioners should exercise caution when using the MENT and interpret scores above the ≥10 cutoff, secondary to other empirical validity tests and relevant clinical information. We hope these findings spur further research into cross-validating this cutoff in other DID samples.

References

Ambrose, T., Giarratano, K. J., McCue, M. L., Brand, B. L., & Dalenberg, C. J. (2023). Utility of the Minnesota Multiphasic Personality Inventory-2 in differentiating genuine from feigned dissociative identity disorder. *Psychological Trauma: Theory, Research, Practice, and Policy*. Advance online publication. https://doi.org/10.1037/tra0001611

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425787

Barth, M. R., Brand, B. L., & Nester, M. S. (2023). Distinguishing clinical and simulated dissociative identity disorder using the Miller Forensic Assessment of Symptoms Test. *Psychological Trauma: Theory, Research, Practice, and Policy*, *15*(5), 846–852. https://doi.org/10.1037/tra0001413

Bass, C., & Wade, D. T. (2019). Malingering and factitious disorder. *Practical Neurology*, *19*(2), 96–105. https://doi.org/10.1136/practneurol-2018-001950

Bernstein, E. M., & Putnam, F. W. (1986). Development, reliability, and validity of a dissociation scale. *Journal of Nervous and Mental Disease*, *174*(12), 727–735. https://doi.org/10.1097/00005053-198612000-00004

Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal*, *5*(4), 404–417. https://doi.org/10.1016/j.spinee.2004.11.016

Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, *23*(4), 729–741. https://doi.org/10.1080/13854040802427803

Boone, K. B. (2021). *Assessment of feigned cognitive impairment* (2nd ed.). Guilford Press.

Bourget, D., & Whitehurst, L. (2007). Amnesia and crime. *The Journal of the American Academy of Psychiatry and the Law*, *35*(4), 469–480. https://web.archive.org/web/20170809030225id_/http://jaapl.org/content/jaapl/35/4/469.full.pdf

Brand, B. L., Classen, C., Lanins, R., Loewenstein, R., McNary, S., Pain, C., & Putnam, F. (2009). A naturalistic study of dissociative identity disorder and dissociative disorder not otherwise specified patients treated by community clinicians. *Psychological Trauma: Theory, Research, Practice, and Policy*, *1*(2), 153–171. https://doi.org/10.1037/a0016210

Brand, B. L. (2023). *The concise guide to the assessment and treatment of trauma-related dissociation*. American Psychological Association.

Brand, B. L., Barth, M., Schlumpf, Y. R., Schielke, H., Chalavi, S., Vissia, E. M., Nijenhuis, E. R. S., Jäncke, L., & Reinders, A. A. T. S. (2021). The utility of the Structured Inventory of Malingered Symptomatology for distinguishing individuals with Dissociative Identity Disorder (DID) from DID simulators and healthy controls. *European Journal of Psychotraumatology*, *12*(1), Article 1984048. https://doi.org/10.1080/20008198.2021.1984048

Brand, B. L., & Chasson, G. S. (2015). Distinguishing simulated from genuine dissociative identity disorder on the MMPI-2. *Psychological Trauma: Theory, Research, Practice, and Policy*, *7*(1), 93–101. https://doi.org/10.1037/a0035181

Brand, B. L., Chasson, G. S., Palermo, C. A., Donato, F. M., Rhodes, K. P., & Voorhees, E. F. (2016). MMPI-2 item endorsements in dissociative identity disorder vs. simulators. *The Journal of the American Academy of Psychiatry and the Law*, *44*(1), 63–72. https://www.researchgate.net/publication/298715127_MMPI-2_Item_Endorsements_in_Dissociative_Identity_Disorder_vs_Simulators

Brand, B. L., McNary, S. W., Loewenstein, R. J., Kolos, A. C., & Barr, S. R. (2006). Assessment of genuine and simulated dissociative identity disorder on the structured interview of reported symptoms. *Journal of Trauma & Dissociation*, *7*(1), 63–85. https://doi.org/10.1300/J229v07n01_06

Brand, B. L., Schielke, H. J., Brams, J. S., & DiComo, R. A. (2017). Assessing trauma-related dissociation in forensic contexts: Addressing trauma-related dissociation as a forensic psychologist, part II. *Psychological Injury and Law*, *10*(4), 298–312. https://doi.org/10.1007/s12207-017-9305-7

Brand, B. L., Tursich, M., Tzall, D., & Loewenstein, R. J. (2014). Utility of the SIRS-2 in distinguishing genuine from simulated dissociative identity disorder. *Psychological Trauma: Theory, Research, Practice, and Policy*, *6*(4), 308–317. https://doi.org/10.1037/a0036064

Brand, B. L., Webermann, A. R., Snyder, B. L., & Kaliush, P. R. (2019). Detecting clinical and simulated dissociative identity disorder with the Test of Memory Malingering. *Psychological Trauma: Theory, Research, Practice, and Policy*, *11*(5), 513–520. https://doi.org/10.1037/tra0000405

Carlson, E. B., & Putnam, F. W. (1993). An update on the Dissociative Experiences Scale. *Dissociation*, *6*(1), 16–27.

Coons, P. M., & Milstein, V. (1994). Factitious or malingered multiple personality disorder: Eleven cases. *Dissociation*, *7*, 81–85.

Dalenberg, C. J., Brand, B. L., Gleaves, D. H., Dorahy, M. J., Loewenstein, R. J., Cardeña, E., Frewen, P. A., Carlson, E. B., & Spiegel, D. (2012). Evaluation of the evidence for the trauma and fantasy models of dissociation. *Psychological Bulletin*, *138*(3), 550–588. https://doi.org/10.1037/a0027447

Erdodi, L. A. (2023). Multivariate models of performance validity: The Erdodi Index captures the dual nature of non-credible responding (continuous and categorical). *Assessment*, *30*(5), 1467–1485. https://doi.org/10.1177/10731911221101910

Erdodi, L. A., Roth, R. M., Kirsch, N. L., Lajiness-O'neill, R., & Medoff, B. (2014). Aggregating validity indicators embedded in Conners' CPT-II outperforms individual cutoffs at separating valid from invalid performance in adults with traumatic brain injury. *Archives of Clinical Neuropsychology*, *29*(5), 456–466. https://doi.org/10.1093/arclin/acu026

Finley, J. C. A., Brook, M., Kern, D. M., Reilly, J. L., & Hanlon, R. E. (2023). Profile of embedded validity indicators in criminal defendants with verified valid neuropsychological test performance. *Archives of Clinical Neuropsychology*, *38*(4), 513–524. https://doi.org/10.1093/arclin/acac073

Finley, J. C. A., Brooks, J. M., Nili, A. N., Oh, A., VanLandingham, H. B., Ovsiew, G. P., Ulrich, D. M., Resch, Z. J., & Soble, J. R. (2023). Multivariate examination of embedded indicators of performance validity for ADHD evaluations: A targeted approach. *Applied Neuropsychology: Adult*. Advance online publication. https://doi.org/10.1080/23279095.2023.2256440

Finley, J. C. A., Cerny, B., Obolsky, M., Brooks, J., Haneda, A., Ovsiew, G., Ulrich, D. M., Resch, Z. J., & Soble, J. (2024). Cross-validating the clinical assessment of attention deficit-adult symptom validity scales in adults with attention deficit/hyperactivity disorder. *Journal of Clinical and Experimental Neuropsychology*, *46*(2), 111–123. https://doi.org/10.1080/13803395.2023.2283940

Finley, J. C. A., Cladek, A., Gonzalez, C., & Brook, M. (2024). Perceived cognitive impairment is related to internalizing psychopathology but unrelated to objective cognitive performance among nongeriatric adults presenting for outpatient neuropsychological evaluation. *The Clinical Neuropsychologist*, *38*(3), 644–667. https://doi.org/10.1080/13854046.2023.2241190

Foote, B., & Park, J. (2008). Dissociative identity disorder and schizophrenia: Differential diagnosis and theoretical issues. *Current Psychiatry Reports*, *10*(3), 217–222. https://doi.org/10.1007/s11920-008-0036-z

Frewen, P., & Lanius, R. (2015). *Healing the traumatized self: Consciousness, neuroscience, treatment*. W.W. Norton.

Gonzalez, C., Finley, J. A., Khalid, E., Basurto, K. S., VanLandingham, H. B., Frick, L. A., Brooks, J. M., Ellison, R. L., Ulrich, D. M., Soble, J. R., & Resch, Z. J. (2024). The impact of adverse childhood experiences on symptom and performance validity tests among a multiracial sample presenting for ADHD evaluation. *Archives of Clinical Neuropsychology*. Advance online publication. https://doi.org/10.1093/arclin/acae006

González, H., Santamaría, P., & Fernández, P. (2010). Precisión predictive del Inventario Estructurado de Simulación de Síntomas—SIMS en el contexto medicolegal [Predictive accuracy of the Structured Inventory of Malingered Symptomatology—SIMS in the medicolegal setting]. *EduPsykhé:Revista de Psicología y Psicopedagogía*, *9*(1), 3–22. http://repositorio.ucjc.edu/bitstream/handle/20.500.12020/17/C00042383.pdf?sequence=1

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley. https://doi.org/10.1002/9781118548387

Ingram, P. B., & Ternes, M. S. (2016). The detection of content-based invalid responding: A meta-analysis of the MMPI-2-Restructured Form's (MMPI-2-RF) over-reporting validity scales. *The Clinical Neuropsychologist*, *30*(4), 473–496. https://doi.org/10.1080/13854046.2016.1187769

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*(4), 625–630. https://doi.org/10.1017/S1355617712000240

Merckelbach, H., Boskovic, I., Pesy, D., Dalsklev, M., & Lynn, S. J. (2017). Symptom overreporting and dissociative experiences: A qualitative review. *Consciousness and Cognition*, *49*, 132–144. https://doi.org/10.1016/j.concog.2017.01.007

Merckelbach, H., Giesbrecht, T., van Heugten-van der Kloet, D., de Jong, J., Meyer, T., & Rietman, K. (2015). The overlap between dissociative symptoms and symptom over-reporting. *European Journal of Psychiatry*, *29*(3), 165–172. https://doi.org/10.4321/S0213-61632015000300001

Merten, T., Kaminski, A., & Pfeiffer, W. (2020). Prevalence of overreporting on symptom validity tests in a large sample of psychosomatic rehabilitation inpatients. *The Clinical Neuropsychologist*, *34*(5), 1004–1024. https://doi.org/10.1080/13854046.2019.1694073

Merten, T., & Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law*, *6*(2), 122–137. https://doi.org/10.1007/s12207-013-9155-x

Miele, A. S., Gunner, J. H., Lynch, J. K., & McCaffrey, R. J. (2012). Are embedded validity indices equivalent to free-standing symptom validity tests? *Archives of Clinical Neuropsychology*, *27*(1), 10–22. https://doi.org/10.1093/arclin/acr084

Miller, H. A. (2000). *M-FAST: Miller forensic assessment of symptoms test professional manual*. Psychological Assessment Resources.

Morel, K. R. (1998). Development and preliminary validation of a forced-choice test of response bias for posttraumatic stress disorder. *Journal of Personality Assessment*, *70*(2), 299–314. https://doi.org/10.1207/s15327752jpa7002_8

Morel, K. R. (2008). Comparison of the Morel Emotional Numbing Test for posttraumatic stress disorder to the word memory test in neuropsychological evaluations. *The Clinical Neuropsychologist*, *22*(2), 350–362. https://doi.org/10.1080/13854040701300531

Morel, K. R., & Shepherd, B. E. (2008). Meta-analysis of the Morel Emotional Numbing Test for PTSD: Comment on Singh, Avasthi, and Grover. *German Journal of Psychiatry*, *11*(3), 128–131. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ce8c15cec12d908d6e82f3010d7d67c2605d3e70

Mychailyszyn, M. P., Brand, B. L., Webermann, A. R., Şar, V., & Draijer, N. (2021). Differentiating dissociative from non-dissociative disorders: A meta-analysis of the structured clinical interview for DSM dissociative disorders (SCID-D). *Journal of Trauma & Dissociation*, *22*(1), 19–34. https://doi.org/10.1080/15299732.2020.1760169

Palermo, C. A., & Brand, B. L. (2019). Can the Trauma Symptom Inventory-2 distinguish coached simulators from dissociative disorder patients? *Psychological Trauma: Theory, Research, Practice, and Policy*, *11*(5), 477–485. https://doi.org/10.1037/tra0000382

Pilton, M., Varese, F., Berry, K., & Bucci, S. (2015). The relationship between dissociation and voices: A systematic literature review and meta-analysis. *Clinical Psychology Review*, *40*, 138–155. https://doi.org/10.1016/j.cpr.2015.06.004

Reinders, A. A. T. S., & Veltman, D. J. (2021). Dissociative identity disorder: Out of the shadows at last? *The British Journal of Psychiatry*, *219*(2), 413–414. https://doi.org/10.1192/bjp.2020.168

Rogers, R. (1997). Researching dissimulation. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 398–426). Guilford Press.

Rogers, R. (2018). Detection strategies for malingering and defensiveness. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 14–35). Guilford Press.

Rogers, R., Payne, J. W., Correa, A. A., Gillard, N. D., & Ross, C. A. (2009). A study of the SIRS with severely traumatized patients. *Journal of Personality Assessment*, *91*(5), 429–438. https://doi.org/10.1080/00223890903087745

Sellbom, M., Wygant, D., & Bagby, M. (2012). Utility of the MMPI-2-RF in detecting non-credible somatic complaints. *Psychiatry Research*, *197*(3), 295–301. https://doi.org/10.1016/j.psychres.2011.12.043

Sharf, A. J., Rogers, R., Williams, M. M., & Henry, S. A. (2017). The effectiveness of the MMPI-2-RF in detecting feigned mental disorders and cognitive deficits: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment*, *39*(3), 441–455. https://doi.org/10.1007/s10862-017-9590-1

Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology*, *35*(6), 735–764. https://doi.org/10.1093/arclin/acaa019

Shura, R. D., Ord, A. S., & Worthen, M. D. (2022). Structured inventory of malingered symptomatology: A psychometric review. *Psychological Injury and Law*, *15*(1), 64–78. https://doi.org/10.1007/s12207-021-09432-y

Simotas, S. (2000). Memory and clinical features in college students with high and low dissociative tendencies. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *60*, 35–80. https://www.proquest.com/docview/304512340?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses

Steinberg, M. (1994). *Structured clinical interview for DSM-IV dissociative disorders (SCID-D)*. American Psychiatric Press.

Steinberg, M. (2000). Advances in the clinical assessment of dissociation: The SCID-D-R. *Bulletin of the Menninger Clinic*, *64*(2), 146–163.

Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, *9*(3), 260–268. https://doi.org/10.1037/1040-3590.9.3.260

Tse, P. K. Y., Finley, J.-C., Frick, L., Guilfoyle, J., Brooks, J., Khalid, E., Charara, R., Resch, Z. J., Ulrich, D. M., Ovsiew, G. P., & Soble, J. R. (2023). Cross-validating the embedded performance validity indicators in the Rey Auditory Verbal Learning Test in mixed neuropsychiatric and attention-deficit/hyperactivity disorder clinical samples. *Psychology & Neuroscience*, *16*(2), 125–137. https://doi.org/10.1037/pne0000302

Widows, M. R., & Smith, G. P. (2005). *Structured Inventory of Malingered Symptomatology*. Psychological Assessment Resources.

Received November 1, 2023
Revision received February 21, 2024
Accepted May 2, 2024 ■