Psychological Injury and Law
https://doi.org/10.1007/s12207-021-09430-0

Memory Complaints Inventory: Review of Psychometric Properties

Patrick Armistead-Jehle[1] · **Robert D. Shura**[2,3,4]

Received: 11 June 2021 / Accepted: 23 October 2021

Abstract
The Memory Complaints Inventory (Green in *Users' manual for the Memory Complaints Inventory (MCI)*. 2019) is a 58-item, stand-alone symptom validity test that measures exaggerated memory complaints. Psychometric properties of the MCI are adequate, and the test manual provides foundational validation data in large samples, which have been replicated in several independent studies. Scoring software and the optional AI program offer an array of score reporting options. MCI scores tend to be moderately relate to performance validity scores on memory-based measures, with stronger relationships to other cognitive symptom validity scales. Additionally, MCI scores tend to be high in those with non-neurological disorders (e.g., depression, pain), and MCI scores were not related to scores on performance-based memory tests. The current paper reviews all studies on the MCI currently published and synthesizes a recommended approach for interpreting the MCI. Strengths, weakness, and areas for future research are also reviewed.

Keywords Memory complaints inventory · Review · Symptom validity testing

Backgrounds and Conditions of Use

Measure Overview

The Memory Complaints Inventory (MCI; Green, 2019) is a 58-item computer administered self-report questionnaire of subjective memory complaints. Originally developed as a DOS program in 1996 by Dr. Paul Green, it was converted to a Microsoft Windows program in 2003. Per the test manual, the MCI allows for (1) standardized data collection regarding self-reported memory problems, (2) quantification of memory complaints into different categories, and (3) comparison of individual responses to various comparison groups (Green, 2019). Item content was based on the author's clinical experience of expressed patient memory concerns over decades of practice.

✉ Patrick Armistead-Jehle
Patrick.j.armistead-jehle.civ@mail.mil

1 Munson Army Health Center, Fort Leavenworth, KS, USA

2 VA Mid-Atlantic Mental Illness Research, Education, and Clinical Center (MA-MIRECC), Salisbury, NC, USA

3 Research & Academic Affairs Service Line, Salisbury Veterans Affairs Medical Center, Salisbury, NC, USA

4 Department of Neurology, Wake Forest School of Medicine, Winston-Salem, NC, USA

The respondent completing the MCI rates statements related to potential memory problems over the last month on a 0 (*Not at All*) to 4 (*Extremely*) scale. The measure consists of nine scales rationally designed to tap specific types of reported memory problems: General Memory Problems (GMP), Numeric Information Problems (NIP), Visuospatial Memory Problems (VSMP), Verbal Memory Problems (VMP), Pain Interferes with Memory (PIM), Memory Interferes with Work (MIW), Impairment of Remote Memory (IRM), Amnesia for Complex Behavior (ACB), and Amnesia for Antisocial Behavior (AAB). The first six scales include plausible memory complaints. In contrast, the latter three scales were intentionally designed to consist of items that would be considered implausible for most individuals with memory problems secondary to an organic etiology. The endorsement of such symptoms in clinical practice is typically thought to reflect either psychiatric origins or exaggerated/feigned memory complaints. The MCI is scored as a percentage of the maximum possible score on each of the nine scales.

Concerning factor structure, per the test manual (Green, 2019), all scales loaded onto one factor. This was considered consistent with the high correlations among scales, with correlations between the average of all MCI scores and each individual subscale ranging from 0.91 (GMP) to 0.63 (AAB). A recent study by Disner, Mattson, Nelson, and Armistead-Jehle (in press) evaluated MCI data in 699

Published online: 11 November 2021

service members. MCI data was fit to a bifactor confirmatory factor analysis (CFA). The CFA was composed of an "Overall MCI factor," which loaded onto all MCI scales and was consistent with the one factor solution outlined in the test manual. Disner et al. also found a residual factor, which emphasized the variance uniquely associated with the IRM, ACB, and AAB subscales. Adequate model fit was defined as the root mean square of approximation < 0.08, standardized root mean residual < 0.08, and comparative fit index > 0.90. Overall, data from these two factor analyses appear to complement each other and ultimately support a single main factor with a residual factor consistent with the original theoretical conceptualization of the measure (e.g., plausible versus implausible scales).

Appropriate Populations

As outlined in the test manual, there are two large samples by which the MCI was developed. Sample 1 data was obtained from 1728 adult patients seen consecutively for outpatient neuropsychological evaluation in the author's (Dr. Green) private practice over an 18-year period (Green, 2019). A variety of primary diagnoses are listed for Sample 1, with mild TBI at the highest frequency. Unfortunately, the manual does not provide further details of the standardization sample. Sample 2 data were seen independently by Dr. Roger Gervais and consisted of consecutive outpatients seen for compensation eligibility. Sample 2 is noted to be comprised predominantly of patients diagnosed with chronic pain, major depression, posttraumatic stress disorder, or orthopedic injuries, with reference that additional information is available in Gervais et al. (2007) RBS validation paper. The largest apparent differences are the predominance of mTBI in Sample 1 and the disability nature of Sample 2.

To date, there are no known studies of this measure with children or adolescents. This is logical as some of the MCI items are clearly designed for adults (e.g., items related to transportation or occupation) and would not lend themselves well to pediatric patients. The current version of the MCI is available in six languages (English, Dutch, Spanish, French, Portuguese, and German). The manual provides no information on how these translations were made. The German version was translated via low-key adaptation (without using formal blinded back-translation) circa 2004 by Dr. Thomas Merten (T. Merten, Personal Communication, August 25, 2021). Unfortunately, there is no available information on how the MCI was translated into the other available languages. To date, there are no studies utilizing the translated versions of the MCI. One study was published in a German language journal, but the English version of the MCI was employed (Green et al., 2005). This study was largely descriptive in nature, and data came from a subset of the sample ultimately used for analysis in the manual.

Program Reports of Results

The MCI program includes three options for reporting examinee results. The per-question report allows the user to see all examinee responses grouped by scale, order of presentation, or the degree to which they were endorsed. The comparative report allows examinee responses to be compared to various clinical groups as a function of MCI subtest. The default on this report is a "best fit" option (showing groups with mean scores that are the closest to the respondent); however, the user can custom select groups. Finally, the build chart option produces a line graph of examinee subtest scores compared to selected groups.

MCI scores can also be reported in Green's Advanced Interpretation (AI) program (Green, 2008a, b). Data from the MCI are automatically collected by the AI program if the AI and MCI are installed on the same computer. Two MCI reports can be generated within the AI program, a complete report and a condensed report. The complete report includes all eight reporting options available in AI for the MCI (Table 1). The condensed report allows the user to select which reporting options he/she wishes to view.

Convergent and Incremental Validity

Manual

As noted above, the MCI manual (Green, 2019) presents results from two different clinical samples. Again, the primary sample (Sample 1) was derived from 1728 outpatients seen in Dr. Green's private practice. A second sample (Sample 2) consisted of 1212 examinees from Dr. Roger Gervais' private practice. Across these samples, four main themes emerge. First, the MCI scales (or average of all MCI scales) are not elevated in those with neurological disease, including severe TBI. However, MCI scores are elevated in samples of those diagnosed with psychiatric disorders (most notably depression). This reflects the trend where individuals with psychiatric conditions present with more memory complaints than those diagnosed with neurological disease and severe TBI (Rohling et al., 2002; Smith et al., 1996). In those with a primary depressive diagnosis from Sample 1 ($n = 193$), the correlation between mean MCI and the Beck Depression Inventory total score was $r = 0.53$ (both before and after removing those with invalid Word Memory Test (Green, 2003) scores). This is further displayed when MCI scores are presented by diagnosis. The highest mean MCI scores were seen in chronic fatigue syndrome (42.2%), chronic pain syndrome (38.0%), and mild TBI (34.5%), while the lowest mean MCI scores were observed in police applicants (2.9%), anxiety disorders (22.5%), and orthopedic injuries (24.4%). Of note, although mild traumatic brain injury (TBI) mean MCI was 34.5%, moderate/severe TBI was 26.1%, thus demonstrating an inverted

Table 1 MCI reporting options in AI program

Reporting option	Definition
1. Overall memory complaints and levels of effort on PVTs	Level of WMT score associated with the MCI GMP subtest per data in Armistead-Jehle et al. (2012a)
2. Overall level of memory complaints relative to healthy adults	Mean MCI score reported as a function of SD above or below the healthy adult mean
3. Overall level of memory complaints relative to patient groups (weighted average values)	The program chooses five groups with MCI profiles most similar to the examinee
4. Overall level of memory complaints relative to patients with a history of severe TBI	Mean MCI score reported as a function of SD above or below patients with a history of severe TBI
5. Scores on plausible and implausible MCI subscales	Mean MCI scores on plausible and implausible scales compared to health adults and severe TBI groups
6. Top four scale scores	Top four MCI subscale scores as a function of SDs from the healthy adult subscale mean
7. Notable item endorsement	Items rarely endorsed by those with genuine neurological disease or items potentially requiring follow up questioning
8. *Z*-score table	Examinee scores in terms of SDs above or below an identified diagnostic group

MCI Memory Complaints Inventory, *AI* Advanced Interpretation, *TBI* traumatic brain injury, *PVT* performance validity test

relationship to head injury severity. This was further highlighted in Sample 1 via brain scan data. Mean MCI scores for normal scans were 38% compared to abnormal scans of 27%. In sum, MCI scores do not relate to neurological disease, which would be expected to result in a higher risk of cognitive impairment; rather, MCI scores are elevated in those with psychiatric diagnoses.

The second main theme culled from the MCI manual is the relationship between MCI scores and performances on measures of verbal memory. More specifically, in those with valid performance validity test (PVT) scores, the MCI is not related to performance on objective measures of verbal memory as assessed by the California Verbal Learning Test (CVLT; Delis et al., 1987). Thus, a divergence between self-reported memory complaints and objectively measured memory performance was established. In Sample 2, after excluding invalid WMT (leaving $n = 1{,}099$), the correlation between VMP and CVLT 1–5 Total was $r = -0.02$. These data indicate that in very simple terms, the MCI is not a memory test.

Third, there is a moderate negative correlation between MCI and memory-based PVTs. In Sample 1, the mean MCI score correlated to WMT scores at IR $r = -0.42$, DR $r = -0.44$, and CNS $r = -0.45$. Additionally, as scores increased on the WMT (i.e., trend towards valid performance), individuals endorsed fewer complaints on the MCI. More specifically, in those with a mean easy subtest WMT score of 91 to 100, the mean MCI score was 25%, and in those with a mean easy subtest WMT score of 60 or lower, the mean MCI score was 55% (Green, 2019). None of the individuals with mean MCI score of < 10% produced invalid scores on the WMT; however, 76% of those with mean MCI scores of 70% + produced invalid scores on the WMT. Although the relationship between the MCI and WMT is not perfect (nor should it be given that these instruments measure somewhat different constructs), the relationship of the MCI to the WMT ($r = -0.42$ to -0.45) is far greater than to the CVLT 1–5 Total score ($r = -0.02$), further establishing validity.

Finally, one should expect the MCI scores to show a stronger relationship to other memory-related symptom validity tests (SVTs) than to performance validity tests. Initial convergent validity data are presented with Sample 2 (Green, 2019). With other SVTs purporting to measure exaggerated memory complaints/impairment ($n = 1550$), the mean MCI score correlated mostly strongly with Response Bias Scale (RBS; Gervais et al., 2007) on Minnesota Multiphasic Personality Inventory, 2nd edition Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008), with $r = 0.69$. The mean MCI scores were however also moderately correlated with other MMPI-2-RF over reporting scales. More specifically, the correlations F-r, Fs-r, and FBS were 0.63, 0.54, and 0.53, respectively. Among the MMPI-2-RF over-reporting scales, the lowest mean MCI scale correlation came with Fp-r ($r = 0.40$). As anticipated the mean MCI score was negatively correlated with MMPI-2-RF under-reporting scales (L-r = −0.04 and K-r = −0.38). The relationships among MCI scores and both PVTs and other SVTs are further established in independent studies published since the release of the MCI, which will now be reviewed.

Published Papers

MCI Convergence to Symptom Validity Tests

At the time of writing the current manuscript, there have been five published studies with data pertinent to the validity of the MCI. The first studies of this kind were completed by

Gervais and colleagues across two articles. First, Gervais et al. (2008) compared MMPI-2 RBS scores to MCI scales in a sample of 1550 adult examinees referred for disability evaluation. RBS scores were strongly correlated with each individual MCI scale, as well as the mean of all MCI scales ($r = 0.69$). Overall, MCI scales were more strongly correlated with the RBS relative to other MMPI-2 over-reporting scales (F, Fb, Fp, and FBS). Moreover, ANOVA results contrasting each MCI scale and the mean MCI score demonstrated significantly higher reported memory problems across increasing RBS T-score ranges with large effect sizes (η^2) ranging from 0.19 to 0.46. In a subsequent study, Gervais et al. (2010) evaluated the MMPI-2-RF RBS to the MCI in a sample of disability seeking adults. Among 908 subjects, the RBS correlated more strongly with each MCI scale (save AAB) and the mean of MCI scales than other MMPI-2-RF over-reporting scales (F-r, Fp-r, Fs, and FBS-r). Moreover, after controlling for performance validity via the WMT ($n = 823$), the researchers correlated that mean MCI score with MMPI-2-RF over-reporting scales. Among these scales, the MCI correlated highest with the RBS ($r = 0.63$). Armistead-Jehle et al. (2016) evaluated the classification statistics of the MCI relative to PVT and SVTs in a sample of 339 active duty service members with a remote history of concussion. For SVTs, this study examined the Personality Assessment Inventory (PAI; Morey, 1991) Negative Impression Management Scale (NIM), Malingering Index (MAL) and Rogers Discriminant Function, and several MMPI-2-RF over-reporting scales (F-r, Fp-r, Fs, FBS-r, RBS). The PAI NIM and MAL were significantly correlated with all MCI scales, as were all MMPI-2-RF over-reporting scales. MCI scales were not significantly correlated the PAI Rogers Discriminant Function scale. ROC analysis for the mean of the MCI scales and the PAI and MMPI-2-RF SVT evidence AUCs ranging from 0.77 (FBS-r) to 0.86 (F-r). Overall, this study demonstrated improved classification statistics for the MCI on SVTs relative to PVTs (described below).

MCI Convergence to Performance Validity Tests

Armistead-Jehle et al. (2012a) evaluated MCI scores as a function of performance validity testing in two large samples of individuals referred for disability evaluations (Ns = 1597 and 2118). The study included four stand-alone PVTs (WMT, Medical Symptom Validity Test [MSVT; Green, 2004], Non Verbal Medical Symptom Validity Test [NV-MSVT; Green, 2008a, b], and the Test of Memory Malingering [TOMM; Tombaugh, 1996]) and two embedded PVTs (Reliable Digit Span [Greiffenstein et al., 1994] and CVLT Recognition Hits [Curtis et al., 2006]). The data across both samples clearly demonstrated that as disability seeking examinees' performances across PVTs worsened, there was a corresponding increase in memory complaints on all MCI subscales. For instance, when comparing WMT easy subtest scores between 91–100% and 71–80% and WMT easy subtest scores between 91–100% and < 51%, effect sizes for the mean of the MCI subscales ranged from medium ($d = 0.52$) to large ($d = 1.93$), respectively.

A replication study by Armistead-Jehle et al. (2012b) compared MCI scores to various PVTs (MSVT, NV-MSVT, TOMM, and the Effort Index) (Silverberg et al., 2007) from the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998) in a clinical sample. This study consisted of 191 active and retired military service members and their adult family members without overt evidence of secondary gain. Overall, these data showed that as examinees' scores across PVTs declined, there was a corresponding increase in subjective memory complaints on the MCI. However, this study differed from the original in two ways. First, in the non-disability seeking military sample, the AAB and ACB subscales were less consistently influenced by poor PVT performance, which stands in contrast the disability-seeking, civilian sample in Armistead-Jehle et al. (2012a) study. Second, the embedded PVTs in the 2012b study had a much less robust relationship to MCI scores, relative to embedded PVTs in the 2012a study. This may be a function of varying sensitivities of the measures used. Nonetheless, these relationships between MCI scores and PVT performances across both Armistead-Jehle et al., 2012a, b studies demonstrate convergent validity for the MCI.

As noted above, Armistead-Jehle et al. (2016) evaluated the MCI in relation to both SVTs and PVTs. With regard to PVTs, relative to those who passed the MSVT and NV-MSVT, those who failed had significantly higher scores on all MCI scales. Effects sizes (d) for the mean of the MCI scales on the MSVT and NV-MSVT were 0.94 and 0.78, respectively. AUC for the MSVT was 0.75 (95% CI = 0.70–0.80) and 0.72 (95% CI = 0.67–0.77) for the NV-MSVT.

MCI in Criterion Group Studies

In addition to studies related to MCI validity, a handful of investigations have employed the MCI as either an outcome or a criterion measure. Rohling et al. (2002) examined the impact of depression on cognitive performances in patients who passed PVTs. The authors employed a range of cognitive ability measures and several self-report instruments to include the MCI. In their disability sample of 420 subjects, no differences were found on objective cognitive performances in examinees with depression. However, self-report measures indicated that patients with higher levels of depression reported an elevated degree of emotional, somatic, and cognitive problems. All MCI scales differed significantly between the low and high depression groups, with moderate to large effect sizes ranging from *Hedges' g* = −0.82

to −1.34. Next, Elias et al. (2019) employed the MCI as one of several measures to inform group membership in a study of exaggerated functional impairment following mild TBI (disability sample, $N = 76$). The MCI and several standalone PVTs were used as the basis for determining Slick criteria in their sample (Slick et al., 1999). Subjects assigned to the probable malingering group evidenced considerably higher mean total MCI scores (41.5 [$SD = 12.7$]) relative to the two non-malingering groups (12.5 [$SD = 10.2$] and 21.0 [$SD = 15.0$]). Lastly, Denby et al. (2019) sought to estimate the long-term impact of vestibular dysfunction on neurobehavioral functioning and disability in UK military veterans. Among their criterion measures, the MCI was employed to assess symptom exaggeration. Fifty-four of the 110 subjects with a history of at least one mild TBI exceeded the MCI cut off score for symptom exaggeration as defined by the authors (< 40%). These three independent studies further establish validity of the MCI in mTBI and depression samples, with results as expected based on findings in the manual (Green, 2019).

Cut Scores and Hit Rates

MCI Compared to Performance Validity Tests

Several studies have evaluated diagnostic accuracy for various cutoff scores for the MCI compared to PVT results. Common methodology of these studies is to use the mean of the nine MCI scales when predicting a given criterion. Initial analyses are included in the manual using Samples 1 and 2 compared to invalid scores on the WMT. When those two samples were pooled together ($n = 4397$), a mean MCI score of 56% or more was associated with WMT failure at sensitivity = 0.24 and specificity = 0.95. The manual additionally presents AUC data for the MCI predicting invalidity on several individual PVTs. AUC values in Sample 1 were as follows: MSVT AUC = 0.71 ($n = 1731$), WMT AUC = 0.75 ($n = 1731$), NV-MSVT AUC = 0.77 ($n = 525$), CARB AUC = 0.81 ($n = 1101$), and TOMM AUC = 0.83 ($n = 297$). AUC values in Sample 2 were similar in magnitude: WMT AUC = 0.73 ($n = 2{,}666$), NV-MSVT AUC = 0.77 ($n = 868$), MSVT AUC = 0.79 ($n = 1146$), CARB AUC = 0.81 ($n = 1512$), and TOMM AUC = 0.85 ($n = 1603$).

Two other studies evaluated mean MCI scores in clinical samples. As noted above, Armistead-Jehle et al. (2016) evaluated the MCI mean score in a sample of active duty outpatients and against both the MSVT and NV-MSVT. For both PVTs, AUCs were in the acceptable range, but sensitivities were poorer than for results predicting the WMT in the MCI manual. Huber et al. (2020) used a mix clinical sample and criterion of invalid on two or more stand-alone PVTs. Sensitivity of the mean MCI score had the highest sensitivity of all four samples. Table 2 presents diagnostic accuracy findings for mean MCI scores across the two manual samples and the two independent samples. In general, a mean MCI cutoff score in the range of 40 to 50% appears best for maximizing sensitivity while maintaining specificity at 0.90 or above.

One possible factor contributing to the lower sensitivities in the above studies is the use of an overall mean score. As described above, the first six scales contain items reflecting plausible memory complaints, whereas the last three scales are implausible complaints. Combining across these two types of scales might dilute relationships among scales and PVTs, especially for the implausible scales. In the manual, the plausible scales mean in Sample 1 was 37.2%, which was over double the mean of implausible scales (17.0%). The study by Huber et al. (2020) also examined these two means, both resulting in AUCs of 0.69 predicting the group

Table 2 Diagnostic accuracy of mean MCI scores compared to performance validity tests

	Green (2019) Sample 1	Green (2019) Sample 2	Armistead-Jehle et al. (2016)	Huber et al. (2020)
N	1731	2666	339	244
Sample	Civilian; mixed clinical outpatients	Civilian; civil forensic outpatients	Army active duty outpatients	Civilian; mixed clinical outpatients
Criterion	WMT	WMT	MSVT[1] NV-MSVT[2]	Invalid on ≥ 2: WMT, MSVT, NV-MSVT, TOMM
AUC	0.75	0.73	0.75[1] 0.72[2]	0.69[a]
Cutoff	52%	47%	50%[1,2]	40%
Sen/Spec	0.33/0.90	0.37/0.90	0.23/0.96[1] 0.16/0.93[2]	0.65/0.90

MCI Memory Complaints Inventory, *WMT* Word Memory Test, *MSVT* Medical Symptom Validity Test, *NV-MSVT* Non-Verbal Medical Symptom Validity Test, *TOMM* Test of Memory Malingering

[a]Partial AUC

[1]MSVT; [2]NV-MSVT

with invalid scores on two or more PVTs. Like Sample 1 in the manual, the cutoff for the plausible mean (53%, sensitivity = 0.65, specificity = 0.89) was far higher than the implausible mean (30%, sensitivity = 0.45, specificity = 0.90). In other words, the threshold of invalid is far lower for the implausible scores given the infrequency for endorsement of those items.

In addition to evaluating the three multi-scale mean MCI scores, the manual presents data across MCI subtests compared to WMT results (Green, 2019). In Sample 1, AUC values ranged from 0.67 (IRM) to 0.74 (VSMP). Sample 2 AUCs ranged from 0.65 (AAB) to 0.71 (NIP, VSMP, MIW, and ACB). Despite the WMT involving recognition of word pairs (verbal), the VSMP best predicted invalid WMT scores. In contrast, Huber et al. (2020) found the best diagnostic accuracy for the GMP scale (AUC = 0.69; sensitivity = 0.55 and specificity = 0.91 for cutoff score of 55%) and worst for the AAB scale (AUC = 0.58; sensitivity = 0.31 and specificity = 0.91 for cutoff score of 19%). Also somewhat discrepant from the manual results was the military sample results (Armistead-Jehle et al., 2016), which evaluated various cutoff scores from ≥ 10 to ≥ 50% at 10% intervals. Compared to the MSVT, the VMP had the best diagnostic accuracy while maximizing specificity (sensitivity = 0.75, specificity = 0.56 at ≥ 50%) and the PIM had the worst sensitivity, but reached adequate specificity (sensitivity = 0.22, specificity = 0.91 at ≥ 50%). Like the MSVT, the VMP also had the best sensitivity for the NV-MSVT (sensitivity = 0.82, specificity = 0.54 at ≥ 50%); however, the AAB showed the lowest sensitivity of all (sensitivity = 0.12, specificity = 0.96 at ≥ 30%). Synthesizing across all the samples, the VSMP, VMP, and GMP performed the best when comparing to PVTs. Of note, the implausible scales tended to result in lower accuracy than the plausible scales. One hypothesis is that the general, visual, and verbal plausible complaints may be more closely related to the memory task in PVTs, in a face valid manner. In other words, the items on those scales are a self-report proxy for the types of tasks performed during PVTs (short-term verbal/visual recognition). In contrast, it is difficult to create a performance-based proxy of the items on the implausible scales (e.g., remote memory impairment).

MCI Compared to Symptom Validity Tests

Although the MCI has shown utility in predicting invalid scores on memory-based PVTs, (admittedly with somewhat lower sensitivity), the MCI is conceptually a symptom validity test. One might then expect improved diagnostic accuracies if compared to another SVT designed for identifying exaggerated self-reported memory complaints. The problem is that there are few memory-based SVTs available, and other cognitive SVTs (e.g., MMPI-2/RF RBS) have used the MCI itself as the criterion for comparison. Of note, the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005) is a 75-items, stand-alone SVT that includes a 15-item Amnestic Disorders (AM) scale; however, no studies were found comparing the MCI and SIMS AM.

The MMPI-family of instruments includes both the RBS (Gervais et al., 2007) on the MMPI-2-RF/3, which was designed to predict those with invalid scores on PVTs (WMT, Computerized Assessment of Response Bias [CARB; Allen et al., 1997], and/or TOMM), and the Symptom Validity Scale (FBS) that was designed to sample exaggerated cognitive and somatic complaints more commonly exaggerated in civil litigation contexts. Armistead-Jehle et al. (2016) specifically set out to assess diagnostic accuracy of the MCI compared to both PVTs and SVTs, hypothesizing a stronger relationship to the SVT scales. Results confirmed the hypothesis using the mean MCI scores. Whereas AUCs for PVTs were 0.72 and 0.75, AUCs across the five overreporting scales of the MMPI-2-RF ranged from 0.77 (FBS-r at ≥ 40% MCI cutoff: sensitivity = 0.48, specificity = 0.90) to 0.86 (F-r at MCI cutoff ≥ 40%: sensitivity = 0.56, specificity = 0.93), and the AUC for the PAI was 0.85 (NIM at MCI cutoff ≥ 50%: sensitivity = 0.44, specificity = 0.96). Interestingly, the highest AUC was for MMPI-2-RF F-r, a scale of overreporting psychopathology, not to RBS or FBS-r. One possibility for this relationship is that those exaggerating psychopathology (e.g., depression) might commonly include exaggerated memory complaints as part of that presentation. Indeed, as the manual indicates, the depression group of Sample 1 produced a mean MCI score of 32.7%, which was the fifth highest group and a score higher than both neurological illness and moderate/severe TBI groups. In sum, results highlight that although the MCI is significantly related to memory-based PVT scores, the MCI better predicts overreporting than underperformance.

Interpretive Approach

The MCI manual does not provide hard rules on cutoff scores for MCI scales, but instead emphasizes interpreting scores in comparison to various clinical groups. The AI program can provide best fit profiles across dozens of identified groups to assist in this process. Green also notes that if mean MCI scores are over 30%, the two main hypotheses are either symptoms exaggeration (which can be confirmed by other evidence) or emotional explanations. Synthesizing across studies reviewed above, the following approach can be considered to supplement results produced by the AI software or MCI program itself:

1. Mean MCI score ≥ 50% indicates globally exaggerated memory complaints, with increasingly likelihood of also producing invalid scores on memory-based PVTs.
2. Plausible mean ≥ 50% suggests exaggerated memory complaints: Implausible mean ≥ 30% suggests fabricated memory complaints.
3. Plausible subtest scores ≥ 50% indicates exaggerated complaints specific to that category (e.g., verbal or visual memory): Implausible subtest scores ≥ 30% indicate fabricated complaints specific to that category (remote memory, or complex or antisocial behaviors).
4. Interpretation of *z* score comparisons produced by the scoring software or AI allows for contextualizing of the results based on those who are valid, neurologically-impaired, and/or invalid on PVTs.

Strengths and Weaknesses

Strengths

As with any measure, the MCI has both strengths and weaknesses. The most obvious strength is that the MCI is among very few stand-alone SVTs available. A stand-alone format allows more flexibility for situations where, for example, administration of a lengthy multi-scale, self-report measure such as an MMPI or PAI is neither feasible nor ideal. Related, administration time is about 5 min, making the MCI one of the least burdensome SVTs available to administer. Scoring is also relatively immediate, and included score reports do not use a credit based system (though, AI reports are based on credits), further making the measure cost-efficient. Other useful features of the software program include availability in six languages (English, French, Spanish, Dutch, Portuguese, and German), three different report formats (item-level report, comparative report, and chart format), increased ease of interpretation with the AI program, and the over 50 available comparison groups for interpretation, which can be used either in the software-forced best fit or for customized selections. In addition to aspects related to practicality, administration, and scoring, initial validation research in the manual involves large sample sizes. As noted, psychometrics appear adequate, and the measure has been validated against a range of both PVTs and SVT scales in several independent samples.

Based on these findings, commentary on the clinical utility of the MCI is warranted. As there are already a number of well-validated PVTs that can reliably distinguish between credible and non-credible memory deficits (WMT, MSVT, NV-MSVT, TOMM, etc.), one may ask what, then, is the contribution of the MCI to clinical or forensic evaluations? The answer lies in the distinction between performance and symptom validity testing. Several studies have demonstrated that PVTs and SVTs measure distinct, but related constructs and that the administration of both are necessary for a comprehensive evaluation of neuropsychological functioning (see for example, Sabelli et al., 2021; Ord et al., 2021). This is further emphasized in the updated Slick criteria for malingering neuropsychological dysfunction, in which overreporting spans psychiatric, somatic, and cognitive complaints, and is considered a separate type of information from underperformance (Sherman et al., 2020). To this end, examinees who fail PVTs and the MCI are different from those who pass PVTs and fail the MCI. This difference could have implications for diagnosis and clinical management. Such a multi-modal assessment with well-validated PVTs and SVTs can also increase the user's confidence that non-credible findings are in fact non-credible. Finally, for examinees who generate failing scores on multiple PVTs, the problem of common method variance could be implicated. By employing an SVT like the MCI (which, relative to PVTs, provides another method of evaluation), the influence of common method variance can be better managed.

Weaknesses

The most pressing limitation of the MCI is that very few published studies have been produced on the measure. Despite high quality of initial studies presented in the manual across several samples, the MCI would benefit from additional research with other samples and in different contexts (e.g., veteran, forensic, and general clinical practice settings). Another arguable weakness is that the measure has lacked other gold-standard memory- or cognitive-based SVTs to be compared to as criteria. In lieu of such, the MCI was initially compared to the WMT, which is a performance measure and thus a different construct. In other words, one could argue that the MCI has not been validated against a gold standard, memory SVT. Another potential limitation is the lack of well-defined and replicated cutoff scores. Finally, the scoring program and AI are based on comparing MCI scores to average scores of various groups, which may be somewhat non-intuitive given most SVTs have formal cutoffs beyond which a scores is considered invalid (e.g., *T* scores on MMPI-family scales). In fact, until relatively recently, there was no published manual for the measure, which may have reduced the popularity of the instrument.

Future Perspectives

As noted previously, identifying a suitable criterion for the MCI is problematic. One approach to overcoming this issue is to use simulation designs, despite this method being a somewhat less-stringent approach than known group designs. No simulation studies were found in which the MCI

was used, and a well-designed, simulation study would add to the existing MCI literature. Next (and as noted above) there have historically been few gold standard cognitively based SVTs by which to compare the MCI. However, recent advances, such as the Inventory of Problems – Memory (IOP-M; Giromini et al., 2020), could provide such a basis of comparison. Third, Slick criteria were recently updated (Sherman et al., 2020) and now formally incorporate exaggerated cognitive symptoms on SVTs into the criteria. Known group studies, particularly in forensic samples, evaluating the MCI compared to the criterion group as defined by the updated Slick criteria would be beneficial to further evaluating the MCI. Fourth, although the standard MCI includes 58 items, there is the option of a short form. The short form does not reduce number of items per scale, rather it only includes four of the nine scales, or 22 items (Verbal Memory Problems, Pain Interferes with Memory, Memory Interferes with Work, and Impairment of Remote Memory). Only one of these scales is categorized as implausible, and it is unclear why these four scales were selected. The short from is not covered in the manual, and the psychometric properties are not at all defined, most notably for the mean MCI short form score. Research informing the utility of the short form compared to the standard form would be potentially useful to MCI users, and the ability to flexibly choose which scales to include in customized short forms would add to the uniqueness and utility of the measure (given the scales are separately validated). Finally, Huber et al (2020) produced memory complaint profiles across cognitively impaired, depressed, and non-credible groups. Further work to establish and replicate such profiles amongst known groups would be a useful addition to the MCI literature.

Declarations

Disclaimer The views, opinions, and/or findings contained in this article are those of the authors and should not be construed as an official Department of the Army, Defense Health Agency, Department of Veterans Affairs, or US Government position, policy, or decision unless so designated by other official documentation.

References

Allen, L. M., Conder, R. L., Green, P., & Cox, D. R. (1997). *CARB 97 manual for the Computerized Assessment of Response Bias*. Durham, NC: CogniSyst.

Armistead-Jehle, P., Gervais, R. O., & Green, P. (2012a). Memory Complaints Inventory results as a function of symptom validity test performance. *Archives of Clinical Neuropsychology, 27*(1), 101–113.

Armistead-Jehle, P., Gervais, R. O., & Green, P. (2012b). Memory complaints and symptom validity test performance in a clinical sample. *Archives of Clinical Neuropsychology, 27*(7), 725–734.

Armistead-Jehle, P., Grills, C. E., Bieu, R., & Kulas, J. (2016). Clinical utility of the Memory Complaints Inventory to detect invalid test performance. *The Clinical Neuropsychologist, 30*(4), 610–628.

Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF (Minnesota multiphasic personality inventory-2-restructured form) manual for administration, scoring, and interpretation*. Minneapolis: The University of Minnesota Press.

Curtis, K. L., Greve, K. W., Bianchini, K. J., & Brennan, A. (2006). California Verbal Learning Test indicators of malingered neurocognitive dysfunction: Sensitivity and specificity in traumatic brain injury. *Assessment, 13*, 46–61.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test (CVLT) manual*. The Psychological Corporation.

Denby, E., Murphy, D., Busuttil, W., Sakel, M., & Wikinson, D. (2019). Neuropsychiatric outcomes in UK military veterans with mild traumatic brain injury and vestibular dysfunction. *Journal of Head Injury Rehabilitation, 35*(1), 57–65.

Disner, S. G., Mattson, E. K., Nelson, W. N., & Armistead-Jehle, P. (in press). The role of posttraumatic stress symptoms on memory complaints and performance in active-duty service members. *The Clinical Neuropsychologist*.

Elias, D. A., MacLaren, V. V., Brien, E. K., & Metcalfe, A. W. S. (2019). Exaggerated functional impairment due to malingered neurocognitive dysfunction following mild traumatic brain injury. *Archives of Clinical Neuropsychology*, *34*(5), 648–656.

Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2007). Development and validation of a Response Bias Scale (RBS) for the MMPI-2. *Assessment, 14*(2), 196–208. https://doi.org/10.1177/1073191106295861

Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green. (2008). Differential sensitivity of the Response Bias Scale (RBS) and MMPI-2 validity scales to memory complaints. *The Clinical Neuropsychologist, 22*(6), 1061–1079.

Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Sellbom, M. (2010). Incremental validity of the MMPI-2-RF over-reporting scales and the RBS in assessing the veracity of memory complaints. *Archives of Clinical Neuropsychology, 25*(4), 274–284.

Giromini, L., Viglione, D. J., Zennaro, A., Maffei, A., & Erdodi, L. A. (2020). SVT meets PVT: Development and initial validation of the Inventory of Problems – Memory (IOP-M). *Psychological Injury and Law, 13*, 261–274.

Green, P. (2003). *Green's Word Memory Test for Windows: User's manual*. Green's Publishing.

Green, P. (2008a). *The Advanced Interpretation program*. Green's Publishing.

Green, P. (2004). *Green's Medical Symptom Validity Test (MSVT) for Microsoft Windows: User's manual*. Green's Publishing.

Green, P. (2008b). *Manual for the nonverbal Medical Symptom Validity Test*. Green's Publishing.

Green, P. (2019). *Users' manual for the Memory Complaints Inventory (MCI)*. Green's Publishing.

Green, P., Gervais, R., & Merten, Th. (2005). Das Memory Complaints Inventory (MCI): Gadachtnisstorungen. *Beschwerdenschilderung Und Leistungsmotivation. Neurol Rehabil, 11*(3), 139–144.

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

Huber, B. N., Jones, R. G., Capps, S. C., & Buchanan, E. M. (2020). Memory complaints inventory profiles: Differentiating neurocognitive impairment, depression, and non-credible performance. *Applied Neuropsychology: Adult*. Advanced online publication. https://doi.org/10.1080/23279095.2020.1735388

Morey, L. C. (1991). *Personality Assessment Inventory*. Psychological Assessment Resources Inc.

Ord, A. S., Shura, R. D., Sansone, A. R., Martindale, S. L., Taber, K. H., & Rowland, J. A. (2021). Performance and symptom validity tests: Are they measuring different constructs? *Neuropsychology*. https://doi.org/10.1037/neu0000722

Randolph, C. (1998). *Repeatable Battery for the Assessment of Neuropsychological Status manual*. The Psychological Corporation.

Rohling, M. L., Green, P., Allen, L. M., III., & Iverson, G. L. (2002). Depressive symptoms and neurocognitive test scores in patients passing symptom validity tests. *Archives of Clinical Neuropsychology, 17*(3), 205–222.

Sabelli, A. G., Messa, I., Giromini, L., Lichtenstein, J. D., May, N., & Erdodi, L. A. (2021). Symptom versus performance validity in patients with mild TBI: Independent sources of non-credible responding. *Psychological Injury and Law*. https://doi.org/10.1007/x12207-021-09400-6

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

Silverberg, N. D., Wertheimer, J. C., & Fichtenberg, N. L. (2007). An effort index for the tepeatable dattery for the Assessment of Neuropsychological Status (RBANS). *The Clinical Neuropsychologist, 21*(5), 841–854.

Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13*(4), 545–561.

Smith, G. E., Petersen, R. C., Ivnik, R. J., Malec, J. F., & Tangalos, E. G. (1996). Subjective memory complaints, psychological distress, and longitudinal change in objective memory performance. *Psychology and Aging, 11*(2), 272–279.

Tombaugh, T. N. (1996). *The test of memory malingering*. Los Angeles, CA: Western Psychology Corporation.

Widows, M. R. & Smith, G. P. (2005). *Structured Inventory of Malingered Symptomatology: Professional manual*. PAR, Inc.