Psychometric Characteristics of The Memory for Intentions Screening Test

Steven Paul Woods[1,*], **Lisa M. Moran**[1], **Matthew S. Dawson**[1], **Catherine L. Carey**[1], **Igor Grant**[1], and **The HIV Neurobehavioral Research Center (HNRC) Group**

[1] Department of Psychiatry, University of California, San Diego

Abstract

The construct of prospective memory (ProM), or "remembering to remember," is hypothesized to play a critical role in normal activities of daily living and has increasingly been the focus of clinical research over the past 10 years. However, the assessment of ProM as part of routine clinical care is presently hampered by the paucity of psychometrically sound, validated ProM tests available in the neuropsychological literature. The Memory for Intentions Screening Test (MIST; Raskin, 2004) is a user-friendly, comprehensive measure of ProM that demonstrates preliminary evidence of construct validity. Extending this research, this study evaluated the psychometric characteristics of the MIST in a sample of 67 healthy adults. Despite a mildly restricted range of scores, results revealed excellent inter-rater reliability, adequate split-half reliability, and satisfactory inter-relationships between the MIST summary score, subscales, and error types. Analysis of demographic correlates showed that the MIST was independently associated with both age and education, but not with sex or ethnicity. These findings broadly support the psychometric properties of the MIST, specifically its reliability and expected relationships with demographic characteristics. Recommendations are provided regarding future research to enhance the clinical usefulness of the MIST.

Keywords

reliability; test construction; neuropsychological assessment; episodic memory

The construct of prospective memory (ProM) has garnered considerable interest within the clinical neuropsychology literature over the past 10 years (McDaniel & Einstein, 2007). ProM is an aspect of declarative (i.e., episodic) memory that describes the formation, maintenance, and execution of future intentions, and is highly dependent upon prefrontal (e.g., Brodmann's area 10; Simons et al., 2006), as well as medial temporal (e.g., hippocampal) neural systems (e.g., Martin et al., 2007a). Colloquially, ProM is referred to as "remembering to remember" and is commonly illustrated by such daily activities as remembering to take a medication at the appropriate time, remembering to return a telephone call, or remembering to pay the monthly household bills. It is therefore understandable why clinical researchers so often herald the ecological validity of ProM. In fact, it has been posited that ProM plays a central role in the independent performance of instrumental activities of daily living (IADLs; e.g., medication adherence), perhaps even more so than other higher-level cognitive functions (e.g., retrospective episodic memory). For example, misremembering the name of a prescribed medication after being queried by

*Corresponding author. Department of Psychiatry (0847), HIV Neurobehavioral Research Center, University of California, San Diego, 150 West Washington Street, 2nd floor, San Diego, California, 92103, USA. Telephone: (619) 543-5004. Fax: (619) 543-1235. spwoods@ucsd.edu.

your physician (i.e., a failure of retrospective memory) is arguably less critical to optimal health than is not fulfilling the intention to take that medication as prescribed (i.e., a failure of ProM).

Accordingly, formal assessment of ProM abilities is arguably an integral aspect of a comprehensive clinical or research neuropsychological evaluation. However, tests of ProM have yet to find their way into the armamentarium of clinical neuropsychologists; to wit, no measures of ProM ranked among the most commonly used assessment instruments in a recent survey of neuropsychological practitioners (Rabin, Barr, & Burton, 2005). Of the top 40 memory assessments listed, only one test, which was endorsed by a mere 6.4% of respondents, includes even a cursory measurement of ProM (i.e., Rivermead Behavioral Memory Test; Wilson, Cockburn, & Baddeley, 1991). The striking absence of ProM tests used in clinical practice may reflect the paucity of published, user-friendly measures of ProM, as well as the restricted clinical usefulness of many of the available techniques (e.g., insufficiently standardized experimental procedures with unknown psychometric properties, no demographically-adjusted normative standards, and limited evidence of construct validity in clinical populations). The development of a standardized, psychometrically-sound measure of ProM that allows for a comprehensive analysis of component processes (e.g., detailed error coding) would be of considerable value to both clinicians and researchers working with a variety of psychiatric, medical, and neurological populations.

The Memory for Intentions Screening Test (MIST; Raskin, 2004) was expressly developed for this purpose and meets the basic parameters of a ProM task (McDaniel & Einstein, 2007); namely, the MIST: 1) requires the delayed execution of intended actions; 2) requires that the intended action be performed in the context of an ongoing foreground (i.e., distractor) task; and 3) provides a constrained window of time in which the intention may be initiated and executed. Specifically, the MIST is a standardized measure in which participants are asked to perform eight different ProM tasks over a 30-min period, during which time they are engaged in a word-search puzzle that serves as the foreground (i.e., distractor) task. The eight ProM trials are balanced in terms of delay interval (i.e., either a 2-min or 15-min delay), cue (i.e., either a time-based or event-based cue), and response modality (i.e., either a verbal or a physical response). Incorrect responses are coded using a detailed, comprehensive scoring system that operationalizes common errors of omission (e.g., loss of time) and commission (e.g., task substitution errors). The MIST also includes an 8-item multiple-choice recognition post-test from which a retrieval index may be generated (Carey et al., 2006). Finally, a more naturalistic 24-hour delay trial may be administered in which the examinee is asked to telephone the clinician the next day and report how many hours they slept the night after the evaluation.

An emergent literature provides preliminary support for the construct validity of the MIST. For example, the MIST discriminates healthy adults from persons with HIV infection (Carey et al., 2006) and individuals with schizophrenia (Woods et al., 2007b). In these clinical samples, MIST impairment correlates with disease severity, including elevated biomarkers of neuropathogenesis in HIV (e.g., monocyte chemoattractant protein-1; Woods et al., 2006) and greater negative psychotic symptoms in schizophrenia (Woods et al., 2007b). The convergent validity of the MIST is supported by its correlations with measures of executive functions, verbal working memory, and retrospective memory in persons with HIV infection and healthy adults (Carey, et al. 2006; Woods et al., 2007a). The MIST has also demonstrated utility in monitoring the efficacy of cognitive rehabilitation efforts in persons suffering from traumatic brain injuries (Fleming, Shum, Strong, & Lightbody, 2005). Perhaps most importantly, the MIST demonstrates evidence of incremental ecological validity relative to traditional tests of retrospective memory (e.g., list learning and recall) in the prediction of IADL declines in persons with HIV infection (Woods et al., in press).

Although the MIST shows promise as a measure of ProM, its ultimate construct validity and clinical usefulness necessitate demonstration of its psychometric characteristics, including reliability, inter-relationships among subscales, and associations with relevant demographics (e.g., age). However, no published studies have heretofore evaluated the basic psychometric properties of the MIST. Accordingly the aims of this investigation were threefold: 1) to determine the inter-rater reliability of the MIST; 2) to evaluate the split-half reliability and internal consistency of the MIST; and 3) to investigate the relationships between the MIST and demographic factors.

Method

Participants

Study participants included 67, English-speaking, healthy adults who were enrolled in a clinical research protocol at the San Diego HIV Neurobehavioral Research Center. Potential participants with prior histories of major psychiatric (e.g., mental retardation, psychosis, and recent substance dependence) or neurological (e.g., seizure disorders, closed head injuries with LOC > 30 minutes, and cerebrovascular accidents) conditions were excluded. Table 1 displays the sample's demographic characteristics.

Materials and Methods

All study volunteers provided informed, written consent and were subsequently administered the MIST as part of a larger neurobehavioral, neurological, medical, and psychiatric evaluation. As described above and displayed in Table 2, the MIST is a standardized test comprised of eight ProM tasks that are balanced on the following characteristics: 1) a 2-min or 15-min delay; 2) a verbal (e.g., "In 2 minutes, ask me what time this session ends") or physical (e.g., "In 15 minutes, use that paper to write down the number of medications you are currently taking") response; and 3) a time-based (e.g., "In 15 minutes, tell me that it is time to take a break") or event-based (e.g., "When I show you a postcard, self-address it") cue. The cognitive load of each item (i.e., the total number of other intentions "online" at the time each intention is supposed to be recalled) is also displayed in Table 2. A series of word search puzzles are provided as a foreground (i.e., distractor) task to prevent overt rehearsal of the prescribed intentions. Each ProM trial on the MIST is worth two possible points: one point is awarded for a correct response and one point for responding (in some manner) at the appropriate time (± 15% of the target) or to the appropriate cue. For example, if a participant is 3 min tardy in asking what time the session ends, only one point is awarded for that trial. Similarly, one point is earned if, for example, the participant signs their name instead of self-addressing the displayed postcard (NB., this differs from the Raskin, 2004 instructions, which award zero points for an incorrect event-based trial).

Individual ProM trials contribute to three of the MIST's six subscales (range = 0–8), as determined by each trial's specific delay, cue, and response characteristics. Each subscale therefore contains four individual ProM trials (see Table 2). The six subscales are then summed to create a summary score, which ranges from 0 to 48. Additionally, participants complete a 3-choice recognition test immediately following the completion of the MIST (range = 0–8). Note that, our recognition protocol also differs slightly from Raskin (2004) in that participants were administered all eight recognition trials, regardless of their performance during the free recall phase of the test. A Retrieval Index (Carey et al., 2006) was also created by subtracting free recall accuracy from correct recognition for each item and summing the difference scores. Finally, a 24-hour probe was administered in which participants were instructed to leave a telephone message for the examiner the following day specifying the number of hours slept the night after the assessment (range = 0–2). Unlike the

other MIST items, participants are allowed to use any mnemonic strategy they wish for the 24-hour probe (e.g., a note in their electronic organizer or assistance from a significant other), but are not explicitly instructed to do so.

The MIST also provides guidelines for standardized qualitative error coding, which is a unique feature of this test. If a participant fails to make a response to a ProM cue (i.e., an error of omission), it is coded as a no response error (NR). Task Substitution (TS) errors can be coded in several situations, including when: 1) the participant substitutes an action for a verbal response (or *vice versa*); 2) a prior task from an earlier part of the test is enacted (i.e., a repetition); or 3) a novel response is performed (i.e., an intrusion). If a participant recognizes a ProM cue, but indicates that they have forgotten all or part of the task, a Loss of Content (LC) error is coded. Loss of Time (LT) errors are recorded when a correct task is performed at the incorrect time (i.e., ± 15% of the target execution time). Place Losing Omissions (PLO) occur when the participant performs only a part of the task or gets distracted prior to task completion. Finally, Random (R) errors are coded for those responses that cannot otherwise be classified.

Data Analyses

A series of Shapiro-Wilk W tests revealed that the MIST data were nonnormally distributed (*ps* < .05). Therefore we adopted a nonparametric approach to data analysis, including Spearman's rho coefficients for correlational analyses and Wilcoxon Ranked Sums tests for between-group comparisons (e.g., sex). No participant committed either PLO or R errors and therefore these data were not useable for the subsequent analyses. Intraclass correlation coefficients (ICCs), Wilcoxon Ranked Sums tests, and percent agreement statistics were all used to analyze the interrater reliability of the MIST. For this aspect of this study, a research assistant transcribed each participant's eight MIST responses verbatim onto two identical blank test protocols. Two psychometrists trained on the administration and scoring of the MIST (LMM and MSD) then independently scored the blinded protocols in a retrospective fashion using the guidelines detailed above. Given the descriptive, exploratory nature of this psychometric study, the critical alpha level was set at .01 and our data interpretations emphasize effect sizes (e.g., Cohen's *d* values) whenever possible.

Results

Descriptive data for the MIST summary score, subscales, and error types are presented in Table 3.

Table 4 displays the inter-rater reliability results, which revealed generally excellent agreement in coding the MIST summary score, subscales, and error types. No MIST variable was significantly different between raters (all *ps* > .10), who achieved 88.1% agreement for the summary score and total errors alike. The median rate of agreement was 0.99 (interquartile range = 0.99, 0.99) across individual items and 0.97 (interquartile range = 0.87, 0.99) for the error types. The ICCs for summary score and total errors were .99 and .97, respectively (*ps* < .0001). Across individual items, the median ICC was .97 (interquartile range = 0.92, 0.99), whereas the median inter-rater reliability coefficient for error types was .98 (interquartile range = 0.82, 1.0).

The split-half reliability of the MIST was .70 as measured by the Spearman-Brown coefficient. Internal consistency analyses revealed that the inter-item reliability of the eight individual MIST trials was generally poor (Cronbach's α = .475); however, the reliability of the six MIST subscales was considerably better (Cronbach's α = .886). The inter-correlations between subscales are displayed in Table 5, which shows that these indices were strongly correlated with the summary score (median *ρ* = .83), with the exception of the

2-min subscale ($\rho = .53$). Total errors and the retrieval index were both highly correlated with the summary score ($ps = -0.97$ and -0.92, respectively), but the individual error types, recognition trial, and distractor total were more modestly associated with the summary score. In contrast, the 24-hour probe did not correlate with any other MIST variable (median $\rho = 0.03$).

Regarding its relationship with demographic factors, Wilcoxon Rank Sums tests showed that no MIST variable was associated with sex or ethnicity ($ps > .10$). However, as displayed in Table 6, several MIST variables (e.g., summary score) were significantly correlated with age ($ps < .01$). Small, trend-level correlations were also observed between a few MIST variables (e.g., summary score and time-based scale) and education ($ps < .05$). A post-hoc linear regression demonstrated that age and education were significant, independent predictors of the MIST summary score ($R^2 = .31$, $p < .0001$).

Discussion

Although literature on ProM in clinical populations has greatly increased in recent years, there remains a gap between research and clinic practice, perhaps in part due to the paucity of clinic-ready ProM tests with evidence supporting their psychometric properties and construct validity. The MIST is a standardized measure of ProM that demonstrates preliminary evidence of construct validity (Carey et al., 2006; Fleming et al., 2005; Woods et al., 2006; 2007a,b; in press), but whose psychometric characteristics had not previously been published. Results of this study generally support the psychometric properties of the MIST in a sample of 67 healthy adults. In particular, the MIST demonstrated excellent inter-rater reliability. Although high levels of inter-rater reliability might be expected of the summary and subscale scores, the excellent reliability coefficients and high rates of overall agreement in coding the MIST error types are quite encouraging. This is particularly important because, as noted above, a unique feature of the MIST is its operationalization of various ProM errors, thus allowing the user to take a component process approach to elucidate the underlying mechanisms of ProM failures. Indeed, the analysis of error types has previously shown to be informative in clinical research on ProM impairment in persons with HIV infection (e.g., Carey et al., 2006; Woods et al., in press) and schizophrenia (Woods et al., 2007b).

Despite the relative brevity of the MIST (i.e., only 8 trials), the split-half reliability coefficient was satisfactory (i.e., .70), thereby supporting the reliability of this novel test. Furthermore, although internal consistency was fairly poor for the individual MIST trials, the subscales demonstrated a much stronger, acceptable level of inter-scale reliability. Such findings are consistent with general psychometric principles, which highlight the importance of test length and variability in scores as important determinants of a test's reliability (e.g., Sattler, 1992). The low reliability of the individual trials on the MIST may be a function of a restricted range of scores for each trial (i.e., 0–2 versus 0–8 for the subscales), as well as possible ceiling effects in this sample of healthy adults. In fact, the upper and lower ends of the interquartile range were at ceiling for all of the individual trials except trial 8 (see Table 4). Accordingly, MIST users should be cautious in interpreting analyses of item-level responses, instead favoring examination of the summary score and subscales whenever possible.

Inter-correlations between the six MIST subscales were generally in the large range, with only the 2-min subscale demonstrating medium associations with the other subscales, probably as a function of ceiling effects (see Table 3). No response, task substitution, and loss of content errors displayed medium correlations with the summary score. Loss of time errors were not significantly related to any other MIST variable, although a few very small,

trend-level correlations were observed (e.g., with total errors). It is unclear exactly why loss of time errors were so weakly associated with the other MIST indices, but one possible explanation is the relative infrequency of these error types (see Table 3), which were significantly less common in this group of healthy adults than were other errors (all $ps < .05$).

Relative to other tests of ProM, another distinctive feature of the MIST is its inclusion of a standardized distractor test and post-test recognition trials, which afford the user a clearer understanding of the cognitive mechanisms of ProM failures (e.g., differentiating consolidation versus retrieval deficits). In the current study, these variables showed small-to-medium correlations with the summary score, as well as with the MIST subscales and error types. The retrieval index, which reflects the number of intentions that were incorrectly recalled, but accurately recognized, was highly correlated with the MIST summary score and subscales. Although the retrieval index was arguably collinear with the summary score in these healthy adults, smaller correlations may be evident in clinical samples with retrieval deficits (e.g., Woods et al., 2007b). Among the different error types, the retrieval index was most strongly correlated with loss of content and task substitutions errors. Such results are commensurate with the relatively greater retrospective memory component inherent in these error types; for example, loss of content errors occur when an individual recognizes the ProM cue, acknowledges that it is time to execute the intention, but cannot remember the details of the intended action. Encoding and retrieval errors such as this may be particularly amenable to correction in a recognition trial.

The 24-hour probe did not correlate with any other MIST variable or demographic characteristic. Small correlations between laboratory and semi-naturalistic tasks is a common occurrence in ProM research (e.g., Rendell & Thompson, 1999), as well as in the broader neuropsychological literature on ecological validity, which struggles to strike the appropriate balance between experimental control and psychometric rigor on the one hand, and accurately reflecting the complexities of daily life on the other (see Chaytor & Schmitter-Edgecombe, 2003 for a review). Case in point, the 24-hour trial of the MIST differs procedurally from the rest of the test items by allowing participants to utilize any mnemonic strategies they deemed appropriate (perhaps more accurately reflecting real-world, everyday circumstances). Unfortunately, no data were gathered regarding whether or not participants used mnemonic tactics (and if so, of what type), which would have been interesting to examine in relation to task success. Prior studies that have used semi-naturalistic ProM tasks such as this typically find that external cues (e.g., electronic reminders) are reliably associated with task completion (e.g., Rendell & Thompson, 1999). Another psychometric issue is the relatively low task completion rate, which is common in semi-naturalistic ProM studies (see McDaniel & Einstein, 2007 for a review). Only 45% of participants in this study actually telephoned the examiner, 30% of whom failed to provide the requested information or called at the wrong time. Importantly, however, the relative absence of evidence in support of the psychometric properties of the 24-hour probe should necessarily not be interpreted as evidence of absence, but rather as a cautionary note, particularly since this index demonstrates some evidence of criterion-related validity (Carey et al., 2006).

A final series of analyses were undertaken to evaluate the relationship of the MIST to demographic factors. Neither sex nor ethnicity were significantly associated with the MIST. Trend-level correlations showed that MIST performance correlated positively with years of education, which is consistent with other higher-order cognitive constructs, such as executive functions (e.g., abstraction; Heaton, Miller, Taylor, & Grant, 2004), which generally increase in association with greater levels of educational attainment. Commensurate with an extensive literature on normal aging and ProM (e.g., Henry,

MacLeod, Phillips, & Crawford, 2004), younger age was also associated with better MIST performance. In fact, research on ProM first gained prominence in the aging literature and has been investigated most extensively with respect to its decline in older adults (Henry et al., 2004). Age-related changes of the brain are predominantly characterized by neuroanatomical alterations in the prefrontal systems (Drachman, 2006; Mielke et al., 1998), which are known to support normal ProM functioning (e.g. Simons et al., 2006). It has been hypothesized that ProM is particularly vulnerable to the effects of aging because of its reliance on internal control mechanisms (e.g., self-initiated retrieval) that depend heavily on prefrontal systems (Craik, 1986). To this end, the relationship between ProM deficits and aging appears to be moderated by several factors, particularly the level of strategic (cf. automatic) processing imposed by the task. Specifically, age effects are typically larger for time- and event-based tasks that place more demands on self-initiated processes (e.g., internally monitoring the passage of time rather than recognizing external cues in the environment; Henry et al., 2004; McDaniel & Einstein, 2007). Interestingly, age was slightly more related to event- versus time-based subscales in this study, perhaps reflecting the relatively stronger processing demands of this event-based task relative to other measures of ProM (e.g., Martin et al., 2007b).

The primary limitation of this study is that the sample was comprised of highly educated, healthy adults, raising questions about the generalizability of these data to individuals with lower levels of education. Relatedly, and as noted throughout the discussion, this sample generated a restricted range of scores, which may have negatively impacted some of the statistical analyses (e.g., Type II error for the 24-hour probe analyses). Indeed, the integrity of reliability analyses depends upon the heterogeneity and general level of performance within the study sample (Anastasi, 1988). As such, findings reported herein may differ in impaired clinical samples (e.g., Alzheimer's disease), which are more likely to generate a broader distribution of scores and error types (e.g., Delis, Jacobson, Bondi, Hamilton, & Salmon, 2003). This issue is particularly relevant to the construction of a ProM task, which must include a sufficient number of trials to enhance reliability, whilst also ensuring that the interval between encoding and the response cue is long enough for the participant to engage in the distractor test (and thereby differentiating ProM from working memory), but not so long as to unnecessarily extend the time of the neuropsychological battery or encroach on other cognitive tests (i.e., out of concern that the increased cognitive load of ProM may interfere with performance on the other tests).

In summary, ProM is important aspect of daily functioning and its clinical assessment represents an important future direction for the field of clinical neuropsychology. The MIST is a user-friendly, comprehensive measure of ProM that demonstrates burgeoning construct validity. The current study provides the first published evidence for the psychometric integrity of the MIST, including its inter-rater reliability, internal consistency, and association with demographic factors (i.e., age and education). The clinical usefulness of the MIST is nevertheless still hampered by the lack of demographically-adjusted normative standards, especially considering its demonstrated association with age and education. Until such data are published using an adequate sample size, studies using the MIST are well advised to consider possible age and education confounds in interpreting their findings. Data on test-retest and alternate form reliability (e.g., reliable change indices) will also be invaluable in enhancing the MIST's usefulness in clinic, as well as in longitudinal research. Finally, it is important to mention that the evaluation of a test's construct validity is an ongoing process and clinical research on the MIST is still nascent. It is our hope, however, that these early studies will serve as a catalyst for future investigations into the construct validity of the MIST in a variety of clinical populations in which episodic memory deficits are prevalent and interfere with everyday functioning.

Acknowledgments

References

Anastasi, A. Psychological testing. 6. New York: Macmillan Publishing Co., Inc; 1988.

Carey CL, Woods SP, Rippeth JD, Heaton RK, Grant I. The HNRC Group. Prospective memory in HIV-1 infection. Journal of Clinical and Experimental Neuropsychology 2006;28:536–548. [PubMed: 16676475]

Chaytor N, Schmitter-Edgecombe M. The Ecological Validity of Neuropsychological Tests: A Review of the Literature on Everyday Cognitive Skills. Neuropsychology Review 2003;13:181–197. [PubMed: 15000225]

Craik, FIM. A functional account of age differences in memory. In: Klix, F.; Hagendorf, H., editors. Human Memory and Cognitive Capabilities: Mechanisms and Performances. New York, NY: Elsevier Science; 1986. p. 409-422.

Delis DC, Jacobson M, Bondi MW, Hamilton JM, Salmon DP. The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: Lessons from memory assessment. Journal of the International Neuropsychological Society 2003;9:936–946. [PubMed: 14632252]

Drachman DA. Aging of the brain entropy and Alzheimer disease. Neurology 2006;67:1340–1352. [PubMed: 17060558]

Fleming JM, Shum D, Strong J, Lightbody S. Prospective memory rehabilitation for adults with traumatic brain injury: A compensatory training programme. Brain Injury 2005;19:1–13. [PubMed: 15762096]

Heaton, RK.; Miller, SW.; Taylor, MJ.; Grant, I. Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African-American and Caucasian adults Scoring Program. Lutz, FL: Psychological Assessment Resources, Inc; 2004.

Henry JD, MacLeod MS, Phillips LH, Crawford JR. A meta-analytic review of prospective memory and aging. Psychology and Aging 2004;19:27–39. [PubMed: 15065929]

Martin T, McDaniel MA, Guynn MJ, Houck JM, Woodruff CC, Bish JP, et al. Brain regions and their dynamics in prospective memory retrieval: a MEG study. International Journal of Psychophysiology 2007a;64:247–258. [PubMed: 17126436]

Martin EM, Nixon H, Pitrak DL, Weddington W, Rains NA, Nunnally G, et al. Characteristics of prospective memory deficits in HIV-seropositive substance-dependent individuals: Preliminary observations. The Journal of Clinical and Experimental Neuropsychology 2007b;29:496–504.

McDaniel, MA.; Einstein, GO. Prospective memory: An overview and synthesis of an emerging field. Los Angeles: Sage Publications; 2007.

Mielke R, Kessler J, Szelies B, Herholz K, Wienhard K, Heiss WD. Normal and pathological aging-findings of positron-emission-tomography. Journal of Neural Transmission 1998;105:821–837. [PubMed: 9869321]

Rabin LA, Barr WB, Burton LA. Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN and APA Division 40 members. Archives of Clinical neuropsychology 2005;20:33–65. [PubMed: 15620813]

Raskin S. Memory for intentions screening test [abstract]. Journal of the International Neuropsychological Society 2004;10(Suppl 1):110. [PubMed: 14751013]

Rendell P, Thomson DM. Aging and prospective memory: Differences between naturalistic and laboratory tasks. Journals of Gerontology. Series B, Psychological Sciences and Social Sciences 1999;54:256–269.

Sattler, JM. Assessment of children. 3. San Diego, CA: Author; 1992.

Simons JS, Scholvinck ML, Gilbert SJ, Frith C, Burgess P. Differential components of prospective memory? Evidence from fMRI. Neuropsychologia 2006;44:1388–1397. [PubMed: 16513147]

Wilson, BA.; Cockburn, J.; Baddeley, AD. The Rivermead Behavioral Memory Test Manual. 2. Suffolk, UK: Thames Valley Test Company; 1991.

Woods SP, Carey CL, Moran LM, Dawson MS, Letendre SL, Grant I. The HNRC Group. Frequency and predictors of self-reported prospective memory complaints in individuals infected with HIV. Archives of Clinical Neuropsychology 2007a;22:187–195. [PubMed: 17289343]

Woods SP, Iudicello JE, Moran LM, Carey CL, Dawson MS, Grant I. The HNRC Group. HIV-associated prospective memory impairment increases risk of dependence in everyday functioning. Neuropsychology. (in press).

Woods SP, Morgan EE, Marquie-Beck J, Carey CL, Grant I, Letendre SL. The HNRC Group. Markers of macrophage activation and axonal injury are associated with prospective memory in HIV-1 disease. Cognitive and Behavioral Neurology 2006;19:217–221. [PubMed: 17159619]

Woods SP, Twamley EW, Dawson MS, Narvaez JM, Jeste DV. Deficits in cue detection and intention retrieval underlie prospective memory impairment in schizophrenia. Schizophrenia Research 2007b;90:344–350. [PubMed: 17175138]

Table 1

Demographic characteristics of the 67 study volunteers.

Variable	Mean	SD	Range
Age (years)	41.2	12.5	19, 74
Education (years)	14.7	2.3	11, 20
Sex (% men)	55.2	--	--
Handedness (% right)	83.6	--	--
Ethnicity (%)	--	--	--
Caucasian	58.2	--	--
African-American	31.3	--	--
Hispanic	6.0	--	--
Other	4.5	--	--

Note. SD = standard deviation.

Table 2

The Memory for Intentions Screening Test (MIST)

Order of Presentation	Instruction	Cue	Response Modality	Time Delay (min)	Order of Execution	Cognitive Load
1	"In 15 minutes, tell me it is time to take a break." - Recognition Foils: *"At any point during this test, were you supposed to: Tell the examiner to turn off the lights? Tell the examiner to leave the room?"*	Time	Verbal	15	4	4
2	"When I show you a red pen, sign your name on your paper." - Recognition Foils: *"When the examiner gave you a red pen, were you supposed to: Write your date of birth? Take it home with you?"*	Event	Action	15	5	3
3	"In 2 minutes, ask me what time this session ends today." - Recognition Foils: *"At any point during this test, were you supposed to: Ask what time the office closes? Ask for your medical records?"*	Time	Verbal	2	1	4
4	"When I show you a postcard, self-address it." - Recognition Foils: *"When you were handed a postcard, were you supposed to: Write today's date? Write a note to the examiner?"*	Event	Action	15	6	2
5	"When I show you a Request for Records form, write your doctors' names on it." - Recognition Foils: *"When the examiner handed you a Request for Records form, were you supposed to: Write your phone number? Fold the form?"*	Event	Action	2	2	4
6	"In 15 minutes, use that paper [examiner points to word search puzzle] to write the number of medications you are currently taking." - Recognition Foils: *"At any point during this test, were you supposed to: Write a list of your past hospitalizations? Write the number of children in your family?"*	Time	Action	15	8	1
7	"When I show you the tape recorder, tell me to rewind the tape." - Recognition Foils: *"When the examiner showed you a tape recorder, were you supposed to: Press the stop button? Tell the examiner to check the battery?"*	Event	Verbal	2	3	5
8	"In 2 minutes, please tell me two things you forgot to do this past week." - Recognition Foils: *"At any point during this test, were you supposed to: "Tell the examiner 2 grocery items? Tell the examiner 2 things you have to do tonight?"*	Time	Verbal	2	7	2

Table 3

Descriptive data for the Memory for Intentions Screening Test (*N* = 67).

MIST variable	Mean	SD	Median	IQR	Range
Summary score	43.1	4.8	42.0	39.0, 48.0	30.0, 48.0
Time-based	6.7	1.2	7.0	6.0, 8.0	4.0, 8.0
Event-based	7.6	0.6	8.0	7.0, 8.0	6.0, 8.0
Verbal	7.5	0.8	8.0	7.0, 8.0	4.0, 8.0
Action	6.9	1.1	7.0	6.0, 8.0	5.0, 8.0
2-min	7.8	0.7	8.0	8.0, 8.0	4.0, 8.0
15-min	6.6	1.3	7.0	6.0, 8.0	3.0, 8.0
Total errors	1.3	1.3	1.0	0.0, 2.0	0.0, 5.0
PM errors	0.3	0.6	0.0	0.0, 1.0	0.0, 2.0
TS errors	0.3	0.6	0.0	0.0, 1.0	0.0, 3.0
LC errors	0.6	1.0	0.0	0.0, 1.0	0.0, 4.0
LT errors	0.1	0.4	0.0	0.0, 0.0	0.0, 2.0
PLO errors	0.0	0.0	0.0	0.0, 0.0	0.0, 0.0
R errors	0.0	0.0	0.0	0.0, 0.0	0.0, 0.0
Recognition	7.8	0.5	8.0	8.0, 8.0	6.0, 8.0
Retrieval index	1.2	1.1	1.0	0.0, 2.0	−1.0, 3.0
Distractor words	20.5	7.7	18.0	16.0, 26.0	10.0, 42.0
24-hour probe	0.7	0.9	0.0	0.0, 2.0	0.0, 2.0

Note. IQR = interquartile range; LC = loss of content; LT = loss of time; MIST = Memory for Intentions Screening Test; NR = no response; PLO = place losing omission; R = random; SD = standard deviation; TS = task substitutions.

Table 4

Inter-rater reliability of the Memory for Intentions Screening Test (*N* = 67).

MIST variable	Rater 1	Rater 2	ICC*	95% CI	Agreement (%)
Summary	43.5 (39.0,48.0)	43.5 (39.0,48.0)	.99	.98, .99	88.1
Trial 1	2.0 (2.0, 2.0)	2.0 (2.0, 2.0)	.98	.97, .99	98.5
Trial 2	2.0 (2.0, 2.0)	2.0 (2.0, 2.0)	.92	.87, .95	98.5
Trial 3	2.0 (2.0, 2.0)	2.0 (2.0, 2.0)	1.00	1.0, 1.0	100.0
Trial 4	2.0 (1.0, 2.0)	2.0 (1.0, 2.0)	.95	.93, .97	97.0
Trial 5	2.0 (2.0, 2.0)	2.0 (2.0, 2.0)	.97	.95, .98	98.5
Trial 6	2.0 (2.0, 2.0)	2.0 (2.0, 2.0)	.89	.82, .93	95.5
Trial 7	2.0 (2.0, 2.0)	2.0 (2.0, 2.0)	---	---	98.5
Trial 8	1.0 (1.0, 2.0)	1.0 (1.0, 2.0)	.99	0.99, 0.99	98.5
Total errors	1.0 (0.0, 2.0)	1.0 (0.0, 2.0)	.97	.95, .98	88.1
PM errors	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	1.0	1.0, 1.0	100.0
LC errors	0.0 (0.0, 1.0)	0.0 (0.0, 1.0)	.98	.96, .99	97.0
LT errors	0.0 (0.0, 1.0)	0.0 (0.0, 1.0)	.98	.96, .99	97.0
TS errors	0.0 (0.0, 0.8)	0.0 (0.0, 1.0)	.76	.64, .85	80.6

Note. Data are presented as medians with the interquartile ranges in parentheses. ICC = intraclass correlation coefficient; CI = confidence interval for the ICCs; LC = loss of content; LT = loss of time; NR = no response; R = random; TS = task substitutions.

*
all *ps* < 0.0001.

Table 5

Correlations between the MIST summary score, subscales, and error types ($N = 67$).

MIST Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Summary score	--															
2. Time-based	0.94 §	--														
3. Event-based	0.75 §	0.51 §	--													
4. Verbal	0.71 §	0.71 §	0.46 §	--												
5. Action	0.91 §	0.84 §	0.73 §	0.37 ‡	--											
6. 2-min	0.53 §	0.38 ‡	0.54 §	0.55 §	0.33 ‡	--										
7. 15-min	0.94 §	0.95 §	0.64 §	0.60 §	0.91 §	0.25 *	--									
8. Total errors	−0.97 §	−0.89 §	−0.77 §	−0.76 §	−0.84 §	−0.54 §	−0.91 §	--								
9. NR errors	−0.58 §	−0.63 §	−0.29 *	−0.18	−0.69 §	−0.08	−0.58 §	0.37 ‡	--							
10. TS errors	−0.54 §	−0.44 §	−0.52 §	−0.50 §	−0.40 §	−0.33 ‡	−0.52 §	0.61 §	0.00	--						
11. LC errors	−0.60 §	−0.51 §	−0.63 §	−0.52 §	−0.48 §	−0.43 §	−0.55 §	0.70 §	−0.01	0.20	--					
12. LT errors	−0.16	−0.22	0.03	−0.29 *	−0.06	0.02	−0.22	0.26 *	−0.20	0.19	0.01	--				
13. Recognition	0.46 §	0.44 §	0.46 §	0.37 ‡	0.38 ‡	0.32 ‡	0.41 §	−0.42 §	−0.33 ‡	−0.26 *	−0.29 *	0.01	--			
14. Retrieval index	−0.92 §	−0.84 §	−0.73 §	−0.74 §	−0.79 §	−0.54 §	−0.86 §	0.97 §	0.29 *	0.59 §	0.71 §	0.27 *	−0.22	--		
15. Distractor total	0.40 §	0.36 ‡	0.34 ‡	0.29 *	0.38 ‡	0.27 *	0.35 ‡	−0.38 ‡	−0.23	−0.31 *	−0.14	−0.18	0.14	−0.38 ‡	--	
16. 24-Hour probe	−0.03	0.02	−0.15	0.05	−0.05	0.02	−0.05	0.02	0.02	0.12	0.01	0.03	−0.11	−0.02	0.08	--

Note. LC = loss of content; LT = loss of time; MIST = Memory for Intentions Screening Test; NR = no response; TS = task substitutions

§ $p < 0.001$.

‡ $p < 0.01$.

* $p < 0.05$.

Table 6

Correlations between the MIST and demographic characteristics ($N = 67$).

MIST Variable	Age	Education
Summary score	−0.49 §	0.27 *
Time-based	−0.40 §	0.26 *
Event-based	−0.51 §	0.17
Verbal	−0.28 *	0.21
Action	−0.50 §	0.26 *
2-min	−0.22	0.06
15-min	−0.47 §	0.27 *
Total errors	0.48 §	−0.23
PM errors	0.27 *	−0.24 *
TS errors	0.23	−0.26 *
LC errors	0.38 ‡	−0.04
LT errors	−0.03	0.02
Recognition	−0.16	0.13
Retrieval index	0.48 §	−0.21
Distractor words	−0.37 ‡	0.13
24-hour probe	−0.09	−0.05

Note. LC = loss of content; LT = loss of time; MIST = Memory for Intentions Screening Test; NR = no response (omissions); TS = task substitutions.

§ p < 0.001.

‡ p < 0.01.

* p < 0.05.