Article

Preliminary Validation of the Global Neuropsychological Assessment in Alzheimer's Disease and Healthy Volunteers

Assessment
2022, Vol. 29(4) 817–825

DOI: 10.1177/1073191121991221
journals.sagepub.com/home/asm
SAGE

Lauren T. Olson[1], Alan Smerbeck[2], Christina M. Figueroa[3], Jeremy M. Raines[4], Kinga Szigeti[1], David J. Schretlen[5], and Ralph H. B. Benedict[1]

Abstract

Methods. We administered the Global Neuropsychological Assessment (GNA), an abbreviated cognitive battery, to 105 adults aged 73.0 ± 7.1 years, including 28 with probable Alzheimer's disease, 9 with amnestic mild cognitive impairment, and 68 healthy controls. We examined group differences in baseline performance, test–retest reliability, and correlations with other conventional tests. *Results.* Healthy adults outperformed patients on all five GNA subtests. Test–retest intraclass correlation coefficients were significant for all GNA subtests. Among patients with healthy controls, GNA Story Memory correlated best with Wechsler Memory Scale–Revised (WMS-R) Logical Memory for learning and delayed recall, GNA Digit Span correlated most highly with the Wechsler Adult Intelligence Scale–Third Edition (WAIS-III) Digit Span, GNA Perceptual Comparison correlated most highly with the Trail Making Test, and GNA Animal Naming correlated most highly with Supermarket Item Naming. *Conclusions.* Preliminary findings suggest that the GNA shows good test–retest validity, clear convergent and discriminant construct validity, and excellent diagnostic criterion validity for dementia and mild cognitive impairment in an American sample.

Keywords

GNA, cognitive assessment, Alzheimer's disease, amnestic MCI, neuropsychology

Multinational studies reveal that in addition to contributions of age, sex, and education, individual differences in language, culture, and race/ethnicity affect cognitive test performance. Consequently, there is growing interest in accounting for these factors in research on cognitive functioning in disease and its treatment. The development and standardization of a brief yet comprehensive cognitive test battery that is translated into many languages and normed in many countries would improve accessibility to clinical neurocognitive evaluations and facilitate clinical trials where the primary outcomes involve cognitive abilities, such as processing speed and episodic memory. Even within the United States, demographic and sociopolitical shifts highlight the rapidly growing need for culturally competent neuropsychological services (Rivera Mindt et al., 2010).

A particularly salient niche of neuropsychological assessment are the so-called story memory tests, the gold standard of which is the Logical Memory subtest from the Wechsler Memory Scale (WMS; Wechsler, 1945). As utilized in the Calibrated Neuropsychological Normative System, there are two WMS–Revised (WMS-R) stories. Story A has 66 words of which 25 are scored, and Story B has 68 words of which 25 are scored items. The total test score is the sum of scored items from both stories. The text passages contain proper nouns (e.g., city names) and culture-specific contents (e.g., CB radio "handles") that hinder their translation, norming in different countries, and suitability for generating alternate, equivalent forms.

Practice effects for the Logical Memory test are well known (Gavett et al., 2016; McCaffrey et al., 1993). Until recently, alternate versions of story memory tests were relegated to single experiments with little normative data or validity research subject to peer review. Trifilio et al. (2020) recently published psychometric data supporting alternate versions of Logical Memory developed by Newcomer (Newcomer, Craft, et al., 1999; Newcomer, Farber, et al., 1999; Newcomer, Selke, et al., 1999). These passages showed good convergent validity with the Logical Memory

[1]University at Buffalo, SUNY, Buffalo, NY, USA
[2]Rochester Institute of Technology, Rochester, NY, USA
[3]Arizona Neurology Associates, Phoenix, AZ, USA
[4]University of North Dakota, Grand Forks, ND, USA
[5]Johns Hopkins School of Medicine, Baltimore, MD, USA

Corresponding Author:
Ralph H. B. Benedict, UBMD Neurology, Conventus Building, 1001 Main Street, 4th Floor, Buffalo, NY 14203, USA.
Email: benedict@buffalo.edu

results and correlated with magnetic resonance imaging (MRI) abnormalities in healthy older-aged adults. However, evidence of their test–retest reliability and diagnostic validity are not yet available. Also, the passages include proper nouns and culture-specific contents (e.g., locations).

Cognitive processing speed (CPS) also is very sensitive to the effects of aging and neurodegenerative disease (Benedict et al., 2017; Lee et al., 2016; Salthouse, 2010; Steffener et al., 2013). Brief tests of attention, working memory, and CPS also can be interposed between learning and delayed recall trials of episodic memory tests. In conventional neuropsychological testing, instruments such as the Trail Making Test (Reitan, 1958), the Digit Span (Wechsler, 1939), and the Symbol Digit Modalities Test (Smith, 1982) are often administered between learning and delayed recall trials of memory tests. We followed this convention for a newly developed story memory test with five alternate forms in the Global Neuropsychological Assessment (GNA).

In this study, we hypothesized that the GNA test battery would (1) show acceptable test–retest reliability; (2) distinguish among groups of older adults with probable Alzheimer's disease (AD), amnestic MCI (aMCI), or no known cognitive disorder; and (3) demonstrate convergent and discriminant validity vis-à-vis standard neuropsychological tests of related and unrelated constructs. Confirming these hypotheses in an English-speaking American sample is critical to moving forward with plans to translate the GNA into multiple languages and norm it in multiple countries. Here, we examine whether preliminary findings support the further pursuit of these goals.

Methods

Subjects

Comprising the sample were 105 healthy volunteers and patients diagnosed with either probable AD ($n = 28$) or aMCI ($n = 9$), all between 60 and 91 years old. Diagnoses were made independently, prior to GNA testing based on a workup consisting of a neurological exam, Mini-Mental State Exam (Folstein et al., 1975), brain MRI scan, laboratory blood studies, and neuropsychological testing in all cases. Diagnoses were made independently, prior to GNA testing, at a consensus case conference that included at least one board-certified neurologist, a board-certified neuropsychologist, a neurology nurse practitioner, and a licensed social worker specializing in dementia.

Healthy controls (HCs) were recruited from the community and paid for participating at the University at Buffalo (UB) clinic ($n = 20$), or volunteered to take the GNA for free while waiting for a family member or friend being seen in the Johns Hopkins Hospital (JHH) Medical Psychology Clinic ($n = 46$). Analysis of demographic data for HCs recruited from the two research sites indicated that site groups did not differ significantly in sex, $\chi^2(2, n = 66) = 0.297$, $p = .400$, or age, $F(1, 65) = 1.37$, $p = .247$, but they did differ in educational achievement, $F(1, 65) = 4.46$, $p = .039$. These findings likely reflect a difference in enrollment procedures used across the two sites and are not unexpected. HCs were excluded if they reported any history of illness that could compromise cognitive function.

Assessments: The Global Neuropsychological Assessment

The GNA consists of seven subtests (Story Memory, Digit Span, Perceptual Comparison, Verbal Fluency [VF], Category Switching, Spatial Span, and the four-item Patient Health Questionnaire or PHQ-4), of which the first four were administered in this study. There are five alternate forms of the GNA for most subtests. A separate investigation of interform reliability is underway and will be reported elsewhere.

The investigators (mainly DJS and RHBB) independently created universally themed, culturally sensitive short stories, resulting in 16 initial compositions. Similar stories were eliminated or merged. By intention, there were no proper nouns, references to locales, or/and no geographic or culturally specific verbiage in these passages. Over several months of piloting and feedback, 12 stories underwent extensive revision to meet specified requirements. Each story contains 28 to 32 words in total, of which 14 comprise the target (to-be-remembered) words for each story. The target words in each story consist of nouns, pronouns, verbs, and adjectives. No adverbs, conjunctions, prepositions, or injunctions serve as target words. Translation software was also used to assess and eliminate story elements that were problematic in multiple languages. Seven studies were rejected based on difficulties encountered during preliminary translation into Japanese, Arabic, Spanish, and Burmese. This left us with five stories to be used for psychometric validation.

The Story Memory GNA subtest includes three trials. The story is read aloud by the examiner at a normal pace. After the first presentation, participants are asked to recall the passage using the exact words that were spoken, avoiding synonyms or gist recollection. The number of target words recalled correctly out of 14 possible is recorded. The story is then repeated with the same instructions for a second learning trial.

After two intervening subtests (Digit Span and Perceptual Comparison) are administered, participants are asked to recall the story presented earlier, without repeated exposure. Recorded are the number correct for each trial, as well as the sum of Trials 1 and 2 for a Total Learning score.

The Digit Span subtest is presented in the conventional manner, beginning with a forward span of two digits and concluding with two consecutive failures up to a ceiling of

nine digits. For backward span, the ceiling score is eight digits.

The GNA VF subtest consists of two parts. The first is Animal Naming, wherein participants are asked to name as many animals as possible in 60 seconds. The second, Category Switching, was not administered in this study.

The Perceptual Comparison subtest is similar to the task developed by Salthouse (2010). Participants compare two sets of shapes, each containing two simple shapes, and verbally state whether the shape sets are the "same" or "different" as quickly as possible. Responses are oral, and the examiner records each response on a response form. There are 54 items presented in three columns on the page. The total number of items attempted and the total number of correct items are scored; the latter is used for analysis.

Assessments: Standard Cognitive Testing

Healthy volunteers and patients from UB completed a conventional neuropsychological battery at an appointment separate from the GNA assessment. The neuropsychological battery was used for clinical dementia workups and employs the North American Adult Reading Test (NAART; Spreen & Strauss, 1998) to estimate the premorbid IQ. Confrontation naming is assessed using the Boston Naming Test (BNT; Kaplan et al., 1983). Letter fluency and category fluency are assessed using the Calibrated Ideational Fluency Assessment (CIFA), using the sum of words provided for the letters "S" and "P" and the sum of words for the categories animals and supermarket items (Schretlen & Vannorsdall, 2010), although Animal Naming was administered as part of the GNA. Visuospatial perception/construction was assessed using the Beery Visuomotor Integration (VMI) measure where subjects copy increasingly difficult figures, the raw score being the correct copies out of 30 (Beery, 2004). Memory for word lists was assessed using the Hopkins Verbal Learning Test–Revised (HVLT-R) that includes learning over three trials, followed by delayed recall and recognition tasks (Benedict et al., 1998). As recommended in the Calibrated Neuropsychological Normative System (CNNS) battery, we used the Wechsler Memory Scale–Revised (WMS-R) Logical Memory and Visual Reproduction subtests, the former also assessing immediate and delayed recall of story passages (Schretlen et al., 2010). Working memory was measured using the Wechsler Adult Intelligence Scale–Third Edition (WAIS-III) Digit Span Test, computing raw scores for highest forward digit recall, highest backward digit recall, and the sum of the longest forward span and longest backward span (Zhu et al., 2001). Processing speed and mental set switching were assessed using the Trail Making Test Parts A and B (Reitan, 1958). Executive function was evaluated with the Delis–Kaplan Executive Function System (DKEFS) Sorting Test, yielding raw scores for the number of correct sorts, total description scores, and the number of repeated sorts (Delis et al., 2001).

Procedure and Analysis

Prior to enrollment, UB clinic patients were administered the standard neuropsychological battery used for purposes of diagnosis and treatment planning. Only patients diagnosed with probable AD or aMCI and underwent standard testing within the preceding 6 months were invited to participate in this study.

The study was reviewed and received institutional review board approval through UB and JHH. Patients and healthy adults at UB gave written informed consent to participate. HCs at JHH gave oral informed consent to participate. Exclusion criteria were as follows: (a) diagnosis of neurocognitive disorder other than probable AD or aMCI; (b) history of major cortical infarction diagnosed by either MRI or clinical stroke with persistent focal neurological deficits; (c) diagnosis of major depressive disorder, bipolar disorder, or any psychiatric disorder judged to impact cognitive function; (d) history electroconvulsive therapy or any somatic therapy that could influence cognitive performance; and (e) diagnosis of other medical illnesses that may impair cognition.

At the time of their enrollment in this study, UB patient participants agreed to take the GNA test battery twice and were randomly assigned to one of two counterbalanced testing orders. For this substudy, the VF task was not readministered. Half of the patients completed the same GNA form at both assessments, and the other half completed alternate forms of the GNA. The test–retest interval was 2 to 4 weeks ($M = 22.4$; $SD = 7.8$ days) for all patients.

Using IBM SPSS 23.0 statistical software, we first determined that all GNA measures showed acceptable skewness and kurtosis as reported in Figure 1. Each variable was also examined for outliers, and two HCs produced one or more GNA subtest scores that fell more than 2 *SD*s from the control group means. These same subjects were nonnative English speakers and were thus excluded from further analysis. To assess the discriminability of the GNA subtest measures, we conducted a one-way analysis of variance, followed by post hoc tests using the Student–Newman–Keuls procedure. Due to the number of statistical tests conducted, we adopted a conservative probability threshold of $p < .01$ to designate statistical significance.

Results

Descriptive demographic data for the probable AD group ($n = 28$), the aMCI group ($n = 9$), and the HC group ($n = 66$) are shown in Table 1. Analyses of demographic data

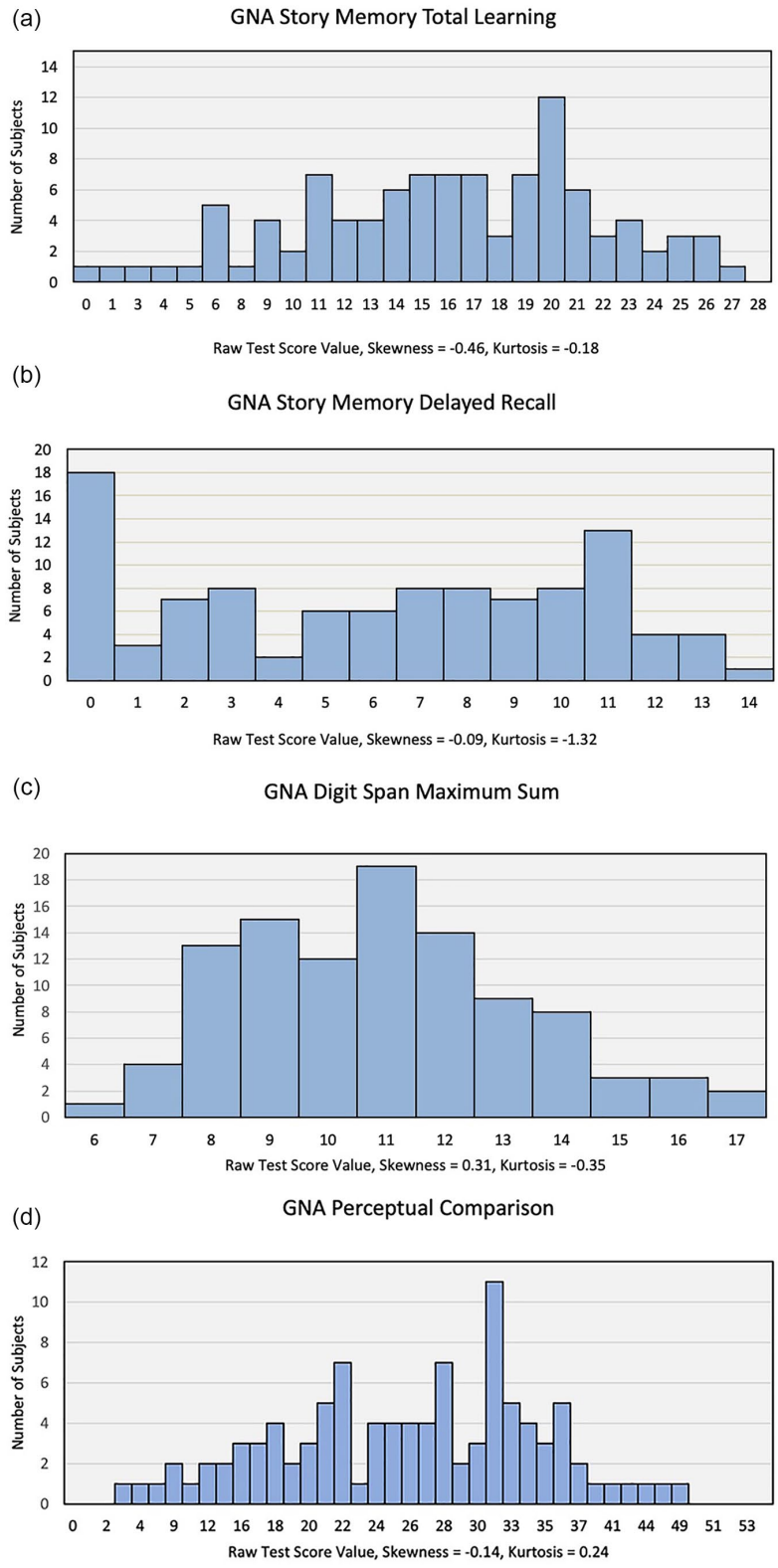

Figure 1. The frequency distributions of raw test scores for the Global Neuropsychological Assessment (GNA) Story Memory Learning (a), Story Memory Delayed Recall (b), Digit Span (c), and Perceptual Comparison (d) tests. These distributions reflect the combined sample of healthy controls, amnestic mild cognitive impairment, and Alzheimer's disease patients ($n = 103$).

Table 1. Demographics and GNA Characteristics for 103 Patients and Controls at Baseline.

Demographic/GNA characteristic	Probable AD (*N* =28)		Amnestic MCI (*N* = 9)		Healthy control (*N* = 66)			
	M	*SD*	*M*	*SD*	*M*	*SD*	Overall model	Group differences
Age (years)	71.32	6.87	73.89	9.94	71.64	6.73	$p = .025$	AD = MCI = HC
Sex (male/female)	14/14		4/5		23/43		$p = .373$	AD = MCI = HC
Race[a] (White/Black)	25/3		9/0		19/1		$p = .499$	AD = MCI = HC
Education	14.32	2.57	16.78	2.54	15.50	2.60	$p = .027$	AD = MCI = HC
GNA SM TL	10.14	4.61	14.78	5.49	18.58	4.55	$p < .001$	AD < MCI < HC
GNA SM Delay	2.04	2.38	2.11	2.32	8.47	3.30	$p < .001$	AD = MCI < HC
GNA DSpan Max Sum	9.50	2.03	10.89	1.96	11.55	2.39	$p < .001$	AD < MCI < HC
GNA PC	18.64	7.69	25.00	7.31	29.49	7.49	$p < .001$	AD < MCI = HC
GNA VF	9.50	3.40	17.33	6.61	21.30	4.57	$p < .001$	AD < MCI < HC

Note. GNA = Global Neuropsychological Assessment; AD = Alzheimer's disease; MCI = mild cognitive impairment; HC = health control; SM Total Learning and Delay = Story Memory sum of total learning trials 1 and 2 and delayed recall, respectively; DSpan = Digit Span; Max Sum = sum of maximum digit spans, forward and backward; PC = Perceptual Comparison (simple processing speed); VF = Verbal Fluency Animal Naming.
[a]Race/ethnicity was not recorded at the Johns Hopkins site prior to March 1, 2020, and cannot be obtained due to the absence of personal identifying information, as participants gave oral informed consent.

indicated that the diagnostic groups did not significantly differ in sex, $\chi^2(2, n = 103) = 1.97$, $p = .373$, and race/ethnicity, $\chi^2(3, n = 57) = 1.39$, $p = .499$. Groups did differ in age, $F(2, 100) = 3.84$, $p = .025$, and educational achievement, $F(2, 100) = 3.75$, $p = .027$.

GNA Performance in Patients Versus Healthy Controls

Figure 1 displays the frequency distributions of baseline GNA scores for all subjects. Table 1 displays the means and standard deviations for each GNA subtest measure by diagnostic group membership. For GNA Story Memory Total Learning, performance significantly differed between groups, $F(2, 102) = 32.66$, $p < .001$, with AD significantly lower than aMCI ($d = -0.9$) and HCs ($d = -1.8$) and aMCI significantly lower than HCs ($d = -0.8$). For GNA Story Memory Delayed Recall, performance also differed between groups, $F(2, 102) = 54.08$, $p < .001$. On this measure, probable AD ($d = -2.3$) and aMCI ($d = -2.3$) groups differed significantly from HCs but not from each other.

For GNA Digit Span Maximum Sum, performance significantly differed between groups, $F(2, 102) = 41.14$, $p < .001$, with AD significantly lower than aMCI ($d = -0.7$) and HCs ($d = -0.9$).

GNA Perceptual Comparison significantly distinguished groups, $F(2, 102) = 20.43$, $p < .001$, such that AD scored significantly lower than aMCI ($d = -0.8$) and HCS ($d = -1.4$), but performance did not significantly differ between aMCI and HCs.

GNA Animal Naming significantly distinguished groups, $F(2, 102) = 43.05$, $p < .001$, with AD significantly lower than aMCI ($d = -1.6$) and HCs ($d = -3.0$).

All models were repeated with analysis of covariance controlling for effects of age and educational achievement; controlling for these demographic variables did not significantly influence the main effects reported above.

Test–Retest Reliability

Test–retest reliability was evaluated with the intraclass correlation coefficient (ICC) for each GNA subtest for subjects with a diagnosis of probable AD or aMCI who completed both baseline and follow-up measures of the GNA ($n = 35$) across an approximately 30-day period ($M = 24.03$ days, $SD = 8.86$ days). Figure 2 displays mean scores for each GNA variable at baseline and follow-up. For the GNA Story Memory, ICC = 0.94 ($p < .001$) for Total Learning and 0.79 ($p < .001$) for Delayed Recall. For the GNA Digit Span task, test–retest reliability was ICC = 0.84 ($p < .001$). For the GNA Perceptual Comparison task, test–retest reliability was ICC= 0.93 ($p < .001$). The reliability coefficients did not significantly differ between the same- and alternate-form groups.

Construct Validity

Correlations between GNA subtests and conventional neurocognitive measures for patients (Table 2) indicate that GNA Story Memory Learning correlated most robustly with WMS-R Logical Memory Immediate Recall, $r = .65$, $p < .001$. GNA Story Memory Delayed Recall correlated most robustly with HVLT-R Delayed Recall, $r = .58$, $p < .001$. Results also indicate that the GNA Digit Span Max Sum was most robustly correlated with WAIS-III Digit Span Max Sum, $r = .75$, $p < .0001$. GNA Perceptual

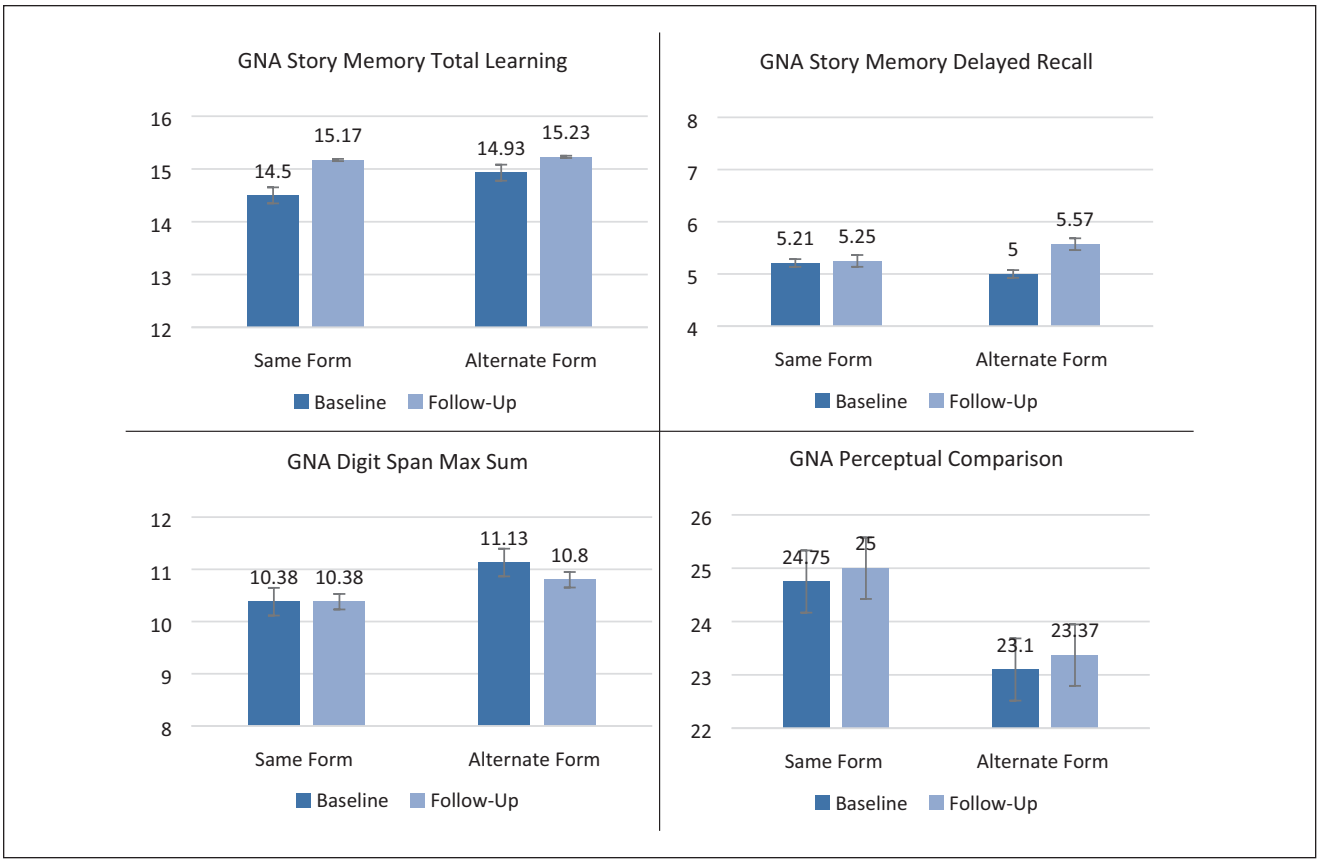

Figure 2. Test–retest means for same and alternate form administration of the Global Neuropsychological Assessment across subjects ($n = 35$). Error bars denote the standard error of the mean.

Table 2. Correlations Between GNA and Other Cognitive Measures in 37 Patients With Probable Alzheimer's Disease or Amnestic Mild Cognitive Impairment.

Standard test	GNA SM Total Learning	GNA SM Delay	GNA DSp Max Sum	GNA PC	GNA VF
WMS-R LM Immediate	0.65*	0.39	0.24	0.29	0.53*
WMS-R LM Delay	0.43*	0.43*	0.20	0.10	0.27
HVLT-R Learning	0.45*	0.51*	0.34	0.45*	0.47*
HVLT-R Delay	0.36	0.58*	0.20	0.17	0.32
WAIS-III DSp Max Sum	0.47*	0.07	0.75*	0.36	0.39*
TMT Part A	−0.31	−0.10	−0.08	−0.73*	−0.33*
TMT Part B	−0.24	0.21	−0.17	−0.40	−0.47*
CIFA Supermarket	0.47*	0.09	−0.19	0.56*	0.70*

Notes. Shaded cells denote measures hypothesized for convergent validity. GNA = Global Neuropsychological Assessment; SM Total Learning and Delay = Story Memory sum of total learning trials 1 and 2 and delayed recall, respectively; DSpan = Digit Span; Max Sum = sum of maximum digit spans, forward and backward; PC = Perceptual Comparison (simple processing speed); VF = Verbal Fluency Animal Naming; WMS-R LM = Wechsler Memory Scale–Revised Logical Memory, immediate and delayed recall; HVLT-R = Hopkins Verbal Learning Test–Revised, learning and delayed recall; TMT = Trail Making Test; CIFA Supermarket = supermarket item naming from the Calibrated Ideational Fluency Assessment.
*$p < .01$ (two-tailed).

Comparison was found to be most robustly correlated with Trail Making Test Part A, $r = -.73$, $p < .001$. Furthermore, GNA Animal Naming was most robustly correlated with CIFA Supermarket Items, $r = .70$, $p < .0001$.

Convergent validity correlations for HCs are shown in Table 3. Findings indicate that the GNA Story Memory Total Learning was most robustly correlated with WMS-R Logical Memory Immediate Recall, $r = .74$, $p < .001$.

Table 3. Correlations Between GNA and Other Cognitive Measures in 66 Aging Healthy Controls.

Standard test	GNA SM Total Learning	GNA SM Delay	GNA DSp Max Sum	GNA PC	GNA VF
WMS-R LM Immediate	0.74*	0.76*	0.14	0.03	0.14
WMS-R LM Delay	0.68*	0.78*	0.12	−0.06	0.05
HVLT-R Learning	0.36	0.15	0.12	0.54*	−0.01
HVLT-R Delay	0.31	0.12	0.08	0.32	−0.19
WAIS-III DSp Max Sum	−0.21	0.26	0.80*	0.12	0.28
TMT Part A	−0.17	0.03	−0.16	−0.59*	−0.38
TMT Part B	−0.14	−0.33	−0.40	−0.62*	−0.43
CIFA Supermarket	0.35	0.30	0.001	0.30	0.22

Note. Shaded cells denote measures hypothesized for convergent validity. GNA = Global Neuropsychological Assessment; SM Learning and Delay = Story Memory sum of learning trials 1 and 2 and delayed recall, respectively; DSp = Digit Span; Max Sum = sum of maximum digit spans, forward and backward; PC = Perceptual Comparison (simple processing speed); VF = Verbal Fluency Animal Naming; WMS-R LM = Wechsler Memory Scale–Revised Logical Memory, immediate and delayed recall; HVLT-R = Hopkins Verbal Learning Test–Revised, learning and delayed recall; TMT = Trail Making Test; CIFA Supermarket = supermarket item naming from the Calibrated Ideational Fluency Assessment.
*$p < .01$ (two-tailed).

GNA Story Memory Delayed Recall was most robustly correlated with WMS-R Logical Memory Delayed Recall, $r = .78$, $p < .001$. Results also indicate that the GNA Digit Span Max Sum was most robustly correlated with WAIS-III Digit Span Max Sum, $r = .80$, $p < .0001$. GNA Perceptual Comparison was found to be most robustly correlated with Trail Making Test Part A, $r = -.593$, $p < .001$. Furthermore, GNA Animal Naming was most robustly correlated with Trail Making Test Part A, $r = .38$, though this finding was not statistically significant, $p = .100$.

Discussion

To address the need for a repeatable, cross-cultural story memory task, we introduce the GNA. This is the first investigation of its psychometric properties, utilizing samples of AD and aMCI patients and healthy volunteers.

Analyses show good criterion-related validity as story learning and retention, CPS, digit repetition, and VF measures all discriminate groups in the expected direction. These effects are independent of the potentially confounding effects of age and education.

A closer inspection of the group effects reveals differences among pairs that are large and conforming with the degree of AD-like pathology as is well documented in the neuropsychology literature (Mortimer et al., 2004; Rabin et al., 2009; Snyder et al., 2011). The largest effect was found for Story Memory Delayed Recall where the group means of both AD and aMCI patients were roughly 2 *SD*s below that of HCs. This critical measure of verbal, contextual, and episodic memory is sensitive to both aMCI and AD but did not show clear patient subgroup differences in severity of memory dysfunction. In contrast, both patient subgroups performed far more poorly than healthy adults on Animal Naming, but the performance decrement was much greater in the probable AD than the aMCI subgroup (ds = −3.0 vs. −1.6, respectively). Further analyses showed that GNA measures of processing speed and working memory are sensitive to the demented state of AD, but show sparing in aMCI, conforming to the specificity of this diagnosis. The effect size distinguishing AD and HCs in CPS was $d = -1.5$, nearing that of $d = -2.3$ for delayed story recall. These findings conform with the hallmark presentations of aMCI and AD (Cherry et al., 2002).

Turning to our investigation of convergent and divergent validity, all of the patients and roughly one half of the HCs underwent a comprehensive neuropsychological battery encompassing the domains of language, spatial construction, memory, attention, processing speed, and executive function. This battery was presented on a separate appointment prior to exposure to the GNA. The time between exams varied, and no participant took the GNA testing before conventional neuropsychological testing. This failure to counterbalance for order means that different results might have been obtained had the GNA been presented first. That said, our data support the validity of the GNA assessments, as the strongest correlate of each GNA subtest was the standard neuropsychological measure that was predicted a priori. Thus, Story Memory measures correlate best with Logical Memory, Digit Span correlates strongly with Digit Span, GNA Perceptual Comparison correlates highly with the Trail Making Test, and Animal Naming correlates most highly with Supermarket Item Naming. These predicted relationships characterize both patient and HC samples.

The patients enrolled in this study were asked to return to the clinic a third time to repeat the GNA and complete questionnaires pertaining to personality and awareness of impairment (analysis of personality metrics are under analysis at the time of this writing). Most (37 of 39) agreed to do so and produced similar scores on follow-up testing. The mean test–retest interval of 24 days is adequate for measuring

reliability without contamination from maturation effects (e.g., aging, influence of ongoing work, or educational experience) but would not generalize to clinical practice where 1-year follow-ups are more routine. Nevertheless, this preliminary investigation suggests adequate psychometric validity, where the Story Memory reliability coefficients range from .77 to .94, Digit Span coefficient is .84, and the Perceptual Comparison coefficient is .93.

There are several limitations that render this a preliminary study. The sample size is suboptimal thereby limiting statistical power. The HCs were obtained from two different research centers with markedly different sampling techniques and incentives for participating. In one, participants were recruited via advertisement and were paid to participate in the study. In the other, healthy persons accompanying patients to the clinic volunteered to take the GNA without compensation. The GNA project is continuing to acquire normative data from several ongoing projects. As the project continues, we will increase the diversity of our normative sample, which currently is predominantly Caucasian (95.00%). In contrast, among people aged 65 years and older, African Americans have the highest prevalence of AD and related dementias (13.8%), followed by Hispanics (12.2%) and non-Hispanic Whites (10.3%), American Indian and Alaska Natives (9.1%), and Asian and Pacific Islanders (8.4%; Chen & Zissimopoulos, 2018). The mission of the GNA is to assess for dementia and other disorders in diverse, international samples, and our future work will be to target diverse populations.

This is a preliminary work where we endeavoured to pilot the GNA and assess for its acceptability to patients and examiners, and to determine if it meets basic psychometric standards for future use. Excluding the Spatial Span, Category Switching, and PHQ-4 subtests that were not administered for this study, the GNA takes less than 15 min to administer, is easily completed, and appears to be reliable and valid. We are proceeding in three directions. First, an investigation of the test–retest and interform reliability of GNA was recently completed, and the manuscript is in preparation. This study employed a proper Latin square research design that was beyond the resources available for the current project. Second, we have begun translating the GNA into other languages and acquiring normative data as a step toward the goal of creating a free test that will enable access to clinicians and researchers around the world. Third, we are investigating the sensitivity of GNA subtests to other clinical populations and as predictors of disease course and functional outcomes, including activities of daily living.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Lauren T. Olson https://orcid.org/0000-0001-7707-9325

References

Beery, K. E. (2004). Beery VMI: The Beery-Buktenica developmental test of visual-motor integration. Pearson.

Benedict, R. H., DeLuca, J., Phillips, G., LaRocca, N., Hudson, L. D., & Rudick, R., & Multiple Sclerosis Outcome Assessments Consortium. (2017). Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. *Multiple Sclerosis*, *23*(5), 721-733. https://doi.org/10.1177/1352458517690821

Benedict, R. H., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins Verbal Learning Test–Revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist*, *12*(1), 43-55. https://doi.org/10.1076/clin.12.1.43.1726

Chen, C., & Zissimopoulos, J. M. (2018). Racial and ethnic differences in trends in dementia prevalence and risk factors in the United States. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *4*(1), 510-520. https://doi.org/10.1016/j.trci.2018.08.009

Cherry, B. J., Buckwalter, J. G., & Henderson, V. W. (2002). Better preservation of memory span relative to supraspan immediate recall in Alzheimer's disease. *Neuropsychologia*, *40*(7), 846-852. https://doi.org/10.1016/s0028-3932(01)00173-7

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan executive function system*. Psychological Corporation. https://doi.org/10.1037/t15082-000

Folstein, M. F., Folstein, S., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189-198. https://doi.org/10.1016/0022-3956(75)90026-6

Gavett, B. E., Gurnani, A. S., Saurman, J. L., Chapman, K. R., Steinberg, E. G., Martin, B., Chaisson, C. E., Mez, J., Tripodis, Y., & Stern, R. A. (2016). Practice effects on story memory and list learning tests in the neuropsychological assessment of older adults. *PLOS ONE*, *11*(10), Article e0164492. https://doi.org/10.1371/journal.pone.0164492

Kaplan, E. F., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test*. Lea & Febiger.

Lee, S., Habeck, C., Razlighi, Q., Salthouse, T., & Stern, Y. (2016). Selective association between cortical thickness and reference abilities in normal aging. *Neuroimage*, *142*, 293-300. https://doi.org/10.1016/j.neuroimage.2016.06.041

McCaffrey, R. J., Ortega, A., & Haase, R. F. (1993). Effects of repeated neuropsychological assessments. *Archives of Clinical Neuropsychology*, *8*(6), 519-524. https://doi.org/10.1093/arclin/8.6.519

Mortimer, J. A., Gosche, K. M., Riley, K. P., Markesbery, W. R., & Snowdon, D. A. (2004). Delayed recall, hippocampal volume and Alzheimer neuropathology: findings from the Nun

Study. *Neurology*, *62*(3), 428-432. https://doi.org/10.1212/01.WNL.0000106463.66966.65

Newcomer, J. W., Craft, S., Fucetola, R., Moldin, S. O., Selke, G., Paras, L., & Miller, R. (1999). Glucose-induced increase in memory performance in patients with schizophrenia. *Schizophrenia Bulletin*, *25*(2), 321-335. https://doi.org/10.1093/oxfordjournals.schbul.a033381

Newcomer, J. W., Farber, N. B., Jevtovic-Todorovic, V., Selke, G., Melson, A. K., Hershey, T., Craft, S., & Olney, J. W. (1999). Ketamine-induced NMDA receptor hypofunction as a model of memory impairment and psychosis. *Neuropsychopharmacology*, *20*(2), 106-118. https://doi.org/10.1016/S0893-133X(98)00067-0

Newcomer, J. W., Selke, G., Melson, A. K., Hershey, T., Craft, S., Richards, K., & Alderson, A. L. (1999). Decreased memory performance in healthy humans induced by stress-level cortisol treatment. *Archives of General Psychiatry*, *56*(6), 527-533. https://jamanetwork.com/journals/jamapsychiatry/fullarticle/1673779

Rabin, L. A., Paré, N., Saykin, A. J., Brown, M. J., Wishart, H. A., Flashman, L. A., & Santulli, R. B. (2009). Differential memory test sensitivity for diagnosing amnestic mild cognitive impairment and predicting conversion to Alzheimer's disease. *Aging, Neuropsychology, and Cognition*, *16*(3), 357-376. https://doi.org/10.1080/13825580902825220

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*(3), 271-276. https://doi.org/10.2466/pms.1958.8.3.271

Rivera Mindt, M., Byrd, D., Saez, P., & Manly, J. (2010). Increasing culturally competent neuropsychological services for ethnic minority populations: A call to action. *The Clinical Neuropsychologist*, *24*(3), 429-453. https://doi.org/10.1080/13854040903058960

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of International Neuropsychological Society*, *16*(5), 754-760. https://doi.org/10.1017/S1355617710000706

Schretlen, D. J., Testa, S. M., & Pearlson, G. D. (2010). *Calibrated Neuropsychological Normative System professional manual*. Psychological Assessment Resources.

Schretlen, D. J., & Vannorsdall, T. (2010). *Calibrated Ideational Fluency Assessment professional manual*. Psychological Assessment Resources.

Smith, A. (1982). *Symbol digit modalities test: Manual*. Western Psychological Services.

Snyder, P. J., Jackson, C. E., Petersen, R. C., Khachaturian, A. S., Kaye, J., Albert, M. S., & Weintraub, S. (2011). Assessment of cognition in mild cognitive impairment: A comparative study. *Alzheimers Dementia*, *7*(3), 338-355. https://doi.org/10.1016/j.jalz.2011.03.009

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd ed.). Oxford University Press.

Steffener, J., Brickman, A. M., Habeck, C. G., Salthouse, T. A., & Stern, Y. (2013). Cerebral blood flow and gray matter volume covariance patterns of cognition in aging. *Human Brain Mapping*, *34*(12), 3267-3279. https://doi.org/10.1002/hbm.22142

Trifilio, E., Tanner, J. J., Butterfield, L., Mangal, P., Maye, J. E., Marsiske, M., Price, C. C., & Bowers, D. (2020). A tale of two stories: Validity of an alternative story memory test in a sample of older adults. *Clinical Neuropsychology*, *34*(1), 158-173. https://doi.org/10.1080/13854046.2018.1538428

Wechsler, D. (1939). *Wechsler-Bellevue Intelligence Scale*. Psychological Corporation. https://doi.org/10.1037/t06871-000

Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology*, *19*(1), 87-95. https://doi.org/10.1080/00223980.1945.9917223

Zhu, J., Tulsky, D. S., Price, L., & Chen, H.-Y. (2001). WAIS-III reliability data for clinical groups. *Journal of the International Neuropsychological Society*, *7*(7), 862-866. https://doi.org/10.1017/S1355617701777090