Received: 15 July 2020 | Accepted: 28 July 2020 | Published online: 30 September 2020

DOI: 10.1002/trc2.12072

Alzheimer's & Dementia
Translational Research & Clinical Interventions

RESEARCH ARTICLE

Development and validation of language and visuospatial composite scores in ADNI

Seo-Eun Choi[1] | Shubhabrata Mukherjee[1] | Laura E. Gibbons[1] | R. Elizabeth Sanders[1] | Richard N. Jones[2] | Douglas Tommet[2] | Jesse Mez[3] | Emily H. Trittschuh[4,5] | Andrew Saykin[6] | Melissa Lamar[7] | Laura Rabin[8] | Nancy S. Foldi[9] | Sietske Sikkes[10] | Roos J. Jutten[10] | Evan Grandoit[11] | Christine Mac Donald[12] | Shannon Risacher[6] | Colin Groot[10] | Rik Ossenkoppele[10,13] | for the Alzheimer's Disease Neuroimaging Initiative[14] | Paul K. Crane[1]

[1] Department of Medicine, University of Washington, Seattle, Washington, USA

[2] Department of Neurology, Brown University, Providence, Rhode Island, USA

[3] Department of Neurology, Boston University, Boston, Massachusetts, USA

[4] Department of Psychiatry, University of Washington, Seattle, Washington, USA

[5] Puget Sound Veterans Administration, Geriatric Research Education and Clinical Center, VA Puget Sound Health Care System, Seattle, Washington, USA

[6] Department of Radiology and Alzheimer's Research Center, Indiana University, Indianapolis, Indiana, USA

[7] Rush Alzheimer's Disease Center and Department of Behavioral Sciences and Psychiatry, Rush University Medical Center, Chicago, Illinois, USA

[8] Department of Psychology, City University of New York–Brooklyn, New York, USA

[9] Department of Psychology, City University of New York–Queens College, New York, USA

[10] Alzheimer Center, Amsterdam UMC - VU University Medical Center, Amsterdam, the Netherlands

[11] Department of Psychology, Northwestern University, Evanston, Illinois, USA

[12] Department of Neurological Surgery, University of Washington, Seattle, Washington, USA

[13] Clinical Memory Research Unit, Lund University, Lund, Sweden

Correspondence
Paul K. Crane, Department of Medicine, University of Washington, Box 359780, 325 Ninth Avenue, Seattle, WA 98104.
E-mail: pcrane@uw.edu

[14]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Abstract

Introduction: Composite scores may be useful to summarize overall language or visuospatial functioning in studies of older adults.

Methods: We used item response theory to derive composite measures for language (ADNI-Lan) and visuospatial functioning (ADNI-VS) from the cognitive battery administered in the Alzheimer's Disease Neuroimaging Initiative (ADNI). We evaluated the scores among groups of people with normal cognition, mild cognitive impairment (MCI), and Alzheimer's disease (AD) in terms of responsiveness to change, association with imaging findings, and ability to differentiate between MCI participants who progressed to AD dementia and those who did not progress.

Results: ADNI-Lan and ADNI-VS were able to detect change over time and predict conversion from MCI to AD. They were associated with most of the pre-specified magnetic

resonance imaging measures. ADNI-Lan had strong associations with a cerebrospinal fluid biomarker pattern.

Discussion: ADNI-Lan and ADNI-VS may be useful composites for language and visuospatial functioning in ADNI.

KEYWORDS

Alzheimer's Disease Neuroimaging Initiative, cognition, language, longitudinal analysis, magnetic resonance imaging, psychometrics, visuospatial functioning

1 | BACKGROUND

Language and visuospatial functioning impairments often accompany memory impairments among individuals with Alzheimer's disease (AD). Assessment of these domains may be useful in clinical and research settings.[1] For example, scores can be used to track changes over time[2–4] and describe natural history,[5] to evaluate the validity of fluid-based or imaging-based biomarkers,[6,7] or to identify patterns of deficits across domains.[8]

The Alzheimer's Disease Neuroimaging Initiative (ADNI) administered an extensive neuropsychological battery at every study visit.[9] The battery includes measures of language and visuospatial functioning. Investigators using ADNI data are faced with an array of scores. Composite scores for language and visuospatial functioning may be useful for theoretical and empirical/statistical reasons.[10–12] Summarizing data in these domains with a single score facilitates analyses without testing multiple hypotheses. Including multiple indicators minimizes measurement error due to idiosyncratic single items. Scores used across multiple investigations may facilitate comparisons and understanding.[13,14]

Our goal was to develop new composite scores for language (ADNI-Lan) and visuospatial functioning (ADNI-VS) and to evaluate these scores, along with standard scores and with z-score composites. This approach is similar to our prior work for memory[10] and executive functioning.[12]

2 | METHODS

2.1 | Participants and data source

ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations. Study resources and data are available through its website (http://adni.loni.usc.edu). The initial 5-year study (ADNI1) was extended by 2 years in 2009 (ADNIGO), and in 2011 and 2016 by further competitive renewals (ADNI2 and ADNI3). The study was conducted after Institutional Review Board approval at each site. Written informed consent was obtained from study participants or authorized representatives.

To date, 2985 people were evaluated by ADNI for possible enrollment; enrolled participants have been followed for up to 156 months. Of these, 2084 had screening and baseline visits, received a diagnosis, and were enrolled in ADNI. Of these, 1913 were age ≥60 at enrollment from ADNI1, ADNIGO, and ADNI2 and had imaging data that passed quality control.

We analyzed participants with normal cognition (NC), amnestic mild cognitive impairment (MCI), and AD. Diagnostic criteria are on the ADNI website. Briefly, diagnosis of amnestic MCI required participant memory complaints, objective memory deficits, intact activities of daily living, global Clinical Dementia Rating score of 0.5,[15] and Mini-Mental State Examination (MMSE) score ≥24.[16] Participants with AD met National Institute of Neurological Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINDS-ADRDA) criteria for probable AD.[17] Conversion from NC or MCI to AD was a primary outcome (http://adni.loni.usc.edu/methods/documents). All diagnostic determinations were made without referring to the composite scores discussed here.

2.2 | Cognitive and clinical measures

2.2.1 | Language indicators and comparisons

ADNI's neuropsychological battery included the Boston Naming Test[18] and animal and (for ADNI1 only) vegetable fluency. ADNI administered MMSE[16] items including object naming, sentence repetition, sentence reading and writing, and following a three-step command. ADNI administered Alzheimer's Disease Assessment Schedule–Cognition (ADAS-Cog)[19] items including following commands, object naming, and ideational praxis. ADNI administered Montreal Cognitive Assessment (MoCA)[20] items beginning in ADNIGO including phonemic fluency and sentence repetition. ADNI did not use alternate forms for any of its language measures.

We used these language indicators to derive z-scores and the ADNI-Lan composite. We used the three tests given at each wave for a "least common denominator (LCD)" z-score strategy. We used all of the tests within each wave for a "wave specific (WS)" z-score strategy. Methods for the LCD and WS strategies are discussed in detail in Methods S1 in supporting information. We compared ADNI-Lan to the LCD and WS z-scores and to scores for animal

fluency, vegetable fluency, and the Boston Naming Test and to sum scores from language items from the MMSE, ADAS-Cog, and MoCA.

2.2.2 | Visuospatial functioning indicators and comparisons

ADNI's neuropsychological battery included copying a clock, with five scored elements including circular face, symmetrical face, copying numbers for hours, presence of hands, and hands pointing to the requested time. The MMSE includes copying interlocking pentagons. The ADAS-Cog includes a constructional praxis item. ADNI did not use alternate forms for any of its visuospatial measures. We used all of these to develop the ADNI-VS composite. We compared that composite to the standard score for the clock and to a z-score made from all the indicators derived from the mean and standard deviation from all NC participants across all waves.

2.2.3 | Magnetic resonance imaging parameters

All participants were scheduled to have a magnetic resonance imaging (MRI) evaluation at each study visit.[21] ADNI magnetization-prepared rapid acquisition with gradient echo (MPRAGE) sequences have been processed using Freesurfer to characterize volumes and cortical thickness in prespecified regions.[21] We selected specific regions for language[22–25] and visuospatial function[26–29] based on prior literature and downloaded variables from the ADNI website for volumes and cortical thickness for those regions (see Tables 1 and 2). ADNI used different scanning protocols at different study waves. All our imaging analyses were cross-sectional (ie, within a wave) and thus used either 1.5T *or* 3.0T MRI data, but never both in the same analyses.

2.2.4 | Cerebrospinal fluid

Some ADNI1 participants had baseline lumbar punctures for cerebrospinal fluid (CSF), which was evaluated for amyloid $\beta_{1\text{-}42}$ (Aβ), total tau, and phosphorylated tau_{181p} (p-tau). Sample collection and analysis are described in detail in Shaw et al.[30] De Meyer et al.[31] used Aβ and p-tau to classify ADNI participants as having an AD signature or not, and provided us with their categories.

2.3 | Psychometric analyses of baseline data

We performed all psychometric analyses using Mplus[32] or R.[33] Almost all of the language items are either dichotomous (right/wrong) or ordered categorical variables. The Boston Naming Test and the fluency items are counts of correctly identified objects and the number of words in 60 seconds. Count items can be modeled using a Poisson distribution, or categorized and treated as ordinal categorical variables. The maximum likelihood estimator with robust standard errors (MLR) allows Poisson distributions for count variables. Both MLR and weighted least squares with estimated mean and variance estimators can be used for dichotomous or ordinal categorical data. The Poisson distribution model would use fewer parameters but assumes count variables follow Poisson distributions. We categorized responses (Table S1 in supporting information) and compared Poisson distribution and ordinal categorical models using the MLR estimator. Factor loadings for non-count items were similar to each other (Tables S2–S4 in supporting information). However, scores derived from these models were substantially different, defined as >0.3 logit units, for a sizable proportion of the cohort, defined as >5%: 11% to 13% of scores were >0.3 units different for ADNI1, ADNI2/GO, and ADNI3 (see Figure S1 and its note in supporting information). Our final model treated these items as ordinal categorical indicators.

HIGHLIGHTS

- Composite scores can be useful for studies of cognitive functioning.
- Investigators developed scores for language and visuospatial functioning.
- The composite scores showed good evidence of validity across several analyses.
- Language and visuospatial composite scores are now available from the Alzheimer's Disease Neuroimaging Initiative web site.

RESEARCH IN CONTEXT

- **Systematic Review**: The authors reviewed the literature using traditional (eg, PubMed) sources. There have been few articles developing composite scores for language or visuospatial functioning, and fewer still comparing the performance of composite scores to standard scoring.
- **Interpretation**: Our findings support the use of the ADNI-Lan or ADNI-VS scores where composite scores for language or visuospatial functioning may prove useful.
- **Future directions**: Researchers have cited previous articles reporting the development of composite measures of memory (ADNI-Mem) and executive functioning (ADNI-EF). ADNI-Lan and ADNI-VS promise to provide similarly useful composites for language and visuospatial functioning.

TABLE 1 Adjusted associations between language scores and MRI-derived measures of left-sided cortical thickness for selected brain regions[a]

Group and measure	Angular gyrus	Wernicke's area	Temporal pole	Inferior frontal gyrus
MCI				
ADNI1				
ADNI-Lan	**4.09**	**6.18**	**4.05**	**3.15**
LCD z-score	**3.61**	**5.29**	**2.99**	**2.08**
WS z-score	**3.82**	**5.61**	**3.27**	**2.15**
Boston Naming (Total)	**4.34**	**7.14**	**4.57**	**3.41**
Category Fluency – Animals	**3.18**	**4.49**	**2.98**	**3.31**
Category Fluency – Vegetables	**3.22**	**4.77**	**3.35**	1.95
MMSE Language Items	0.05	−0.51	−1.41	−0.65
ADAS-Cog Language Items	**2.07**	**2.75**	1.62	0.30
ADNI2/GO				
ADNI-Lan	**3.48**	**6.01**	**4.49**	**1.56**
LCD z-score	**3.15**	**6.07**	**5.27**	**2.33**
WS z-score	**3.50**	**6.14**	**4.97**	**2.30**
Boston Naming (Total)	1.70	**5.45**	**5.70**	0.89
Category Fluency – Animals	**3.66**	**5.10**	**2.63**	1.50
MMSE Language Items	0.87	**2.12**	1.68	**2.09**
ADAS Language Items	**2.00**	**2.50**	**2.64**	1.38
MoCA – Animal Naming	1.49	**2.51**	**3.27**	0.30
MoCA – All Language Items	**2.16**	**3.59**	**2.54**	1.30
AD				
ADNI1				
ADNI-Lan	**4.77**	**5.82**	**5.63**	**4.11**
LCD z-score	**5.28**	**7.27**	**7.46**	**4.45**
WS z-score	**5.50**	**7.31**	**7.42**	**4.65**
Boston Naming (Total)	**3.88**	**7.37**	**7.46**	**3.60**
Category Fluency – Animals	**4.13**	**3.61**	**2.53**	**3.53**
Category Fluency – Vegetables	**3.50**	**3.91**	**3.64**	**3.16**
MMSE Language Items	0.69	1.41	1.64	1.26
ADAS-Cog Language Items	**5.00**	**5.67**	**6.55**	**3.38**
ADNI2/GO				
ADNI-Lan	**5.45**	**6.03**	**4.72**	**3.09**
LCD z-score	**4.88**	**5.93**	**4.20**	**3.00**
WS z-score	**5.89**	**6.76**	**4.45**	**3.44**
Boston Naming (Total)	**3.92**	**5.86**	**5.80**	**2.78**
Category Fluency – Animals	**4.19**	**3.80**	**2.75**	**2.22**
MMSE Language Items	−0.94	**2.53**	1.80	0.76
ADAS Language Items	**4.75**	**5.81**	**4.92**	**2.66**
MoCA – Animal Naming	**4.70**	**4.77**	**5.37**	**2.21**
MoCA – All Language Items	**5.47**	**6.48**	**4.01**	**3.46**

[a] Test statistics for MRI cortical thickness measures (left hemisphere, adjusted for intracranial volume) from regression models for the language score controlling for age, education, sex, presence of ≥1 *APOE-ε4* alleles, by baseline diagnosis and ADNI phase. ADAS-Cog scores were reversed, so higher implies better performance. Bolded values indicate P values < 0.05.

Abbreviations: ADAS, Alzheimer's Disease Assessment Schedule; ADAS-Cog, Alzheimer's Disease Assessment Schedule–Cognition; ADNI, Alzheimer's Disease Neuroimaging Initiative; ADNI-Lan, ADNI language composite; LCD, least common denominator; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; MRI, magnetic resonance imaging; WS, wave specific.

TABLE 2 Adjusted associations between visuospatial scores and MRI-derived measures of cortical thickness for selected brain regions[a]

Group and measure	Superior temporal gyrus	Inferior parietal	Inferior temporal gyrus	Inferior frontal gyrus	Fusiform	Lingual gyrus	Lateral occipital	Pericalcarine
MCI								
ADNI1								
ADNI-VS	**3.66**	**4.15**	**4.77**	**3.27**	**3.85**	**2.45**	**3.04**	**2.14**
Z-score	**3.76**	**3.97**	**4.58**	**3.15**	**3.77**	**2.43**	**3.05**	1.78
Clock Score	**3.43**	**3.15**	**4.17**	**2.94**	**2.97**	1.38	**2.09**	1.58
ADNI 2/GO								
ADNI-VS	**2.50**	**2.25**	1.26	1.72	**2.03**	1.01	**1.99**	−0.85
Z-score	**2.56**	**2.62**	1.18	1.93	**1.85**	1.36	**2.11**	−0.41
Clock score	**2.72**	1.59	1.41	**2.12**	**2.19**	0.56	0.84	−1.41
AD								
ADNI 1								
ADNI-VS	**2.07**	**4.45**	**1.89**	**2.61**	**2.49**	1.30	**3.79**	1.37
Z-score	**2.80**	**6.27**	**2.96**	**3.00**	**4.40**	**3.42**	**6.09**	**2.67**
Clock score	**2.02**	**5.61**	**2.41**	**2.26**	**3.60**	**3.53**	**5.59**	1.76
ADNI 2/GO								
ADNI-VS	0.37	**4.42**	**2.03**	0.20	**3.48**	**2.68**	**4.05**	1.37
Z-score	1.60	**6.63**	**3.06**	0.09	**5.58**	**4.29**	**6.50**	**2.49**
Clock score	1.25	**5.54**	**3.02**	1.44	**4.11**	1.43	**4.25**	1.30

[a] Test statistics for MRI cortical thickness measures (adjusted for intracranial volume) from regression models for the visuospatial outcomes controlling for age, education, sex, presence of ≥1 *APOE-ε4* alleles, by baseline diagnosis and ADNI phase. Bolded values indicate *P* values < 0.05.
Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; ADNI-VS, ADNI visuospatial composite; LCD, least common denominator; MCI, mild cognitive impairment; MRI, magnetic resonance imaging.

Second, we evaluated dimensionality. The simplest confirmatory factor analysis (CFA) models are single-factor models in which all items are conceptualized as indicators of a single underlying factor. A more complicated model is a bi-factor model, in which, in addition to a factor defined by all the items, distinct subsets of items are also conceptualized as indicators of secondary factors. For example, for visuospatial, there are five clock items that could share methods variance beyond their relationship with visuospatial functioning defined by all the items.

We fit single-factor and bi-factor models to the data. We considered fit statistics including the confirmatory fit index (CFI), the Tucker-Lewis index (TLI), and the root mean squared error of approximation (RMSEA). Useful thresholds for these indices for model fit include CFI >0.95, TLI >0.95, and RMSEA <0.08.[34] We also compared scores using single-factor and bi-factor models, using the approach outlined above; if using the simpler but theoretically less justified single-factor model was associated with differences in scores >|0.3| for >5% of participants, we would choose the bi-factor model. Otherwise we would choose the simpler single-factor model.

For language we compared single-factor models to several candidate bi-factor models. For ADNI1, one of these was theory-driven based on item content. The other two models were data-driven using the Omega package in R[33] with different thresholds for salient loadings for secondary domains (see Methods S2 in supporting information). Fit statistics for all these models were good (see Table S5 in supporting information). Factor loadings for two bi-factor models had negative loadings on secondary domains, suggesting overfitting. There were no participants with scores that differed by as much as 0.3.

For ADNI2/Go, we used two candidate bi-factor models, one developed from Omega and the other using cluster analyses as shown in our recent article,[8] which we evaluated because the theoretical justification for secondary factors from Omega was weaker (see Methods S2 and Table S6 in supporting information). Fit statistics for all models were good, and there were few participants (9 for the cluster analysis-based model, or 0.6%, and 12 for the Omega-based model, or 0.8%). These results led us to choose a single-factor model for language (see Methods S1 and S2).

For the visuospatial domain, the Omega package suggested a single bi-factor structure, with three clock items in one secondary domain and the other two in another, and the remaining items in a third secondary domain. Fit statistics were nearly identical for single- and bi-factor models. There was one participant with a score that was different by >|0.3| across all ADNI enrollment waves (n = 2985). Fit statistics for the single-factor model were good, with TLI 0.97, CFI 0.96, and RMSEA 0.026. These results led us to choose a single-factor model for visuospatial.

TABLE 3 Baseline demographic, clinical data by baseline diagnosis

	Normal Cognition (NC)		
	ADNI1 n = 229	**ADNI2/GO n = 401**	**ADNI3 n = 334**
Baseline demographic characteristics			
Female, n (%)	110 (48%)	226 (56%)	203 (61%)
Age in years, mean (SD)	76.0 (5.0)	72.8 (6.1)	70.8 (5.8)
Education in years, mean (SD)	16.1 (2.9)	16.6 (2.6)	16.9 (2.2)
≥1 *APOE-ε4* allele, n (%)	61 (27%)	87 (30%)	86 (26%)
Baseline language data, mean (SD)			
ADNI-Lan	0.80 (0.71)	0.71 (0.66)	0.80 (0.73)
LCD z-score[a]	−0.06 (0.61)	−0.03 (0.64)	0.08 (0.60)
WS z-score	0.05 (0.54)	0.03 (0.55)	−0.004 (0.56)
MMSE Language Items	7.84 (0.42)	7.82 (0.43)	7.83 (0.41)
ADAS-Cog Language Items	14.80 (0.53)	14.77 (0.57)	14.84 (0.45)
Category Fluency – Animals	19.93 (5.59)	20.93 (5.29)	22.00 (5.63)
Category Fluency – Vegetables	14.70 (3.91)	NA	NA
Boston Naming (Total)	27.89 (2.29)	28.22 (2.10)	NA
MoCA – Animal Naming	NA	2.90 (0.30)	2.93 (0.27)
MoCA – All Language Items	NA	5.44 (0.74)	5.52 (0.75)
Baseline Visuospatial data, mean (SD)			
ADNI-VS	0.27 (0.62)	0.20 (0.56)	0.20 (0.60)
Z-score	0.03 (0.97)	0.01 (0.93)	−0.04 (1.09)
Clock score	4.86 (0.42)	4.87 (0.37)	4.85 (0.41)
	Mild Cognitive Impairment (MCI)		
	ADNI1 n = 385	**ADNI2/GO n = 546**	**ADNI3 n = 102**
Baseline Demographic Characteristics			
Female, n (%)	136 (35%)	241 (44%)	38 (37%)
Age in years, mean (SD)	75.5 (6.8)	72.8 (6.9)	73.5 (7.3)
Education in years, mean (SD)	15.6 (3.0)	16.2 (2.7)	16.6 (2.5)
≥1 *APOE-ε4* allele, n (%)	209 (54%)	210 (48%)	28 (27%)
Baseline language data, mean (SD)			
ADNI-Lan	−0.04 (0.76)	0.34 (0.71)	0.24 (0.57)
LCD z-score[a]	−0.71 (0.89)	−0.42 (0.78)	−0.35 (0.63)
WS z-score	−0.95 (1.18)	−0.39 (0.87)	−0.51 (0.59)
MMSE Language Items	7.65 (0.56)	7.66 (0.59)	7.70 (0.52)
ADAS-Cog Language Items	14.41 (0.86)	14.60 (0.72)	14.67 (0.54)
Category Fluency – Animals	15.82 (4.92)	18.01 (5.19)	17.52 (4.34)
Category Fluency – Vegetables	10.71 (3.45)	NA	NA
Boston Naming (Total)	25.47 (4.10)	26.75 (3.45)	NA
MoCA – Animal Naming	NA	2.82 (0.43)	2.74 (0.48)
MoCA – All Language Items	NA	5.19 (0.97)	4.87 (1.02)
Baseline visuospatial data, mean (SD)			
ADNI-VS	−0.10 (0.80)	0.01 (0.69)	−0.01 (0.68)
Z-score	−0.59 (1.44)	−0.35 (1.18)	−0.39 (1.33)
Clock score	4.64 (0.68)	4.72 (0.56)	4.71 (0.58)

TABLE 3 Continued.

	Alzheimer's Disease (AD)		
	ADNI1 n = 186	ADNI2/GO n = 177	ADNI3 n = 30
Baseline demographic characteristics			
Female, n (%)	85 (46%)	72 (41%)	12 (40%)
Age in years, mean (SD)	76.1 (6.7)	75.6 (7.4)	75.2 (8.7)
Education in years, mean (SD)	14.7 (3.1)	15.8 (2.8)	15.9 (3.1)
≥1 *APOE-ε4* allele, n (%)	123 (66%)	94 (67%)	7 (23%)
Baseline language data, mean (SD)			
ADNI-Lan	−0.75 (0.87)	−0.59 (0.99)	−0.24 (0.87)
LCD z-score[a]	−1.48 (1.39)	−1.58 (1.55)	−1.31 (1.06)
WS z-score	−2.12 (1.86)	−1.80 (1.41)	−1.21 (0.97)
MMSE Language Items	7.51 (0.71)	7.48 (0.88)	7.30 (1.06)
ADAS-Cog Language Items	13.70 (1.49)	13.59 (1.70)	14.06 (0.93)
Category Fluency – Animals	12.38 (4.91)	12.14 (5.13)	12.87 (5.60)
Category Fluency – Vegetables	7.81 (3.34)	NA	NA
Boston Naming (Total)	22.21 (6.26)	21.91 (5.95)	NA
MoCA – Animal Naming	NA	2.42 (0.81)	2.81 (0.40)
MoCA – All Language Items	NA	4.17 (1.53)	4.56 (1.31)
Baseline visuospatial data, mean (SD)			
ADNI-VS	−0.59 (0.93)	−0.38 (0.87)	−0.46 (0.88)
Z-score	−1.68 (2.18)	−1.38 (2.07)	−1.57 (2.26)
Clock score	4.34 (0.98)	4.44 (0.91)	4.20 (1.15)

[a] LCD, "least common denominator" z score strategy; WS, "wave specific" z score strategy; see text for details.
Abbreviations: ADAS, Alzheimer's Disease Assessment Schedule; ADAS-Cog, Alzheimer's Disease Assessment Schedule–Cognition; ADNI, Alzheimer's Disease Neuroimaging Initiative; ADNI-Lan, ADNI language composite; ADNI-VS, ADNI visuospatial composite; APOE, apolipoprotein E; LCD, least common denominator; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; MRI, magnetic resonance imaging; SD, standard deviation; WS, wave specific.

2.4 | Determination of scores from other study visits

We fixed item thresholds and loadings based on the baseline study visit. We used those item parameters to obtain ADNI-Lan and ADNI-VS scores at subsequent visits using Mplus.[32]

2.5 | Comparisons of scores

We performed several analyses to evaluate the validity of our composite scores. For all language analyses, we performed analyses separately by study wave because the battery of language items differed by wave. We also performed imaging analyses separately for 1.5T and 3T MRI platforms. Because all analyses are descriptive, we did not adjust *P* values to account for multiple hypotheses.

2.5.1 | Rates of change

We used mixed effects models with random intercepts and slopes for each individual to characterize rates of change for people with NC, MCI, and AD, adjusting for age, education, sex, and apolipoprotein E (*APOE*) genotype coded as ≥1 versus 0 ε4 alleles. We used slope terms and standard errors to determine sample sizes needed per group to detect a 25% reduction in rate of decline in 12 months with 80% power and $\alpha = 0.05$, assuming a two-sided test.

2.5.2 | Time to conversion for people with NC, MCI

For NC and MCI participants, we used Cox proportional hazard models of time to MCI and to AD controlling for age, education, sex, and *APOE* genotype.

2.5.3 | Strength of association with MRI parameters

We used standard linear regression models to determine strength of association between cognitive scores and selected MRI values from baseline, adjusting for total intracranial volume, age, education, sex, and *APOE* genotype.

2.5.4 | Ability to differentiate trajectories of participants with MCI with the De Meyer et al. AD CSF biomarker profile

We used the CSF data discussed in 2.2.4 to divide people with MCI into those with and without the AD biomarker profile. We used linear mixed effects models to determine the cognitive score's ability to differentiate trajectories of participants with and without AD biomarker profiles. We used random intercepts and slopes and an unstructured covariance matrix, controlling for age, education, sex, and *APOE* genotype.

3 | RESULTS

3.1 | Characteristics of participants

Demographic, cognitive, and imaging data for the 1913 included participants from ADNI1 and ADNI2/GO are shown in Table 3. The ADNI-Lan composite score captured differences in language performance across the three diagnostic groups ($P < 0.0001$). The MMSE and ADAS-Cog language items did not differentiate clinically meaningful differences between people with NC and people with MCI, though scores were lower for people with AD (both $P < 0.0001$ overall). The Boston Naming Test and animal and vegetable fluency distinguish well across diagnostic groups (all $P < 0.0001$).

The ADNI-VS composite score also captures differences in performance across the three diagnostic groups ($P < 0.0001$) and did so with more clinically meaningful differences than the clock copy score (still $P < 0.0001$).

Cortical thickness data for regions associated with language and visuospatial functioning respectively, are shown in the bottom part of Table 3. There were modest differences across groups in some regions (see Table S6).

3.2 | Sample sizes needed to detect changes over time

The 1610 included participants accrued 6598 person-years of follow-up (mean 4.1 years, standard deviation [SD] 2.9 years, range 6 months to 13 years, interquartile range 3 to 6 years). Table 4 shows test statistics from mixed effects models of change over time within each diagnostic group, along with sample sizes that would be needed to show 25% differences across groups. Neither language nor visuospatial functioning showed much change over time among participants with NC (left column). For participants with MCI, the ADNI-Lan score required the smallest sample, and in AD, the WS z-score was smallest, though none of these differences were tested statistically. The z-score required the smallest sample for visuospatial functioning.

3.3 | Ability to predict conversion

Table 5 shows results for analyses of conversions from NC to MCI and from MCI to AD. Baseline variation in language and visuospatial functioning was not strongly associated with risk of conversion from NC to MCI (left column). Point estimates for ADNI-Lan were similar to comparators. Baseline language and visuospatial functioning were associated with increased risk of conversion from MCI to AD. For language, the point estimate for ADNI-Lan was the best of those considered, including the two z-scores, at predicting conversion to AD, while for visuospatial functioning, all point estimates were similar. As in section 3.2, these comparisons are descriptive.

3.4 | Strength of association with a priori selected MRI measures

There were no a priori selected regions with strong associations with language measures among people with NC (not shown). Table 1 shows associations for language scores. ADNI-Lan had statistically significant associations with all regions in all samples, except for one. Table 2 shows strength of associations between visuospatial scores and a priori selected regions for participants with MCI and AD. Strong associations with ADNI-VS were observed for most of these regions, as they were for the z-score and the clock Score.

3.5 | Ability to differentiate trajectories for people with MCI with the AD CSF biomarker pattern

Table 6 shows test statistics for the effect of having the CSF biomarker pattern on the intercept and slope for language (top) and visuospatial functioning (bottom) in adjusted mixed effects models. Having the AD CSF biomarker profile was associated with lower language score intercepts for almost all the measures, though only ADNI-Lan was distinguishable from zero. Having the AD CSF biomarker pattern was associated with more rapid rates of decline for each of the language measures. Test statistics for the ADNI-Lan were larger than those of any comparator (not tested statistically). Having the AD CSF biomarker signature was not associated with visuospatial functioning intercepts. Of the slopes, only the z score was statistically significant.

4 | DISCUSSION

In this article we present methods used to derive composite language and visuospatial scores from ADNI's neuropsychological battery, and present results of analyses addressing validity. The ADNI-Lan measure performed well in comparison to other language scores that can be derived from the ADNI battery. ADNI-Lan has the advantage of incorporating all of the language indicators. This is particularly appealing in the context of the changing battery across ADNI's different waves. Fewer items were administered to assess visuospatial functioning. ADNI-VS, a z-score, and the clock composite score performed similarly in our comparative validity analyses.

There have been few previous composite scores for language or visuospatial functioning in the literature. Deters et al. used principal

TABLE 4 Test statistics, coefficients for time, and resulting sample sizes needed per group to detect a 25% change for language and visuospatial functioning measures[a]

	Normal Cognition		Mild Cognitive Impairment		Late Onset Alzheimer's Disease	
Language scores by study phases	Test statistic (Coefficient) for time	Sample size per group	Test statistic (Coefficient) for time	Sample size per group	Test statistic (Coefficient) for time	Sample size per group
All phases						
ADNI-Lan	−9.4 (−0.06)	17,000	−16.6 (−0.14)	1700	−14.7 (−0.37)	200
LCD z-score	−5.7 (−0.05)	55,000	−12.3 (−0.17)	2500	−9.8 (−0.43)	200
WS z-score	−1.4 (−0.02)	359,000	−11.3 (−0.15)	1800	−11.7 (−0.42)	150
Category Fluency – Animals	−7.2 (−0.05)	38,000	−16.8 (−0.12)	4100	−10.8 (−0.31)	600
MMSE Language Items	−3.2 (−0.03)	278,000	−7.9 (−0.10)	13,600	−6.1 (−0.31)	1200
ADAS-Cog Language Items	−2.9 (−0.03)	207,000	−10.8 (−0.15)	3600	−9.9 (−0.41)	300
ADNI1 only						
ADNI-Lan	−8.2 (−0.08)	11,000	−14.8 (−0.28)	900	−11.7 (−0.37)	200
LCD z-score	−5.1 (−0.06)	38,000	−10.5 (−0.21)	1500	−7.6 (−0.38)	250
WS z-score	0.45 (0.01)	1,318,000	−9.2 (−0.19)	1000	−8.9 (−0.40)	160
Category Fluency – Vegetables	−0.9 (−0.01)	405,000	−10.9 (−0.18)	2200	−9.1 (−0.28)	800
ADNI 1, 2, GO only						
ADNI-Lan	−9.4 (−0.06)	16,000	−16.6 (−0.14)	1700	−14.7 (−0.37)	200
LCD z-score	−5.7 (−0.05)	55,000	−12.3 (−0.17)	2500	−9.8 (−0.43)	220
WS z-score	−1.4 (−0.02)	338,000	−11.3 (−0.15)	1800	−11.8 (−0.42)	150
Boston Naming (Total)	−4.1 (−0.03)	70,000	−11.4 (−0.11)	3600	−11.4 (−0.28)	300
ADNI 2, GO, 3 only						
ADNI-Lan	−4.9 (−0.05)	78,000	−8.9 (−0.10)	4400	−8.8 (−0.38)	200
LCD z-score	−2.9 (−0.04)	165,000	−7.0 (−0.13)	5100	−7.0 (−0.54)	200
WS z-score	−2.5 (−0.04)	66,000	−6.5 (−0.11)	3800	−8.9 (−0.51)	140
MoCA –Animal Naming	−2.0 (−0.03)	184,000	−3.6 (−0.05)	59,100	−0.9 (−0.08)	17,100
MoCA – All Language Items	−1.5 (−0.02)	148,000	−4.1 (−0.05)	38,100	−4.16 (−0.23)	2000
Visuospatial functioning scores (all phases)						
ADNI-VS	−5.4 (−0.05)	70,000	−11.0 (−0.10)	11,700	−8.2 (−0.26)	1200
Z-score	−5.3 (−0.05)	53,000	−11.2 (−0.12)	5700	−8.6 (−0.31)	600
Clock score	−3.8 (−0.04)	123,000	−9.2 (−0.11)	11,200	−9.0 (-0.37)	700

[a] These data were derived from mixed effects models with cognitive score as the dependent variable in each model. Models included terms for age, education, sex, and the presence of ≥1 APOE ε4 allele. The "Test statistic (coefficient for time)" column shows test statistics (ie, estimates divided by their standard errors) for the adjusted rate of change, and shows the beta coefficient itself in parentheses. These values can be used to estimate sample sizes needed to show a 25% difference in the rate of decline with 80% power and using a two-sided α of 0.05. Coefficients were rounded to the nearest 0.1, and sample sizes rounded to the nearest 100 or for sample sizes >10,000, to the nearest 1000.

Abbreviations: ADAS, Alzheimer's Disease Assessment Schedule; ADAS-Cog, Alzheimer's Disease Assessment Schedule–Cognition; ADNI, Alzheimer's Disease Neuroimaging Initiative; ADNI-Lan, ADNI language composite; ADNI-VS, ADNI visuospatial composite; APOE, apolipoprotein E; LCD, least common denominator; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; MRI, magnetic resonance imaging; WS, wave specific.

TABLE 5 Association between composite scores and comparator scores with risk of conversion[a]

Language scores by phase	NC to MCI (95% CI)	MCI to AD (95% CI)
All	107 conversions /509 enrolled	320 conversions / 799 enrolled
ADNI-Lan	0.84 (0.69, 1.04)	**0.75 (0.66, 0.84)**
LCD z-score	**0.76 (0.64, 0.90)**	0.90 (0.80, 1.01)
WS z-score	**0.80 (0.69, 0.94)**	**0.82 (0.73, 0.92)**
Category Fluency – Animals	0.88 (0.71, 1.08)	**0.76 (0.68, 0.86)**
MMSE Language Items	**0.83 (0.72, 0.97)**	1.06 (0.95, 1.18)
ADAS-Cog Language Items	**0.83 (0.72, 0.95)**	0.91 (0.81, 1.01)
ADNI1 only	62 conversions / 222 enrolled	206 conversions / 373 enrolled
ADNI-Lan	0.87 (0.63, 1.19)	**0.79 (0.69, 0.90)**
LCD z-score	**0.64 (0.48, 0.87)**	0.98 (0.84, 1.14)
WS z-score	**0.71 (0.54, 0.93)**	**0.86 (0.74, 0.99)**
Category Fluency – Vegetables	0.91 (0.67, 1.23)	**0.79 (0.69, 0.90)**
ADNI1, 2, GO only	107 conversions / 508 enrolled	320 conversions / 795 enrolled
ADNI-Lan	0.84 (0.69, 1.04)	**0.75 (0.66, 0.84)**
LCD z-score	**0.76 (0.64, 0.90)**	0.90 (0.80, 1.01)
WS z-score	**0.80 (0.69, 0.94)**	**0.82 (0.73, 0.92)**
Boston Naming (Total)	0.95 (0.83, 1.09)	**0.79 (0.69, 0.90)**
ADNI2, GO, 3 only	45 conversions / 287 enrolled	114 conversions / 426 enrolled
ADNI-Lan	0.95 (0.75, 1.21)	**0.65 (0.50, 0.85)**
LCD z-score	0.97 (0.77, 1.23)	**0.81 (0.67, 0.97)**
WS z-score	1.01 (0.81, 1.24)	**0.76 (0.61, 0.96)**
MoCA – Animal Naming	0.91 (0.70, 1.19)	**0.78 (0.66, 0.92)**
MoCA – All Language Items	1.05 (0.83, 1.32)	**0.80 (0.64, 0.99)**
Visuospatial scores across all phases	NC to MCI (95% CI) 107 conversions/509 enrolled	MCI to AD (95% CI)320 conversions / 799 enrolled
ADNI-VS	0.90 (0.74, 1.09)	**0.82 (0.72, 0.92)**
Z-score	0.91 (0.75, 1.09)	**0.78 (0.68, 0.90)**
Clock score	0.98 (0.85, 1.14)	**0.80 (0.70, 0.91)**

[a]The top row in each section provides the number of people who converted and the total number of people at the beginning of that phase of the study. Numbers in subsequent rows represent the adjusted time ratio (95% confidence interval) associated with a one SD higher score for each measure. For example, the top left entry indicates that compared to people who scored at the mean for ADNI-Lan, people who scored one SD above the mean had an adjusted time ratio of 0.84 (95% confidence interval 0.69-1.04).

Abbreviations: AD, Alzheimer's disease; ADAS, Alzheimer's Disease Assessment Schedule; ADAS-Cog, Alzheimer's Disease Assessment Schedule–Cognition; ADNI, Alzheimer's Disease Neuroimaging Initiative; ADNI-Lan, ADNI language composite; ADNI-VS, ADNI visuospatial composite; APOE, apolipoprotein E; LCD, least common denominator; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; MRI, magnetic resonance imaging; NC, normal cognition; WS, wave specific.

components analysis to generate a memory-adjusted composite score for language from ADNI and used that as a phenotype for genome-wide association studies.[35] Wilhalme et al. used exploratory factor analysis and z-scores to derive composite scores in a small study of people with NC and MCI, and found that none of the composites showed change among people with NC.[36] Hassenstab et al. used a z-score approach to generate composite scores from the National Alzheimer's Coordinating Center (NACC) database to evaluate people with AD-type pathology at death, and found differences in several domains between people who did and did not develop AD-type dementia during life.[37] Wang et al. used a z-score approach to generate composite scores for people with autosomal dominant AD from the Dominantly Inherited Alzheimer Network (DIAN) study.[38] Zahodne et al. used a z-score approach to generate composite scores for people in the Washington Heights/Inwood Columbia Aging Project (WHICAP) study.[39] Dong et al. also used a z-score approach to compare profiles of people with mild and moderately severe AD.[40]

We used CFA approaches rather than exploratory factor analysis or z-score approaches. There are several reasons for this. One reason is theoretical. Borsboom addresses theories of science associated with different modeling choices and concludes modern psychometric approaches using CFA analysis are most consistent with modern

TABLE 6 Test statistics for the intercept and slope terms associated with having the AD cerebrospinal fluid (CSF) pattern[a]

Score	Intercept	Slope	*P* value
Language score			
ADNI-Lan	**−1.58**	**−5.38**	1.2×10^{-5}
LCD z-score	0.03	**−3.65**	2.8×10^{-2}
WS z-score	−0.64	**−4.62**	7.7×10^{-4}
Boston Naming Test	−0.81	**−4.49**	9.3×10^{-4}
Animal fluency	−1.89	**−3.60**	2.5×10^{-3}
Vegetable fluency	−1.28	**−4.72**	3.8×10^{-5}
MMSE language items	0.91	−1.56	2.2×10^{-1}
ADAS-Cog language items	0.27	**−3.53**	3.9×10^{-2}
Visuospatial functioning score			
ADNI-VS	−0.76	−1.94	0.847
Z-score	−0.65	**−2.30**	0.727
Clock	+0.59	−1.80	0.129

[a] Data in the left two columns of this table represent test statistics derived from beta coefficients and their standard errors for the main effect of having the AD CSF biomarker pattern on the language or visuospatial functioning score ("Intercept") and for the effect of the AD CSF biomarker pattern on the rate of change of the language or visuospatial score ("Slope"). Each score was included in a mixed effects model using z-scores of each language score. Test statistics greater than +/− 1.96 are shown in bold font. Data in the right-hand column represent *P* values from likelihood ratio tests comparing full models (with main effects for the AD signature term and its interaction with time) to reduced models (without those two effects), using a chi-squared distribution with two degrees of freedom.

Abbreviations: AD, Alzheimer's disease; ADAS-Cog, Alzheimer's Disease Assessment Schedule–Cognition; ADNI, Alzheimer's Disease Neuroimaging Initiative; ADNI-Lan, ADNI language composite; ADNI-VS, ADNI visuospatial composite; LCD, least common denominator; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; WS, wave specific.

scientific inquiry.[41] While exploratory factor analysis techniques can be useful in determining whether a single-factor model is appropriate for observed data,[42] modern psychometric theory emphasizes use of confirmatory rather than exploratory factor analysis models.[14,43] Z-score approaches are used by several investigators. Z-score approaches obscure strong assumptions, especially for a battery that evolves over time. These points are discussed in supporting information.

Modern psychometric approaches readily address different samples at different study waves. Furthermore, the influence of each indicator on the overall score is constant across all waves, regardless of which other items are included at any time point, consistent with intuition of how an indicator should function.

Modern psychometrics provides several means to assess the appropriateness of a modeling strategy. We considered single-factor and bi-factor models, ultimately determining single-factor models were appropriate for both domains. For language we further considered whether some items could be modeled using a Poisson distribution, and determined the greater flexibility of a categorical approach was needed. In cross-sectional datasets from each study wave, fit statistics demonstrated good fidelity to the data by our models. Empirically determined factor loadings provide strong evidence of variation in the strength of relationship between each indicator and the underlying construct measured by all of the indicators in common. These loadings correspond again to our intuition about better and worse indicators of language and visuospatial functioning. For example, animal and vegetable fluency items, the multi-item Boston Naming Test, and the Object Naming indicator from the ADAS-Cog have much higher loadings than any of the six dichotomous language indicators from the MMSE. These differences would be obscured if we used a z-score approach.

Beyond those theoretical and practical considerations, as we demonstrate here, ADNI-Lan and ADNI-VS perform well in a variety of prespecified analyses. Z-scores also did well, so from this single study evaluation we could not conclude that z-score composites were an invalid approach. As discussed in supporting information, z-score approaches especially break down in the context of a changing battery across waves. This issue is even more apparent when faced with trying to harmonize data across different studies with partially overlapping neuropsychological batteries.[8]

Several limitations are relevant to this study. ADNI is a convenience sample with limited ethnic heterogeneity. The ADNI battery was carefully selected but we are limited by the available data. Investigators focused on more specific aspects of language or visuospatial functioning may not wish to use composite scores.

We are making ADNI-VS and ADNI-Lan scores available from the ADNI website as we have for executive functioning (ADNI-EF) and memory (ADNI-Mem). Investigators interested in studying language and/or visuospatial functioning in ADNI should consider using these scores in their analyses.

ACKNOWLEDGMENTS

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California

Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Data management and the specific analyses reported here were supported by NIH grant R01 AG029672 (Paul K. Crane, PI).

The project was also supported by funding from NIH-National Institute of General Medicine SC3.GM122662 (Nancy S. Foldi, PI).

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. Lezak MD. *Neuropsychological Assessment*. 3 ed. NY: Oxford University Press; 1995.
2. Boada M, Lopez O, Nunez L, et al. Plasma exchange for Alzheimer's disease Management by Albumin Replacement (AMBAR) trial: study design and progress. *Alzheimers Dement (N Y)*. 2019;5:61-69.
3. Ho JK, Nation DA. Neuropsychological profiles and trajectories in preclinical Alzheimer's disease. *J Int Neuropsychol Soc*. 2018;24:693-702.
4. Mandal PK, Joshi J, Saharan S. Visuospatial perception: an emerging biomarker for Alzheimer's disease. *J Alzheimers Dis*. 2012;31(Suppl 3):S117-35.
5. McCullough KC, Bayles KA, Bouldin ED. Language performance of individuals at risk for mild cognitive impairment. *J Speech Lang Hear Res*. 2019;62:706-722.
6. Cho H, Choi JY, Lee HS, et al. Progressive tau accumulation in Alzheimer disease: 2-year follow-up study. *J Nucl Med*. 2019;60:1611-1621.
7. Balachandar R, Bharath S, John JP, et al. Resting-state functional connectivity changes associated with visuospatial cognitive deficits in patients with mild Alzheimer disease. *Dement Geriatr Cogn Disord*. 2017;43:229-236.
8. Mukherjee S, Mez J, Trittschuh EH, et al. Genetic data and cognitively defined late-onset Alzheimer's disease subgroups. *Mol Psychiatry*. 2018. Epub ahead of print.
9. Weiner MW, Aisen PS. The Alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimers Dement*. 2010;6:20. 2-11 e7.
10. Crane PK, Carle A, Gibbons LE, et al. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav*. 2012;6:502-516.
11. Crane PK, Narasimhalu K, Gibbons LE, et al. Composite scores for executive function items: demographic heterogeneity and relationships with quantitative magnetic resonance imaging. *J Int Neuropsychol Soc*. 2008;14:746-759.
12. Gibbons LE, Carle AC, Mackin RS, et al. A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging Behav*. 2012;6:517-527.
13. Embretson SE, Reise SP. *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum; 2000.
14. McDonald RP. *Test Theory: A Unified Treatment*. Mahwah, NJ: Erlbaum; 1999.
15. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*. 1993;43:2412-2414.
16. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189-198.
17. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984;34:939-944.
18. Kaplan E, Goodglass H, Weintraub S. Boston Naming Test. 2nd ed. Philadelphia: Lea & Febiger; 1983.
19. Mohs RC, Knopman D, Petersen RC, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord*. 1997;11(Suppl 2):S13-21.
20. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53:695-699.
21. Jack CR, Jr, Bernstein MA, Borowski BJ, et al. Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative. *Alzheimers Dement*. 2010;6:212-220.
22. Fridriksson J, den Ouden DB, Hillis AE, et al. Anatomy of aphasia revisited. *Brain*. 2018;141:848-862.
23. Hickok G, Poeppel D. The cortical organization of speech processing. *Nat Rev Neurosci*. 2007;8:393-402.
24. Hickok G. The functional neuroanatomy of language. *Phys Life Rev*. 2009;6:121-143.
25. Bourguignon NJ, Ohashi H, Nguyen D, Gracco VL. The neural dynamics of competition resolution for language production in the prefrontal cortex. *Hum Brain Mapp*. 2018;39:1391-1402.
26. Tranel D, Rudrauf D, Vianna EP, Damasio H. Does the Clock Drawing Test have focal neuroanatomical correlates?. *Neuropsychology*. 2008;22:553-562.
27. Samton JB, Ferrando SJ, Sanelli P, Karimi S, Raiteri V, Barnhill JW. The clock drawing test: diagnostic, functional, and neuroimaging correlates in older medically ill adults. *J Neuropsychiatry Clin Neurosci*. 2005;17:533-540.
28. Lamar M, Ajilore O, Leow A, et al. Cognitive and connectome properties detectable through individual differences in graphomotor organization. *Neuropsychologia*. 2016;85:301-309.
29. Matsuoka T, Narumoto J, Shibata K, et al. Neural correlates of performance on the different scoring systems of the clock drawing test. *Neurosci Lett*. 2011;487:421-425.
30. Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol*. 2009;65:403-413.
31. De Meyer G, Shapiro F, Vanderstichele H, et al. Diagnosis-independent Alzheimer disease biomarker signature in cognitively normal elderly people. *Arch Neurol*. 2010;67:949-956.
32. Muthen LK, Muthen BO. *Mplus User's Guiade*. 7 ed. LA: Muthen & Muthen; 1998-2012.
33. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2005.
34. Hu L-t, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999;6:1-55.
35. Deters KD, Nho K, Risacher SL, et al. Genome-wide association study of language performance in Alzheimer's disease. *Brain Lang*. 2017;172:22-29.
36. Wilhalme H, Goukasian N, De Leon F, et al. A comparison of theoretical and statistically derived indices for predicting cognitive decline. *Alzheimers Dement (Amst)*. 2017;6:171-181.
37. Hassenstab J, Monsell SE, Mock C, et al. Neuropsychological markers of cognitive decline in persons with Alzheimer disease neuropathology. *J Neuropathol Exp Neurol*. 2015;74:1086-1092.
38. Wang F, Gordon BA, Ryman DC, et al. Cerebral amyloidosis associated with cognitive decline in autosomal dominant Alzheimer disease. *Neurology*. 2015;85:790-798.

39. Zahodne LB, Manly JJ, Narkhede A, et al. Structural MRI predictors of late-life cognition differ across African Americans, Hispanics, and Whites. *Curr Alzheimer Res*. 2015;12:632-639.
40. Dong Y, Gan DZ, Tay SZ, et al. Patterns of neuropsychological impairment in Alzheimer's disease and mixed dementia. *J Neurol Sci*. 2013;333:5-8.
41. Borsboom D. *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge, UK: Cambridge University Press; 2005.
42. Lai JS, Crane PK, Cella D. Factor analysis techniques for assessing sufficient unidimensionality of cancer related fatigue. *Qual Life Res*. 2006;15:1179-1190.
43. Lord FM, Novick MR. *Statistical Theories of Mental Test Scores, With Contributions By Allan Birnbaum*. Reading, MA: Addison-Wesley; 1968.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Choi S-E, Mukherjee S, Gibbons LE, et al. Development and validation of language and visuospatial composite scores in ADNI. *Alzheimer's Dement*. 2020;6:e12072. https://doi.org/10.1002/trc2.12072