

# The Latent Structure and Test–Retest Stability of Connected Language Measures in the Wisconsin Registry for Alzheimer’s Prevention (WRAP)

Kimberly D. Mueller<sup>1,2,\*</sup>, Rebecca L. Kosciak<sup>1</sup>, Lindsay R. Clark<sup>1,4,5</sup>, Bruce P. Hermann<sup>3</sup>,  
Sterling C. Johnson<sup>1,4,5,6</sup>, Lyn S. Turkstra<sup>2,6,7</sup>

<sup>1</sup>Wisconsin Alzheimer’s Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA

<sup>2</sup>Department of Communication Sciences and Disorders, University of Wisconsin – Madison, Madison, WI, USA

<sup>3</sup>Department of Neurology, University of Wisconsin – Madison, Madison, WI, USA

<sup>4</sup>Alzheimer’s Disease Research Center, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA

<sup>5</sup>Geriatric Research Education and Clinical Center, Wm.S. Middleton Veterans Hospital, Madison, WI, USA

<sup>6</sup>Neuroscience Training Program, University of Wisconsin-Madison, Madison, WI, USA

<sup>7</sup>Department of Surgery, University of Wisconsin-Madison, Madison, WI, USA

\*Corresponding author at: Kimberly D. Mueller, Wisconsin Alzheimer’s Institute, School of Medicine and Public Health, University of Wisconsin – Madison, WARF Building, Room 950, 610 Walnut Street, Madison, WI 53726, USA. Tel.: +1-608-263-2860; Fax: +1-608-265-9122.

E-mail address: kdmueller@wisc.edu (K.D. Mueller)

Editorial Decision 31 October 2017; Accepted 7 November 2017

## Abstract

**Introduction:** While it is well known that discourse-related language functions are impaired in the dementia phase of Alzheimer’s Disease (AD), the presymptomatic temporal course of discourse dysfunction are not known earlier in the course of AD. To conduct discourse-related studies in this phase of AD, validated psychometric instruments are needed. This study investigates the latent structure, validity, and test–retest stability of discourse measures in a late-middle-aged normative group who are relatively free from sporadic AD risk factors.

**Methods:** Using a normative sample of 399 participants (mean age = 61), exploratory factor analyses (EFA) and confirmatory factor analyses (CFA) were conducted on 18 measures of connected language derived from picture descriptions. Factor invariance across sex and family history and longitudinal test–retest stability measures were calculated.

**Results:** The EFA revealed a four-factor solution, consisting of semantic, syntax, fluency, and lexical constructs. The CFA model substantiated the structure, and factors were invariant across sex and parental history of AD status. Test–retest stability measures were within acceptable ranges.

**Conclusions:** Results confirm a factor structure that is invariant across sex and parental AD history. The factor structure could be useful in similar cohorts designed to detect early language decline in investigations of preclinical or clinical AD or as outcome measures in clinical prevention trials.

*Keywords:* Alzheimer’s disease; Dementia; Mild cognitive impairment; Speech and speech disorders; Language and language disorders

## Introduction

Alzheimer’s disease (AD), the most common cause of dementia, is a neurodegenerative disease that will affect more than 5 million Americans in 2017 (Alzheimer’s Association, 2017). Unless scientists discover ways to prevent or delay the symptom onset of AD, projected costs for the disease will be over \$1 trillion by 2050 (Alzheimer’s Association, 2017). Advances in imaging and biotechnology have revealed that the neuropathological processes of AD begin years or decades before the onset of clinical symptoms; therefore, researchers have been working to discover the most sensitive measures for detecting cognitive change at the earliest point possible, i.e., in the Mild Cognitive Impairment (MCI) stage or before. To this end, there has been an increasing number of longitudinal studies of asymptomatic people at risk for AD (Beeri & Sonnen, 2016; Coats & Morris, 2005; Ellis et al., 2009; Sager, Hermann, & La Rue, 2005; Soldan et al., 2013). Studies of AD risk groups allow for

monitoring of both biomarker and cognitive profiles over time, and profiles of biomarker positive groups may be informative in making predictions about future cognitive decline (Papp et al., 2016).

The diagnosis of dementia due to probable AD requires that an individual show a gradual and progressive change in memory function (typically) and at least one other cognitive domain, causing interference with social and functional activities of daily living (Dubois et al., 2007; McKhann et al., 2011). One of the other cognitive domains frequently affected is language, with the most common being word-retrieval problems (Clarnette, Almeida, Forstl, Paton, & Martins, 2001). While typical assessment of language includes standardized testing of confrontation naming and word finding, over the last two decades researchers have been interested in moving beyond such limited language assessment to examine speech and language in contexts that more closely approximate everyday communication. Analysis of “connected language,” also referred to as “discourse,” “connected speech,” or “spontaneous speech,” is the examination of spoken language produced in a continuous sequence, as in everyday conversations. Because connected language analysis can assess multiple cognitive-communication processes, it may be more sensitive to language problems in the early stages of AD than standardized psychometric language tests. Researchers elicit connected language using a variety of methods, such as picture description tasks, autobiographical narratives, and story-telling (e.g., “Tell me the story of Cinderella.”). Connected language analysis may yield rich information about cognition, as such tasks require a complex interplay of not only retrieval from semantic and episodic memory stores, but also executive functions for search and retrieval, planning, and organizing the complexity and delivery of the spoken material to meet the characteristics of the listener (Turkstra, 2001).

Studies of connected language in AD have used a multitude of measures due to the multifaceted nature of discourse. Specifically, researchers have analyzed semantic knowledge and retrieval (Ahmed, de Jager, Haigh, & Garrard, 2013; Bird, Ralph, Patterson & Hodges, 2000; Ehrlich et al., 1997), syntactic complexity (Bates, Harris, Marchman, Wulfeck, & Kritchevsky, 1995; deLira et al., 2011), speech fluency (Ahmed, de Jager et al., 2013; Sajjadi et al., 2012), pragmatic aspects of language (e.g., efficiency and coherence; Feyereisen et al., 2006; Chapman, Highley, & Thompson, 1998), and narrative structure (Kemper & Anagnopoulos, 1989). Each of these connected language domains has been linked to several underlying cognitive processes. For example, the semantic aspect of language is purported to reflect the ability to access semantic and episodic memory systems, while complex syntax and fluency may involve executive functions for the coordination of planning, search, and retrieval. Pragmatic and social language functions may require self-monitoring and awareness of the perspective or mental state of others, or “theory of mind” (Byom & Turkstra, 2012).

Although discourse measures have been well studied in the dementia phase of AD, they are largely under-investigated in the MCI or pre-MCI literature. Also, most studies analyze multiple measures at once, which may increase the risk of Type I error. Several researchers have grouped measures based on theories linking measures to underlying linguistic constructs (e.g. Ahmed, de Jager, et al., 2013; Ahmed, Haigh, de Jager, & Garrard, 2013; Rentoumi, Raoufian, Ahmed, de Jager, & Garrard, 2014; Sajjadi, Patterson, Tomek, & Nestor, 2012; Sherratt, 2007; Wilson et al., 2010), but few studies have examined these constructs statistically. “Latent variables” refer to underlying constructs of observable variables inferred from a statistical model, usually Exploratory Factor Analysis (EFA) and/or Confirmatory Factor Analysis (CFA). By identifying a latent structure using clusters of measures for each domain rather than each individual measure, redundancy and risk of Type I error are minimized (Brown, 2014) and the characteristics of everyday speech and language become easier to interpret. This may be especially important for MCI groups, as language impairments in these individuals will be less pronounced, and adjusting for multiple comparisons in a less impaired group may place additional unnecessary constraints on analyses.

In addition to the multiple measures used to assess connected language, the within-group variability of connected language is also problematic when attempting to identify group differences (Duong, Giroux, Tardif, & Ska, 2005). Possible contributing factors to variability in connected language include age, education, and mood (Le Dorze & BÉDard, 1998; Murray, 2010). Whether or not sex contributes to variability of connected language is an area of debate in the literature. While some studies propose that there is a female advantage in language production and verbal fluency (Halpern, 2013; Kimura, 2000), a meta-analysis of language abilities and their anatomical basis in healthy subjects did not find a sex effect (Wallentin, 2009). In general, variability can partially be addressed by within-subject study designs, where an individual’s connected language is compared to his or her baseline over time. However, only a handful of studies have assessed test–retest reliability of elicited discourse, and these have been in adults with acquired aphasia (Brookshire & Nicholas, 1994; Yorkston & Beukelman, 1980). One study (Forbes-McKay & Venneri, 2005) examined test–retest reliability of picture description after 1 week in 40 cognitively normal individuals ages 23–84 years. The authors found good reliability for some measures and significant practice effects for other measures, however the small sample size and wide age range limit generalizability to other cohorts. No study has assessed the stability of discourse measures over time, therefore the variability of these measures within healthy late-middle aged and older adults is largely unknown.

We approached the present study with three aims. The first aim was to identify a latent structure of connected language measures among late-middle-aged largely asymptomatic cohort enriched with risk for AD due to parental history of the

disease. Second, we aimed to determine the test–retest stability of these measures at two study time points, approximately 2 years apart, in a group of cognitively stable individuals. Third, to address the possibility of sex differences in connected language, we aimed to compare the factor structure between males and females, to determine if there was factor invariance (i.e., the factors measured the same construct in the two groups) (Bleecker, Bolla-Wilson, Agnew, & Meyers, 1988).

## Method

### Participants

*WRAP study sample.* Our study sample was drawn from the Wisconsin Registry for Alzheimer’s Prevention (WRAP), a longitudinal study of late-middle-aged adults that is enriched for AD risk based on parental family history. WRAP participants are asymptomatic and are predominantly between the ages of 40 and 65 at the time of enrollment (Sager et al., 2005). Approximately 72% of participants have a parent with either autopsy-confirmed or probable AD as defined by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s disease and Related Disorders Association (NINCDS-ADRDA) criteria (McKhann et al., 1984). The sample consists of 71% women, with a mean age at baseline of 54 years. The longitudinal study design includes a baseline visit and serial follow-up visits every 2 years thereafter. Participants undergo a comprehensive neuropsychological test battery, including collection of a spontaneous speech sample (implemented at the third wave visit in 2012). After each study visit, participant cognitive status is determined by a consensus conference review process that allows study research-clinicians to categorize performance as “cognitively normal”, “early MCI (eMCI),” “clinical MCI,” or “dementia” (Koscik et al., 2016). All study procedures were approved by the University of Wisconsin – Madison’s Institutional Review Board, and procedures were done in accord with the ethical standards of the Human Research Protection Program of the UW-Madison and the 1975 Declaration of Helsinki.

*EFA/CFA sample.* The current study included a subset of 399 WRAP participants who were relatively risk free and had completed language samples and neuropsychological assessments. Participants were drawn from a sample that was created to establish robust internal norms for detecting cognitive change in this group (Clark et al., 2016). Specifically, the normative sample described by Clark and colleagues (2016) included WRAP participants ( $n = 476$ ) with both positive and negative parental history of AD, but excluded those at genetic risk for the disease. The criteria for inclusion in the Clark et al. normative sample were: (1) non-carrier of the *APOE*  $\epsilon 4$  allele, (2) no history of neurological or major psychiatric disorders by self-report, and (3) classified as cognitively normal at baseline and all follow-up visits using normative factor scores derived from a previous study of WRAP (Koscik et al., 2014).

Our study had the additional criteria that (1) participants had completed a language sampling task and (2) they had not been classified as “Early MCI” or “Clinical MCI” in the interim between the Clark and colleagues (2016) study and the collection of the speech sample. The sample had an age range of 43–75 ( $M = 61$  years,  $SD = 7$ ), with an average education level of 16 years ( $SD = 2.7$ ), and 67.2% were female. The race make-up of the sample was 95% non-Hispanic White, 2.3% African-American, 1.8% Spanish/Hispanic, and 0.5% Native American. Sixty-two percent of the sample had a family history of Alzheimer’s Disease.

*Test/retest reliability sample.* To be eligible for test–retest analyses, participants had to have consecutive visits with speech samples and had to have a classification of “cognitively healthy” at both study visits ( $n = 108$ ; mean( $SD$ ) = 2.4(1) years between visits).

### Discourse Collection Procedure

Participants provided informed consent to have their speech recorded while describing the “Cookie Theft” picture from the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983). Participants were instructed to “Tell me everything you see going on in this picture.” Feedback was not provided during the descriptions by the test administrators; however, if responses were unusually brief, evaluators provided the scripted prompt, “Do you see anything else going on?” Language samples had a mean( $SD$ ) duration of 50.4(.02) s, including prompts from the examiner. All responses were recorded using an Olympus VN-6200PC digital audio recorder.

### Transcriptions and Discourse Measures

Language samples were transcribed by a trained speech-language pathologist (KDM), and two trained speech-language pathology students, using Codes for Human Analysis of Transcripts (CHAT) (MacWhinney, 2000). Utterances were segmented into C-Units, an established metric for discourse analysis defined as “an independent clause and all of its modifiers” (Hughes, McGillivray, & Schmidek, 1997; Hunt, 1965). Transcriptions were coded for automatic analyses by the Computer Language Analysis (CLAN) program (MacWhinney, Fromm, Forbes, & Holland, 2011), including codes for filled and unfilled pauses, repetitions, revisions, semantic units, errors (semantic, phonological, lexical), and nonverbal behaviors (coughing, laughing, etc.). Three raters analyzed 15% of samples to calculate inter-rater reliability. Reliability was calculated using the RELY program within CLAN, and agreement was 92.4% for transcription and 98% for coding of semantic units. Semantic units, parts of speech, total utterances, grammatical relations, and other quantifiers were then automatically extracted by the CLAN program using the MOR and MEGRASP programs (MacWhinney, 2000). Variables submitted to the initial exploratory factor analysis are defined in Table 1. Variables were chosen based on findings from previous literature on this (Ahmed, de Jager, Haigh, & Garrard, 2012; Ahmed, de Jager, et al., 2013; Ahmed, Haigh, et al., 2013; Cuetos, Arango-Lasprilla, Uribe, Valencia, & Lopera, 2007; Forbes-McKay, Shanks, & Venneri, 2013; Forbes, Venneri, & Shanks, 2002; Forbes-McKay & Venneri, 2005; Fraser, Meltzer, & Rudzicz, 2015; Mueller et al., 2016).

### Samples

Analyses were performed using a split-sample approach for cross-validation of the factor structure. We used a randomization procedure within SPSS to assign 250 participants to the exploratory factor analysis (EFA) group and 149 participants to the confirmatory factor analysis (CFA) group. Table 2 shows the demographic composition of the two subsamples. The two groups were similar across all variables except for Family History of AD, where the CFA group had significantly more participants with a family history of AD than the EFA group. We address this discrepancy by testing the invariance of the factor solution between family history positive and family history negative (see CFA section).

**Table 1.** Description of 18 language measures submitted to exploratory factor analysis

Name of measure	Description
Total Semantic Units	Ahmed, Haigh and colleagues (2013), which included 23 units consisting of people, objects, actions and attitudes
Semantic Unit Idea Density (SUID)	Semantic Units divided by total number of words
Propositional Idea Density (“density”)	Based on Computerized Propositional Idea Density Rater (CPIDER3), a ratio of propositions corresponding to verbs, adjectives, adverbs, prepositions, and conjunctions to the total number of words (Brown et al., 2008; Covington & McFall, 2010; MacWhinney, 2000)
Percent Nouns	All nouns divided by total words
Percent Verbs <sup>a</sup>	All verbs divided by total words
Verb Index	Total number of verbs divided by total number of utterances
Grammatical Complexity Index	Number of grammatical relations that mark syntactic embeddings divided by the total number of grammatical relations
Total Number of Pronouns <sup>a</sup>	Count of all pronouns in language sample
Pronoun Index	Total number of pronouns divided by nouns plus pronouns
Unique words	Count of all different words, excluding repetitions of words
Total words	all spoken words in language sample
Type Token Ratio	Number of unique words divided by total words
Moving Average Type-Token Ratio	Based on a moving window, an average of type-token ratios for each successive window of 10 (Covington & McFall, 2010)
Total Mazes <sup>a</sup>	Total number of mazes (repetitions, revisions, false starts, filled pauses, unfilled pauses) per language sample
Maze Index	Total number of mazes divided by number of utterances
Filled pauses	Total number of filled pauses
Unfilled pauses	Total number of silent pauses
Inferences <sup>a</sup>	Total number of semantic units expressing an inference (e.g., “mother unconcerned about water overflowing”)

<sup>a</sup>Variables were removed due to collinearity, cross-loading, or failure to load on any factors.

**Table 2.** Comparison of demographic variables by group

Demographic variables	EFA <sup>a</sup> (N = 250)	CFA <sup>b</sup> (N = 149)	<i>t</i> <sup>*</sup>	<i>p</i> -Value <sup>**</sup>
Age (years)				
<i>M</i>	61.28	61.79	-.70	.48
<i>SD</i>	7.03	6.39		
Range				
WASI (verbal IQ) <sup>c</sup>				
<i>M</i>	112.30	112.70	-.392	.69
<i>SD</i>	9.64	9.84		
Range				
Education (years)				
<i>M</i>	16.21	16.28	-.24	.81
<i>SD</i>	2.58	2.92		
Range				
WRAT-3 Reading <sup>d</sup>				
<i>M</i>	108.07	112.70	-.06	.95
<i>SD</i>	8.89	9.84		
Range				
Female				
<i>N</i>	165	103	.414	.52
%	69.1	66.0		
AD family history				
<i>N</i>	81	70	8.44	.004
%	32.4	47.0		

<sup>a</sup>Exploratory factor analysis. <sup>b</sup>Confirmatory factor analysis. <sup>\*</sup>For categorical variables, the Pearson Chi-Square statistic was used. <sup>\*\*</sup>All *p*-values were compared to Bonferroni-adjusted  $\alpha = .0083$ . <sup>c</sup>Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999). <sup>d</sup>Wide Range Achievement Test-3 (Wilkinson, 1993).

### Exploratory Factor Analysis

We examined the latent structure of the discourse measures using a multi-step, exploratory factor analysis procedure within the EFA group ( $N = 250$ ). We extracted the factors using “principal axis factor analysis,” which is recommended when the assumption of normality is violated (Beavers et al., 2013; Fabrigar, Wegener, MacCallum, & Strahan, 1999). To allow factors to be correlated, we performed an oblique (Promax) rotation of the latent factors. We determined the optimal number of factors to retain by using a combination of Kaiser’s K1 rule (Kaiser, 1960) (Eigenvalues greater than 1), Cattell’s scree plot (Cattell, 1966), Very Simple Structure (VSS) determination, and parallel analysis tests. Discourse variables were selected in the final solution if their loadings were greater than .32 (Tabachnick & Fidell, 2001), and cross-loadings were less than .32. All models were compared using the root-mean-squared error of approximation (RMSEA, the discrepancy between the hypothesized model and the population covariance matrix), non-normed fit index (NNFI, the comparison of the hypothesized model and the null model), and the comparative fit index (CFI, measures the discrepancy between the data and the hypothesized model). We selected the final model by using the acceptability criteria defined previously (Dowling, Hermann, La Rue, & Sager, 2010): CFI > .90, NNFI > .90, RMSEA < .08 (Dowling et al., 2010). In addition, interpretability and theoretical significance were considered in final model selection. EFA models were fitted, and scree tests, Very Simple Structure (VSS) (Revelle & Rocklin, 1979), and parallel tests were performed using the *psych* package (Revelle, 2014) in R statistical software version 3.3.2 (Team, 2013).

### Confirmatory Factor Analysis

After selecting the final EFA model, we evaluated the fit of the factor structure within the CFA group using strict constraints based on principles of structural equation modeling, whereby the model is specified ahead of the analysis. All CFA analyses were performed using the R *lavaan* package (Rosseel, 2012). Method of extraction was “unweighted least squares” (ULS), due to the non-normal distribution of some of the variables’ residuals (Jöreskog & Sörbom, 1986). Multiple-group confirmatory factor analysis was then performed to test for measurement invariance between male versus female participants, and positive versus negative family history. Using the R *semtools* package, we tested for the following types of invariance: configural invariance (identical factor structure across groups), metric invariance (identical factor loadings), and scalar invariance (identical means).

### Longitudinal Test–Retest Stability of Factor Structure

In the final step, first we calculated factor scores from the EFA structure by summing the regression-based weights (Grice, 2001; Thurstone, 1931). Next, we measured the test–retest stability of the resulting factor scores within the subset with consecutive discourse samples. Using R, we computed three test–retest reliability estimates: the Pearson’s correlation coefficient ( $r$ ), two-way random effects model intraclass correlation coefficients (ICC) (Cronbach’s alpha) (Cronbach, 1951), and the coefficient of repeatability (Beckerman et al., 2001). In interpreting reliability measures, we used Fleiss’s (Fleiss, 1986) criteria for ICC, which described values from .40 to .75 as “fair to good” and values greater than .75 as “excellent”; however, due to the arbitrary nature of these inferences and the fact that measurements were taken more than 2 years apart, we also considered absolute measures of reliability that take random and systematic error into account. Specifically, we calculated the coefficient of repeatability (CR), also known as the “Smallest Real Difference” (SMD), which is calculated by multiplying the Standard Error of Measurement (SEM) by 2.77, and thus provides the value below which the absolute differences between two measurements would lie with .95 probability (Vaz, Falkmer, Passmore, Parsons, & Andreou, 2013).

### Correlation with Standardized Language Measures

We obtained Spearman’s correlation coefficients among factor scores and standard language tests (verbal IQ; Wechsler, 1999), category and letter fluency (Benton, 1969), and the Boston Naming Test (Goodglass & Kaplan, 1983).

## Results

### Exploratory Factor Analysis

We examined a variety of models and found that some variables were either too strongly correlated with one another (e.g., *total pronouns* with *unique words* ( $r = .85$ ), *total filled pauses* with *maze index* ( $r = .72$ )), or failed to load on any factor within any model iteration (e.g., *total inferences*). In the case of the highly correlated variables, we removed the variable with the weakest evidence in AD/MCI populations, and removed those variables that failed to load on any factor. Scree tests, parallel analyses, and Very Simple Structure analyses revealed that a four-factor model explained 60% of the total variance among the discourse measures (see Fig. 1). Table 3 shows factors and factor loadings. Factor correlations ranged from .07 to .43, which justified the Promax rotation. The four-factor model provided a good fit: empirical  $\chi^2 = 27.86$  ( $p < .05$ ), CFI = .98, NNFI = .90, RMSEA index = .08 (90% confidence interval = .06–.117).

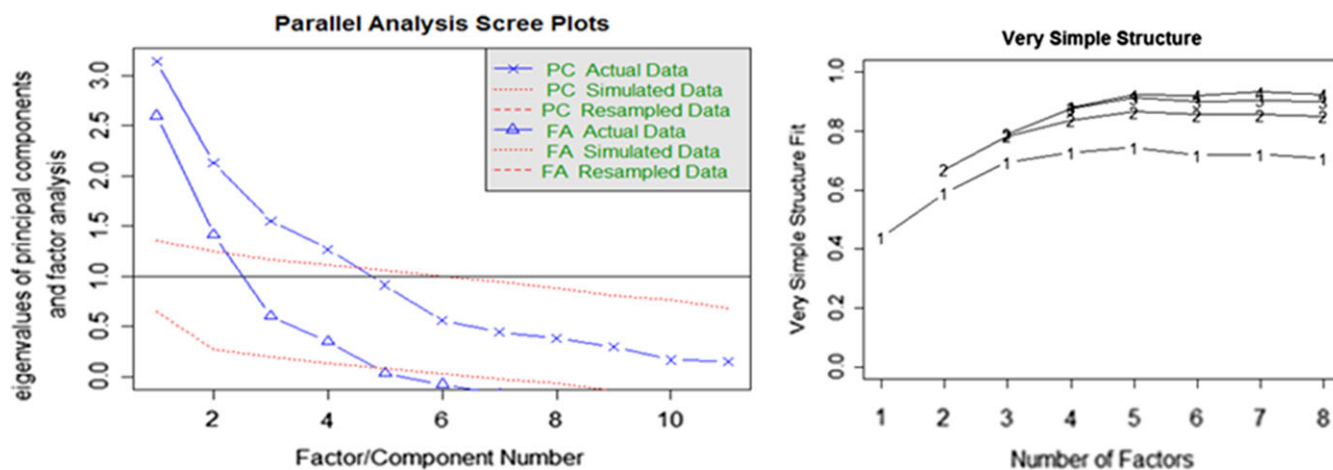
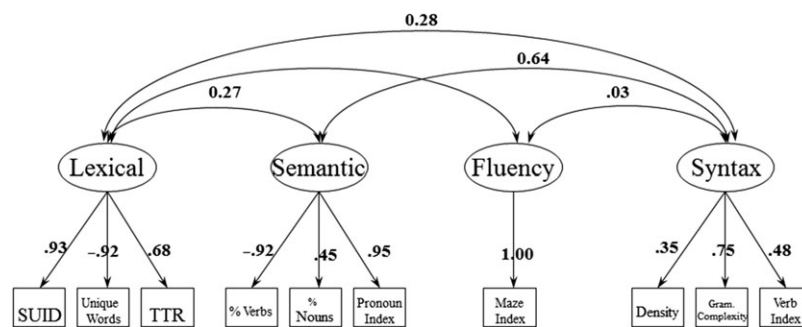


Fig. 1. Scree plot and Very Simple Structure (VSS) plot used to determine the number of factors in the exploratory factor analysis.

**Table 3.** Factor loadings for the exploratory factor analysis with Promax Rotation

Discourse variable	Lexical factor	Syntax factor	Fluency factor	Semantic factor
Density	−0.21	<b>0.34</b>	0.04	0.08
Unfilled Pauses	−0.05	−0.05	<b>0.75</b>	−0.03
% Verbs	0.25	0.31	−0.02	<b>0.38</b>
Maze Index	0.07	0.01	<b>0.76</b>	0.00
Type-Token Ratio	<b>0.75</b>	0.01	0.06	0.13
% Nouns	0.09	0.03	0.02	<b>−0.93</b>
Unique Words	<b>−0.93</b>	−0.03	−0.05	0.04
Pronoun Index	−0.09	−0.11	0.01	<b>0.91</b>
Semantic Unit Idea Density	<b>0.86</b>	0.01	−0.08	−0.20
Grammatical Complexity	−0.16	<b>0.57</b>	0.03	−0.19
Verb Index	0.10	<b>0.81</b>	−0.03	−0.01

Note: Items in bold met criteria for primary factor loadings. Principal Axis factoring was the extraction method. The four-factor model accounted for 60% of the total variance among the discourse measures.

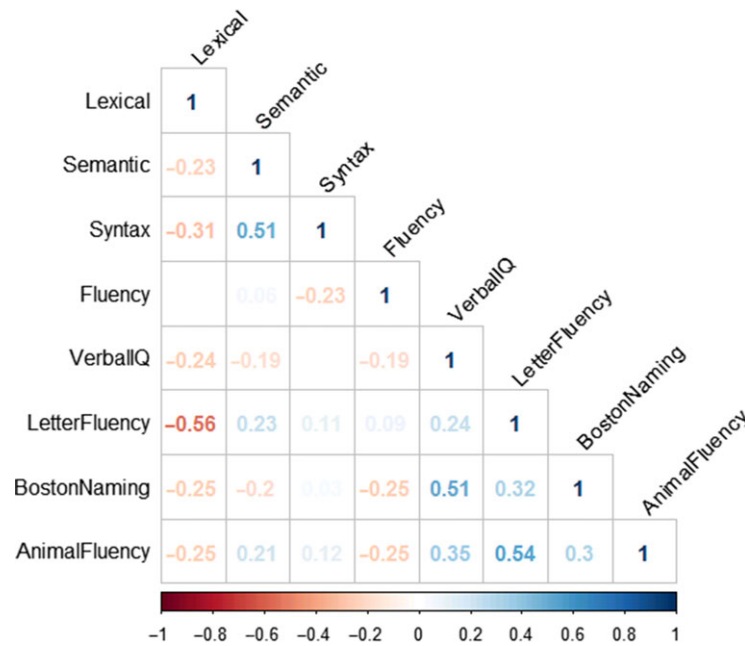
**Fig. 2.** Conceptual model of the confirmatory factor analysis using the CFA subsample ( $n = 149$ ).**Table 4.** Latent structure of confirmatory factor analysis ( $N = 149$ )

Latent variables	Lexical factor	Semantic factor	Fluency factor	Syntax factor
SUID	.93			
Unique Words	−.92			
Type Token Ratio	.68			
% Nouns		−.92		
% Verbs		.45		
Pronoun Index		.96		
Maze Index			1.0	
Verb Index				.70
Gram. Complexity				.36
Density				.46

### Confirmatory Factor Analysis

Based on the results from the EFA, a four-factor conceptual model was submitted to CFA. Initial results of the CFA indicated that variance estimates were negative for *total unfilled pauses*, which may reflect the highly skewed distribution of this variable. Therefore, the submitted model included only *Maze Index* as the variable for the Fluency Factor. Fig. 2 shows the final conceptual model.

Once the variable *total unfilled pauses* was removed, the CFA model proved to be a good fit, with the following satisfactory fit indices:  $\chi^2 = 35.63$ , CFI = 0.99, NNFI = 0.99, and RMSEA = 0.04. Table 4 shows the CFA results using the cross-validation subsample ( $N = 149$ ). The average variance extracted (AVE) for each factor, ranging from .53 to .72, was higher than the square of the correlations among the four factors (ranging from .001 to .37). This provides evidence for discriminant validity among factors (Fornell & Larcker, 1981).



**Fig. 3.** Spearman's correlation coefficients of discourse factors with Verbal IQ, Boston Naming Test, Letter Fluency, and Animal Fluency measures. Correlations reflect scores using the full sample ( $n = 399$ ).

#### Factor Invariance Across Sex and Family History

As the small CFA sample size did not allow factor invariance measures to be computed, we checked for factor invariance using the full sample ( $N = 399$ ). All fit indices yielded acceptable measures of fit for the four types of invariance: configural model (CFI = .95; RMSEA = .08), equal loadings (CFI = .94, RMSEA = .08), equal intercepts (CFI = .95, RMSEA = .07), and equal means (CFI = .95; RMSEA = .07), indicating that there were no sex-based differences in the factor estimates. Similarly, there were no differences between subgroups with positive versus negative family history, with acceptable invariance fit measures of the configural model (CFI = .95; RMSEA = .08); equal loadings (CFI = .95; RMSEA = .07), and equal means (CFI = .96; RMSEA = .06).

#### Correlations Among Language Measures

Spearman's rank correlation coefficients among standard language tests and discourse factors are presented in Fig. 3. The strongest relationship was a negative correlation between the Lexical factor and the letter fluency factor (-.56). The Semantic factor was weakly correlated with letter (.21) and category (.23) fluency, and showed a weak negative correlation with Boston Naming Test (-.20). The Fluency factor was weakly, negatively correlated with verbal IQ (-.19), Boston Naming (-.25), and Animal Naming (-.25). The syntax factor was not significantly correlated with any standard language tests.

#### Test–Retest Stability of the Discourse Factor Scores

The measures of relative and absolute reliability of the factor scores between two time points are presented in Table 5. Overall, all of the factor scores were not significantly different between the two time points, and each factor yielded a coefficient of repeatability that was within the limits of agreement as calculated by the Bland–Altman method, in which 95% of mean differences are less than two standard deviations (Bland & Altman, 1986).

**Lexical factor.** Factor stability from time 1 to time 2 was classified as “good” (Fleiss, 1986), with an alpha coefficient of .77, intraclass correlation coefficient of .76, and a mean difference score of .09 that fell within the repeatability coefficient (CR) of  $\pm .23$ . A paired sample  $t$ -test of the two time points showed that the two scores were not significantly different ( $p = .25$ ).



**Table 5.** Relative and absolute reliability indices of discourse factors

Discourse factor	Time 1		Time 2		ICC	Mean diff (bias)	SDdiff between subject	<i>t</i>	<i>p</i> -Value	95%CI LB <sup>a</sup>	95%CI UB <sup>b</sup>	Within subject variance	SEM <sup>c</sup>	CR <sup>d</sup>
	Mean ( <i>SD</i> )	Mean ( <i>SD</i> )	$\alpha$	<i>r</i>										
Lexical	−0.02 (1)	0.07 (1)	0.77	0.62	0.76	0.09	0.85	−1.2	.25	−1.8 (−2.1–(−)1.5)	1.7 (1.4–1.9)	0.37	0.08	0.23
Semantic	−0.01 (1)	−0.02 (1)	0.50	0.33	0.49	0.02	1.19	0.2	.86	−2.2 (−2.6–(−)1.9)	2.9 (.2–1.9)	0.71	0.11	0.32
Syntax	−0.05 (1)	0.01 (1)	0.49	0.32	0.49	0.07	1.19	−0.6	.56	−0.29 (−2.7–(−)1.9)	2.9 (1.9–2.3)	0.70	0.11	0.32
Fluency	0.26 (1.1)	0.06 (1)	0.69	0.52	0.68	0.20	1.03	2.0	.05	−1.8 (−2.1–(−)1.5)	2.1 (1.8–2.4)	0.55	0.10	0.28

<sup>a</sup>95% Confidence interval, lower bound. <sup>b</sup>95% confidence interval, upper bound. <sup>c</sup>Standard error of the mean. <sup>d</sup>Coefficient of repeatability, 2.77\*SEM.

*Semantic factor.* The semantic factor yielded an intraclass correlation coefficient of .49 (“fair”; Fleiss, 1986), an alpha of .50, and a repeatability coefficient of  $\pm$ .32. The mean difference score was .02, which falls within the CR interval. The paired sample *t*-test showed that the scores between the two time points were not significantly different ( $p = .86$ ).

*Syntax factor.* The syntax factor yielded an ICC of .49 (“fair”), an alpha of .49, and a CR of .32. The mean difference of .07 fell within the CR, and the paired difference test revealed no significant difference ( $p = .32$ ).

*Fluency factor.* The fluency factor had an ICC of .68 (“good”), an alpha of .69, and a CR of .28. The mean difference of .20 fell within the CR of  $\pm$  .28, and the paired difference test revealed no significant difference between the two time points ( $p = .05$ , adjusted alpha = .0125).

## Discussion

Several studies have shown that connected language analysis can contribute to early detection of Alzheimer’s Disease (Ahmed, Haigh et al., 2013) and help distinguish Alzheimer’s type dementia from other disorders (Gorno-Tempini et al., 2011; Murray, 2010; Nicholas, Obler, Albert, & Helm-Estabrooks, 1985). Because of the multifaceted nature of discourse, there are numerous measures that can be derived. We set out to determine the underlying constructs of multiple discourse measures obtained from a picture description task administered to a select subset of participants in the Wisconsin Registry for Alzheimer’s Prevention who were APOE4 negative and cognitively stable over time. Once this latent structure was defined, we tested both the structure’s stability over time in a group of individuals identified as cognitively stable, as well as the invariance of these factors between males and females, and across participants with and without a family history. Correlations between the factor scores and tests of verbal fluency, picture naming, and verbal IQ were small but statistically significant.

Results of the exploratory and confirmatory factor analyses revealed a four-factor solution that explained 64% of the total variance among the connected language measures. The latent factors were similar to those identified in the discourse of adults with AD and other dementias (Ahmed, de Jager et al., 2013; Fraser et al., 2015; Wilson et al., 2010). Fraser et al. performed an exploratory factor analysis of picture descriptions from 240 AD patients (mean age = 71.8) and 233 controls (mean age = 65.2) taken from the DementiaBank corpus (Macwhinney et al., 2011). Following a machine-learning classification of 370 linguistic features, the authors found a four-factor solution that bore some similarities to our EFA/CFA solution. Similar to our findings, Fraser et al. named a semantic factor that included pronoun ratio, percent verbs, and percent nouns; and a syntax factor that included a measure of verbs and some measures of complex syntactic forms (e.g., auxiliaries, participles). Duong and colleagues (2005) performed a cluster analysis of discourse measures with subsequent principle components analysis, and found significant heterogeneity of discourse patterns among adults with AD, due to factors such as education, age, sex, and heterogeneity in cognitive decline over time.

Our study differed from those of Duong and colleagues (2005) and Fraser and colleagues (2015) in several ways. First, we examined linguistic variables in a group of individuals that were cognitively stable across at least two time points, which allows us to establish patterns of unimpaired adults for future comparisons. Second, we performed a confirmatory factor analysis using a split-sample approach, which added validity to the latent factor structure. Third, we performed test–retest reliability analyses for these measures 2 years apart and found that the factor scores were stable within individuals over time. Finally, our study captured a latent structure in a group that was younger than most other previously reported groups (mean age = 61 years). The ability to detect stable discourse patterns in younger adults will allow us to detect subtle change earlier in the disease process, which may be critical in future longitudinal studies of preclinical AD.

One of the problems in the discourse literature is the use of a variety of measures to assess highly similar constructs, which leads to difficulty aggregating findings across samples and limits replicability and reproducibility. We defined a latent structure of the variables commonly reported in previous studies of connected language and AD, which may offer a thorough examination of connected language while at the same time reduce the problem of Type I error associated with multiple comparisons. We used a split-sample approach to EFA/CFA which validated our factor structure, and we validated the model further by testing factor invariance between males and females, and between those with a positive versus negative family history of AD. By demonstrating factor invariance in these two subgroups of interest, we provide evidence that differences in test scores across subgroups will represent true differences rather than measurement bias (Byrne, Shavelson, & Muthén, 1989; Dowling et al., 2010).

Another major finding of our study was the test–retest stability of the factor structure over a 2–3 year window among participants who were cognitively normal at both visits. Of the few studies that examine test–retest reliability, some used picture descriptions from adults with impairments (Brookshire & Nicholas, 1994; Yorkston & Beukelman, 1980), while others examined test–retest stability within short time intervals, varying between one week and one month (Forbes-McKay & Venneri, 2005; Shewan & Henderson, 1988). Very few studies have examined test–retest stability across intervals typically used in longitudinal studies. In addition, because the cookie theft picture is a simple description task, test–retest reliability within short intervals is most likely going to be subject to significant practice effects in healthy adults (Bartels, Wegrzyn, Wiedl, Ackermann, & Ehrenreich, 2010). Our study provides evidence of minimal practice effects in picture descriptions across a 2–3-year interval in late-middle-aged adults who are cognitively stable.

Our approach was to examine the stability of the factor structure in a group of individuals who had remained stable on other cognitive measures for an average of more than 2 years. This method assumes that cognitive stability on standardized measures would yield stability on connected language measures, but research to back this assumption is limited. Statistical methods like Pearson's correlations or Interclass Correlation Coefficients (ICC) would not account for the increased variability and measurement error that could occur over a longer time interval, so in addition we used the coefficient of reliability (Vaz et al., 2013). All factor scores between the two time points yielded repeatability coefficients within acceptable intervals.

### *Limitations and Future Directions*

A limitation of this study is that not all possible dimensions of connected language were submitted to the factor analyses. We chose measures most commonly associated with discourse deficits in early AD or MCI, based on a thorough review of the literature. Accordingly, the measures were related to word-retrieval difficulties, such as production of content units, ratios of nouns to pronouns, or speech disfluencies that could result from poor search and retrieval. Other measures, such as error monitoring, words per minute, or response to word-finding delays were not included and may have loaded on and strengthened one or more of the latent factors, or perhaps loaded on a new factor.

Another limitation was the relatively homogenous sample of participants. The sample was recruited primarily from a highly educated, predominantly non-Hispanic and white population, and our findings may not generalize to other study populations. Future studies can investigate the psychometric properties of connected language (e.g., factor invariance and test–retest reliability) across more samples with greater diversity (e.g., ethnically and linguistically).

An additional limitation is that the sample selection was not informed by AD biomarkers. We did not assess amyloid or tau burden biomarkers and thus cannot predict who in this normative group are exhibiting evidence of AD-related neuropathologic change. Thus generalizability of this factor structure beyond AD risk groups is not known. Future directions include using this factor structure with risk groups defined by biomarker profiles (Papp et al. 2016) to determine its effectiveness in characterizing discourse across groups.

### **Conclusion**

This study addressed an urgent priority of the Alzheimer's Association's Research Roundtable (AARR): to identify appropriately sensitive tools to measure cognitive and *functional* change in "presymptomatic" AD or very early MCI (Snyder et al., 2014). The AARR noted challenges in accomplishing this goal, including that existing tools are insensitive to very mild deficits, and may measure different constructs across cultural and ethnic groups, between men and women, and across lifestyle variables that may influence the development of cognitive decline. Snyder and colleagues (2014) recommended using performance-based assessments in addition to self-report scales, and we propose that connected language assessment can be a reliable, performance-based assessment tool. Standardized assessment of verbal fluency tasks has shown promising results in distinguishing between healthy aging and MCI (Clark et al., 2009; Mueller et al., 2015; Murphy, Rich, & Troyer, 2006;

Nutter-Upham et al., 2008; Papp et al., 2016), but performance on standardized tests has provided little information about everyday function (Kavé & Goral, 2016).

To our knowledge, this is the largest cross-sectional or longitudinal study of connected language in a risk-enriched AD cohort. It is therefore a valuable resource for documenting connected language changes in both normally aging adults and also those showing very early cognitive decline. The next step is to examine the latent factors identified here in those who are declining on other cognitive measures. By combining information from longitudinal standardized assessment and longitudinal connected language analyses, we may be able to identify functional changes associated with early MCI, and not only provide outcome measures for clinical trials but also identify handicapping and potentially treatable language signs and symptoms in adults with MCI.

### Conflict of interest

None declared.

### Acknowledgments

We would like to thank WRAP participants for their contributions to the WRAP study. Gratitude also to Kristina Fiscus and Sarah Riedeman for language transcription and to Davida Fromm and Brian MacWhinney for consultation on CHAT/CLAN. WRAP is supported by NIA grant R01AG27161, Louis Holland Sr. Research Fund. Portions of this research was supported by the Clinical and Translational Science Award (CTSA) program, through the NIH National Center for Advancing Translational Sciences (NCATS), grant UL1TR000427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

### References

- Ahmed, S., de Jager, C. A., Haigh, A. M., & Garrard, P. (2012). Logopenic aphasia in Alzheimer's disease: Clinical variant or clinical feature? *Journal of Neurology, Neurosurgery, and Psychiatry*, *83*, 1056–1062. doi:10.1136/jnnp-2012-302798.
- Ahmed, S., de Jager, C. A., Haigh, A. M., & Garrard, P. (2013). Semantic processing in connected speech at a uniformly early stage of autopsy-confirmed Alzheimer's disease. *Neuropsychology*, *27*, 79–85. doi:10.1037/a0031288.
- Ahmed, S., Haigh, A. M., de Jager, C. A., & Garrard, P. (2013). Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain*, *136*, 3727–3737. doi:10.1093/brain/awt269.
- Alzheimer's Association. (2017). Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *13*, 325–373.
- Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults: A longitudinal study on frequent repetitive cognitive testing. *BMC Neuroscience*, *11*, 118.
- Bates, E., Harris, C., Marchman, V., Wulfeck, B., & Kritchinsky, M. (1995). Production of complex syntax in normal aging and Alzheimer's disease. *Language and Cognitive Processes*, *10*, 487–539.
- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation*, *18* (6), 1–13.
- Beckerman, H., Roebroek, M. E., Lankhorst, G. J., Becher, J. G., Bezemer, P. D., & Verbeek, A. L. (2001). Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *10*, 571–578.
- Beeri, M. S., & Sonnen, J. (2016). Brain BDNF expression as a biomarker for cognitive reserve against Alzheimer disease progression. *Neurology*, *86*, 702–703. doi:10.1212/WNL.0000000000002389.
- Benton, A. (1969). Development of a multilingual aphasia battery: Progress and problems. *Journal of the Neurological Sciences*, *9*, 39–48.
- Bird, H., Ralph, M. A. L., Patterson, K., & Hodges, J. R. (2000). The rise and fall of frequency and imageability: Noun and verb production in semantic dementia. *Brain and Language*, *73*, 17–49.
- Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*, 307–310.
- Bleecker, M. L., Bolla-Wilson, K., Agnew, J., & Meyers, D. A. (1988). Age-related sex differences in verbal memory. *Journal of Clinical Psychology*, *44*, 403–411.
- Brookshire, R. H., & Nicholas, L. E. (1994). Test-retest stability of measures of connected speech in aphasia. *Clinical Aphasiology*, *22*, 19–133.
- Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Publications.
- Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods*, *40*, 540–545.
- Byom, L. J., & Turkstra, L. (2012). Effects of social cognitive demand on theory of mind in conversations of adults with traumatic brain injury. *International Journal of Language & Communication Disorders*, *47*, 310–321.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276.

- Chapman, S. B., Highley, A. P., & Thompson, J. L. (1998). Discourse in fluent aphasia and Alzheimer's disease: Linguistic and pragmatic considerations. *Journal of Neurolinguistics*, *11*, 55–78.
- Clark, L. J., Gatz, M., Zheng, L., Chen, Y. L., McKeary, C., & Mack, W. J. (2009). Longitudinal verbal fluency in normal aging, preclinical, and prevalent Alzheimer's disease. *American Journal of Alzheimer's Disease and other Dementias*, *24*, 461–468. doi:10.1177/1533317509345154.
- Clark, L. R., Kosciak, R. L., Nicholas, C. R., Okonkwo, O. C., Engelman, C. D., Bratzke, L. C., et al. (2016). Mild Cognitive impairment in late middle age in the Wisconsin Registry for Alzheimer's Prevention Study: Prevalence and characteristics using robust and standard neuropsychological normative data. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*. doi:10.1093/arclin/acw024.
- Clarnette, R. M., Almeida, O. P., Forstl, H., Paton, A., & Martins, R. N. (2001). Clinical characteristics of individuals with subjective memory loss in Western Australia: Results from a cross-sectional survey. *International Journal of Geriatric Psychiatry*, *16*, 168–174.
- Coats, M., & Morris, J. C. (2005). Antecedent biomarkers of Alzheimer's disease: The adult children study. *Journal of Geriatric Psychiatry and Neurology*, *18*, 242–244. doi:10.1177/0891988705281881.
- Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: the moving-average type–token ratio (MATTR). *Journal of Quantitative Linguistics*, *17*, 94–100.
- Croisile, B., Ska, B., Brabant, M. J., Duchene, A., Lepage, Y., Aimard, G., & Trillet, M. (1996). Comparative study of oral and written picture description in patients with Alzheimer's disease. *Brain and language*, *53* (1), 1–19.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334. doi:10.1007/bf02310555.
- Cuetos, F., Arango-Lasprilla, J. C., Uribe, C., Valencia, C., & Lopera, F. (2007). Linguistic changes in verbal expression: A preclinical marker of Alzheimer's disease. *Journal of the International Neuropsychological Society*, *13*, 433–439.
- de Lira, J. O., Ortiz, K. Z., Campanha, A. C., Bertolucci, P. H. F., & Minett, T. S. C. (2011). Microlinguistic aspects of the oral narrative in patients with Alzheimer's disease. *International Psychogeriatrics*, *23*, 404–412.
- Dowling, N. M., Hermann, B., La Rue, A., & Sager, M. A. (2010). Latent structure and factorial invariance of a neuropsychological test battery for the study of preclinical Alzheimer's disease. *Neuropsychology*, *24*, 742.
- Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., et al. (2007). Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS–ADRDA criteria. *The Lancet Neurology*, *6*, 734–746.
- Duong, A., Giroux, F., Tardif, A., & Ska, B. (2005). The heterogeneity of picture-supported narratives in Alzheimer's disease. *Brain and Language*, *93*, 173–184. doi:10.1016/j.bandl.2004.10.007.
- Ehrlich, J. S., Obler, L. K., & Clark, L. (1997). Ideational and semantic contributions to narrative production in adults with dementia of the Alzheimer's type. *Journal of Communication Disorders*, *30*, 79–99.
- Ellis, K. A., Bush, A. I., Darby, D., De Fazio, D., Foster, J., Hudson, P., et al. (2009). The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *International Psychogeriatrics/IPA*, *21*, 672–687. doi:10.1017/S1041610209009405.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272.
- Feyereisen, P. (2006). How could gesture facilitate lexical access? *Advances in Speech Language Pathology*, *8*, 128–133.
- Fleiss, J. L. (1986). Reliability of measurement. *The Design and Analysis of Clinical Experiments*. New York, NY: John Wiley & Sons, Inc.
- Forbes-McKay, K., Shanks, M. F., & Venneri, A. (2013). Profiling spontaneous speech decline in Alzheimer's disease: A longitudinal study. *Acta Neuropsychiatrica*, *25*, 320–327.
- Forbes-McKay, K. E., & Venneri, A. (2005). Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, *26*, 243–254. doi:10.1007/s10072-005-0467-9.
- Forbes, K. E., Venneri, A., & Shanks, M. F. (2002). Distinct patterns of spontaneous speech deterioration: An early predictor of Alzheimer's disease. *Brain and Cognition*, *48*, 356–361.
- Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, 382–388.
- Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2015). Linguistic features identify Alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, *49*, 407–422.
- Goodglass, H., & Kaplan, E. (1983). *Boston diagnostic aphasia examination booklet*. Philadelphia, PA: Lea & Febiger.
- Gorno-Tempini, M., Hillis, A., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. E., et al. (2011). Classification of primary progressive aphasia and its variants. *Neurology*, *76*, 1006–1014.
- Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*, 430.
- Hughes, D. L., McGillivray, L., & Schmidek, M. (1997). *Guide to narrative language: Procedures for assessment*. Eau Claire, WI: Thinking Publications.
- Halpern, D. F. (2013). *Sex differences in cognitive abilities*. New York, NY: Psychology press.
- Hunt, K. W. (1965). Grammatical Structures Written at Three Grade Levels. NCTE Research Report No. 3.
- Jöreskog, K. G., & Sörbom, D. (1986). *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*. Skokie, IL: Scientific Software.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141–151. doi:10.1177/001316446002000116.
- Kavé, G., & Goral, M. (2016). Word retrieval in picture descriptions produced by individuals with Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, *38*, 958–966.
- Kemper, S., & Anagnostopoulos, C. (1989). Language and aging. *Annual Review of Applied Linguistics*, *10*, 37–50.
- Kimura, D. (2000). *Sex and cognition*. Cambridge, MA: MIT press.
- Kosciak, R. L., Berman, S. E., Clark, L. R., Mueller, K. D., Okonkwo, O. C., Gleason, C. E., et al. (2016). Intraindividual cognitive variability in middle age predicts cognitive impairment 8–10 years later: Results from the Wisconsin Registry for Alzheimer's prevention. *Journal of the International Neuropsychological Society*, *22*, 1016–1025.

- Koscik, R. L., La Rue, A., Jonaitis, E. M., Okonkwo, O. C., Johnson, S. C., Bendlin, B. B., et al. (2014). Emergence of mild cognitive impairment in late middle-aged adults in the wisconsin registry for Alzheimer's prevention. *Dementia and Geriatric Cognitive Disorders*, 38, 16–30. doi:10.1159/000355682.
- Le Dorze, G., & BÉDard, C. (1998). Effects of age and education on the lexico-semantic content of connected speech in adults. *Journal of Communication Disorders*, 31, 53–71.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, 25, 1286–1307. doi:10.1080/02687038.2011.589893.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). 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*, 34, 939–939.
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 7, 263–269. doi:10.1016/j.jalz.2011.03.005.
- Mueller, K. D., Koscik, R. L., LaRue, A., Clark, L. R., Hermann, B., Johnson, S. C., et al. (2015). Verbal fluency and early memory decline: Results from the Wisconsin Registry for Alzheimer's Prevention. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 30, 448–457. doi:10.1093/arclin/acv030.
- Mueller, K. D., Koscik, R. L., Turkstra, L. S., Riedeman, S. K., LaRue, A., Clark, L. R., et al. (2016). Connected language in late middle-aged adults at risk for Alzheimer's disease. *Journal of Alzheimer's Disease (Preprint)*, 1–12.
- Murphy, K. J., Rich, J. B., & Troyer, A. K. (2006). Verbal fluency patterns in amnesic mild cognitive impairment are characteristic of Alzheimer's type dementia. *Journal of the International Neuropsychological Society*, 12, 570–574.
- Murray, L. L. (2010). Distinguishing clinical depression from early Alzheimer's disease in elderly people: Can narrative analysis help? *Aphasiology*, 24, 928–939.
- Nicholas, M., Obler, L. K., Albert, M. L., & Helm-Estabrooks, N. (1985). Empty speech in Alzheimer's disease and fluent aphasia. *Journal of speech and Hearing Research*, 28, 405–410.
- Nutter-Upham, K. E., Saykin, A. J., Rabin, L. A., Roth, R. M., Wishart, H. A., Pare, N., et al. (2008). Verbal fluency performance in amnesic MCI and older adults with cognitive complaints. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 23, 229–241. doi:10.1016/j.acn.2008.01.005.
- Papp, K. V., Mormino, E. C., Amariglio, R. E., Munro, C., Dagley, A., Schultz, A. P., et al. (2016). Biomarker validation of a decline in semantic processing in preclinical Alzheimer's disease. *Neuropsychology*, 30, 624.
- Racine, A. M., Koscik, R. L., Berman, S. E., Nicholas, C. R., Clark, L. R., Okonkwo, O. C., et al. (2016). Biomarker clusters are differentially associated with longitudinal cognitive decline in late midlife. *Brain*, 139, 2261–2274. doi:10.1093/brain/aww142.
- Rentoumi, V., Raoufian, L., Ahmed, S., de Jager, C. A., & Garrard, P. (2014). Features and machine learning classification of connected speech samples from patients with autopsy proven Alzheimer's disease with and without additional vascular pathology. *Journal of Alzheimer's Disease*, 42 (Suppl. 3), S3–S17.
- Revelle, W. (2014). *psych: Procedures for psychological, psychometric, and personality research* (165). Evanston, Illinois: Northwestern University.
- Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, 14, 403–414.
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48 (2), 1–36.
- Sager, M. A., Hermann, B., & La Rue, A. (2005). Middle-aged children of persons with Alzheimer's disease: APOE genotypes and cognitive function in the Wisconsin Registry for Alzheimer's Prevention. *Journal of Geriatric Psychiatry and Neurology*, 18, 245–249. doi:10.1177/0891988705281882.
- Sajjadi, S. A., Patterson, K., Tomek, M., & Nestor, P. J. (2012). Abnormalities of connected speech in semantic dementia vs Alzheimer's disease. *Aphasiology*, 26, 847–866.
- Sherratt, S. (2007). Multi-level discourse analysis: A feasible approach. *Aphasiology*, 21, 375–393.
- Shewan, C. M., & Henderson, V. L. (1988). Analysis of spontaneous language in the older normal population. *Journal of Communication Disorders*, 21, 139–154.
- Snyder, P. J., Kahle-Wroblewski, K., Brannan, S., Miller, D. S., Schindler, R. J., DeSanti, S., et al. (2014). Assessing cognition and function in Alzheimer's disease clinical trials: Do we have the right tools? *Alzheimer's & Dementia*, 10, 853–860.
- Soldan, A., Pettigrew, C., Li, S., Wang, M. C., Moghekar, A., Selnes, O. A., et al. (2013). Relationship of cognitive reserve and cerebrospinal fluid biomarkers to the emergence of clinical symptoms in preclinical Alzheimer's disease. *Neurobiology of Aging*, 34, 2827–2834. doi:10.1016/j.neurobiolaging.2013.06.017.
- Tabachnick, B., & Fidell, L. S. (2001). Cleaning up your act: Screening data prior to analysis. *Using Multivariate Statistics*, 5, 61–116.
- Team, R. C. (2013). R: A language and environment for statistical computing.
- Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, 38, 406–427.
- Turkstra, L. S. (2001). Partner effects in adolescent conversations. *Journal of Communication Disorders*, 34, 151–162. doi:http://dx.doi.org/10.1016/S0021-9924(00)00046-0.
- Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., & Andreou, P. (2013). The case for using the repeatability coefficient when calculating test–retest reliability. *PLoS One*, 8, e73990.
- Wallentin, M. (2009). Putative sex differences in verbal abilities and language cortex: A critical review. *Brain and Language*, 108, 175–183.
- Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. San Antonio, TX: Psychological Corporation.
- Wilkinson, G. S. (1993). *WRAT-3: Wide range achievement test*. San Antonio, TX: The Psychological Corporation.
- Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., et al. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain*, 133, 2069–2088. doi:10.1093/brain/awq129.
- Yorkston, K. M., & Beukelman, D. R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders*, 45, 27–36.