March 2019

Quantifying Coherence In a Transdiagnostic Sample: A Methodological Investigation of Computationally-Derived Coherence Using Ambulatory Assessment

Taylor L. Fedechko
Louisiana State University and Agricultural and Mechanical College

QUANTIFYING COHERENCE IN A TRANSDIAGNOSTIC SAMPLE: A METHODOLOGICAL INVESTIGATION OF COMPUTATIONALLY-DERIVED COHERENCE USING AMBULATORY ASSESSMENT

A Thesis

Submitted to the Graduate Faculty of
Louisiana State University
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Arts

in

The Department of Psychology

by
Taylor Leanne Fedechko
B.S., University of Maryland College Park, 2014

May 2019

Table of Contents

Abstract

Schizophrenia is a clinical diagnosis assigned to individuals that experience positive (e.g., hallucinations and delusions), negative (e.g., blunted affect), and disorganized (e.g., incoherent speech) symptoms. One particularly disabling symptom is incoherence, which is defined as the meaning-based relationship between ideas. This symptom can drastically affect an individual's quality of life by affecting areas such as social and occupational functioning. Currently, the mechanism behind this symptom is unknown and requires further study. One way to examine incoherence is to understand its level of expression in other clinical populations. With the advent of computationally-derived natural language processing (NLP), coherence can be quantified with more fine-grained detail at potentially lower levels of expression. Latent Semantic Analysis (LSA) is one promising methodology to examine coherence, but many unanswered technical questions about its application, specifically in clinical populations, still remain. Previous research has shown LSA can be used on speech from individuals with schizophrenia, who display the most extreme form of incoherence. To test LSA's utility in other clinical populations and to specify parameters for its use, the current study used LSA on a "transdiagnostic" adult sample with varying forms of psychopathology. The current study aimed to extend previous findings in a different clinical sample and examine how coherence changes over time as a function of treatment. Results suggest that more traditional measures of coherence (i.e., clinician-ratings) were moderately correlated with LSA-measured coherence ($r = 0.51$). The optimal window size to differentiate high from low clinician-rated recalls was the entire recall, rather than eight words, as was previously found. Evidence for LSA-measured coherence's dynamic nature was found as its reliability fell in the moderate range ($\alpha = 0.72$). This was close to clinician-rated coherence, with its reliability falling in the good range ($\alpha = 0.79$). Lastly, evidence supporting

the incremental validity of LSA-measured coherence was not found as it was unable to provide unique variance in a model predicting clinical outcomes. Implications for these findings include additional evidence that newer computerized methodologies are related to traditional clinical measures and may provide insight into the dynamic nature of coherence.

Introduction

Schizophrenia

Definition. Schizophrenia is a heterogeneous disorder characterized by positive/psychotic (e.g., hallucinations, delusions), negative (e.g., catatonic behavior, avolition, blunted affect, anhedonia), and disorganized symptoms (e.g., incoherent speech and bizarre behavior). In order to meet diagnostic criteria for schizophrenia, one of the following symptoms must be present for at least one month: delusions, hallucinations, or incoherent speech. Additionally, at least one more unique symptom is required to be present including: another one of the aforementioned symptoms, negative symptoms, or grossly disorganized/catatonic behavior (American Psychiatric Association, 2013). Along with these specific diagnostic criteria, more general symptoms not specific to this diagnosis include functional impairment as well as cognitive, behavioral, and affective aberrations. These general symptoms commonly occur in conjunction with diagnostic criteria and greatly impact the individual with the illness.

Impact and Importance. For most individuals, severe functional impairments are present that last a lifetime and qualify them for disability assistance (Awad & Voruganti, 2000; Jobe & Harrow, 2005). Outcomes for this disorder, though, vary dramatically depending on variables including age of onset (Hollis, 2000), the delay in treatment after the onset of psychotic symptoms (i.e., duration of untreated psychosis; Crow, MacMillan, Johnson, & Johnstone, 1986; Loebel et al., 1992), and environmental factors (Feigenson, Kusnecov, & Silverstein, 2014; McClellan, Werry, & Ham, 1993). It is possible to recover to pre-morbid functioning, the level of functioning before the onset of the illness, yet that is not common (Carpenter Jr. & Kirkpatrick, 1988; Jobe & Harrow, 2005; Roy, Mérette, & Maziade, 2001).

Treatment. Current evidence-based treatment for schizophrenia generally begins with medication, and more specifically, anti-psychotics (Hasan et al., 2013; Patel, Cherian, Gohil, & Atkinson, 2014). Along with medication, psychotherapy, such as Cognitive Behavioral Therapy for psychosis (CBTp), is also recommended. Additional components of effective treatment include social and either educational or occupational support from a multi-disciplinary team (Patel et al., 2014). Mixed evidence for the effectiveness of treatment exists and the risks of some treatments are high (Hasan et al., 2015; Whitaker, 2004). For example, anti-psychotics tend to have significant side effects, such as weight gain and drowsiness, and often are aversive enough to cause discontinuation (Hasan et al., 2013). Furthermore, some medication side effects are similar to symptoms of schizophrenia (e.g., negative symptoms), therefore, medication use can make it difficult to disentangle treatment effects from the natural course of the illness (Hasan et al., 2013).

Understanding Symptoms. Because schizophrenia is a disabling illness associated with severe symptoms, potentially negative side effects from treatment, and a high comorbidity with other impairing dysfunction (e.g., cognitive difficulties), research in this population has unique challenges. However, this illness is worth understanding so that improved treatments can be developed. One way to understand the symptoms experienced by this population is to examine them among other clinical populations, albeit at differing levels of expression. A good candidate for this type of examination is the symptom of incoherence.

Coherence in Schizophrenia

Definition. One form of disorganized behavior under the criteria for schizophrenia is incoherent speech. Incoherence was included in the original conceptualization of schizophrenia, originally termed 'loose associations' (Bleuler, 1950). This behavior has also been labelled

disorganized speech, Formal Thought Disorder (FTD), and word salad. Even though the specific term has changed over the years and is not standardized across fields and researchers, much research has been done examining the speech patterns in individuals with schizophrenia.

In linguistic terms, coherence in language production is defined as the semantic, or meaning-based, relationship between ideas (Elvevåg, Foltz, Weinberger, & Goldberg, 2007; Foltz, 2007). Coherence is influenced by both semantics (i.e., content-based meaning) and syntax (i.e., the form and structure of language) (Andreasen, 1979a; Elvevåg et al., 2007). A coherent relationship between ideas is clearly understood, both in content and structure, and is characterized as logical. On the other hand, incoherence is unclear in meaning and may even have a structure defying grammar rules; thus it is considered disorganized (Elvevåg et al., 2007). Although the coherence of speech is simple to differentiate in theory, meaning-based relationships are influenced by one's previous knowledge and are subjective in nature (Elvevåg et al., 2007).

Impact and Importance. Incoherence in individuals with schizophrenia not only decreases the effectiveness of communication, it has also been associated with indirect effects such as poor functional outcomes (Andreasen & Grove, 1986; Jobe & Harrow, 2005). Because successful communication is such an essential aspect of social and occupational functioning, incoherence engenders costs for both the individual (Bowie et al., 2010) and society (Rice, Kelman, & Miller, 1992; Wang, Demler, & Kessler, 2002).

Treatment. Generally, current treatment for schizophrenia is palliative at best (Catts & O'Toole, 2016; May, Tuma, Dixon, Thiele, & Kraude, 1981). Unfortunately, the same trend also emerges in the treatment of incoherence. More specifically, neuroleptics have been shown to alleviate some aspects of incoherence, while not affecting others (Goldberg, Dodge, Aloia, Egan,

& Weinberger, 2000; Remberk, Namysłowska, & Rybakowski, 2012). In order to advance the treatment of incoherence, its mechanism must be understood with finer detail.

Understanding Coherence. As described above, coherence reflects the quality of language, from clear and logical to confusing and disorganized. Even though individuals with schizophrenia are associated with producing some of the most incoherent samples of language, subtle forms of incoherence are commonly observed in all clinical populations (Andreasen, 1979a; Covington et al., 2005; Kuperberg, 2010). The primary tool for defining clinical psychiatric disorders, the Diagnostic and Statistical Manual of Mental Disorders-5th Edition (DSM-5), even states that mild incoherence is ubiquitous and not specific to any disorder (American Psychiatric Association, 2013). Moreover, incoherence can be found in samples of language from healthy individuals (Oltmanns, Murphy, Berenbaum, & Dunlop, 1985). Therefore, studying coherence on a continuum in the entire population may be informative with regard to multiple stages of illness. Luckily, various fields of research have been examining speech patterns in many populations.

Coherence Measurement Historical Overview

The measurement of coherence both is influenced by and reflects the different ways in which it's been conceptualized over the years in different fields. Because coherence can be measured in any sample of language, the its current study is influenced by multiple fields including, but not limited to clinical psychology research, primarily in individuals with schizophrenia; neuropsychology research, primarily in individuals with aphasia; and psycholinguistics research in a variety of populations.

Clinical Psychology. Because incoherence has been a part of clinical psychology's conceptualization of schizophrenia since its inception, varying definitions and terms have been

used to describe it. In 1979, Andreasen published formalized definitions that created a standard for the field (Andreasen, 1979a). Incoherence was broadly characterized as aberration in both the meaning and structure of speech. The most extreme and rare form of disorganized speech, known as word salad, occurs most prominently in individuals with schizophrenia, and is characterized by disorganization at the level of words.

With this new lexicon, Andreasen developed the Scale for the assessment of Thought, Language, and Communication (Andreasen, 1986; TLC), which measures incoherence through clinical ratings (i.e., a clinical impression assigned to a numerical value based on anchored descriptions). This scale quantifies a clinician's ability to identify incoherence in a patient purely based on an interaction and rating criteria (Andreasen, 1986). This clinician-rated scale includes twenty forms of incoherence, each with a rating from 0 to 4 (Andreasen, 1986).

Another clinician-rated scale that is considered a gold standard for measuring psychotic symptoms and includes rating of incoherence is the Brief Psychiatric Rating Scale – Expanded Version (BPRS-E; Lukoff, Nuechterlein, & Ventura, 1986). The BPRS-E is widely used in a variety of settings including research with clinical populations in inpatient settings (Burlingame et al., 2006), outpatient settings (Tully et al., 2017), and in the judicial system to track changes in symptoms (Hassan et al., 2011). This measure has a total of 24 items, but only some have references to speech, and more specifically the content and structure of speech. The information used to make the ratings on this scale are from self-reported information gained during a semi-structured interview, observations during the interview, and any collateral information gained about the time frame for which the interview questions ask. Items relevant to coherence include Conceptual Disorganization (i.e., a rating of the quality of speech's meaning and structure), Unusual Thought Content (i.e., a rating that includes speech content and how it reflects

preoccupation), Disorientation (i.e., a rating partially determined by an individual's confusion expressed through the content of speech), and Excitement (i.e., a rating partially determined by the presence of pressured speech) (Lukoff et al., 1986).

Two scales designed as complements to one another, measuring the positive and negative symptoms of schizophrenia, are The Scale for the Assessment of Negative Symptoms (SANS; Andreasen, 1983) and The Scale for the Assessment of Positive Symptoms (SAPS; Andreasen, 1984). The SANS is a measure that contains 25 items, of which only some rate negative symptoms expressed through speech. These items include poverty of speech (i.e., a decrease in speech output that can be severe enough to appear as incoherence) and poverty of content of speech (i.e., producing many words that lack details and ideas)(Andreasen, 1983). The SAPS is a measure that contains 34 items, of which only some rate positive symptoms expressed as incoherence. Two subscales on the SAPS contain items fully or partially rated on the presence of incoherent speech. First, the bizarre behavior subscale contains an item assessing repetitive or stereotyped behavior, which may be present in the unusual use of repetitive language. Secondly, all items under the positive formal thought disorder subscale reflect forms of incoherent speech. These include derailment (i.e., disjointed ideas next to one another that move away from the original topic), tangentiality (i.e., replying to a question or statement with an irrelevant topic or idea), word salad (i.e., incomprehensible speech with disorganization between words or sentences), illogicality (i.e., reaching a conclusion based on false premises or unclear logic), circumstantiality (i.e., long-winded speech that provides an abundance of information but eventually goes back to the original idea), pressured speech (i.e., increased rate of speech), distractible speech (i.e., speech, usually prompted by external distraction, that changes topic without completing an idea), and clanging (i.e., sounds, rather than semantics and meaning,

dictating the words being spoken)(Andreasen, 1984). With all of these well-validated measures that clinicians are being trained to use, the research on coherence in many clinical populations is growing.

Coherence in Other Clinical Populations. The study of incoherence in clinical populations has sometimes aimed to find unique speech patterns, or profiles, associated with each clinical diagnosis. Some evidence has been gathered to suggest that experts can correctly sort a patient's speech into a diagnosis based on clinical judgment alone, but more research and validation of this method is needed (Oxman, Rosenberg, Schnurr, & Tucker, 1988). Relatedly, coherence is sometimes thought to independently manifest itself within each disorder by separate mechanisms. However, little to no evidence has been gathered to support this hypothesis (Andreasen, 1979b; Andreasen & Grove, 1986). Since coherence is studied in varied fields and many populations display this symptom, it is clear that this construct is important to many disciplines.

Bipolar Disorder. The majority of people who are diagnosed with bipolar disorder experience a psychotic symptom during their lifetime (Dilsaver et al., 1997; Dunayevich & Keck Jr, 2000). The defining feature of bipolar disorder is an episode of mania or hypomania, which is a period of at least one week or four days, respectively, in which an individual experiences an expansive or euphoric mood. Along with a clinically significant increase in mood, three or more of the following symptoms must be present: requiring less sleep, having an increase in goal-directed activity, experiencing an expanded sense of self (i.e., grandiosity), being more talkative than usual, experiencing rapidly shifting or racing thoughts (i.e., flight of ideas), being easily distracted, or having excessive involvement in activities with harmful consequences (American Psychiatric Association, 2013). It is important to note that some criteria used to diagnose mania

can also qualify as a psychotic symptom, such as grandiose delusions. Another symptom that occurs in both populations is pressured speech. This type of speech is commonly present during a manic episode and is characterized by quick, loud, and/or incoherent language output (American Psychiatric Association, 2013; Karam et al., 2014; Osher & Bersudsky, 2007). This has led some individuals to theorize that incoherence is actually a symptom of bipolar disorder, not schizophrenia (Raymond Lake, 2007). Little evidence has been gathered to support this idea (Cuesta & Peralta, 2011). Nonetheless, incoherence along with other symptoms are common to both schizophrenia and bipolar disorder.

With overlapping symptoms and the high rate of these diagnoses co-occurring (i.e., comorbidity), many studies have compared the speech patterns of individuals diagnosed with bipolar disorder to individuals diagnosed with schizophrenia (Cuesta & Peralta, 1993; N M Docherty, DeRosa, & Andreasen, 1996; Harrow et al., 2000). Some differences have been documented (Harvey & Brault, 1986; M.A. et al., 1994; Perlini et al., 2012), but a clear consensus has not been reached to determine whether speech during a manic episode or during an acute episode of psychosis is meaningfully different (N M Docherty et al., 1996; Lott, Guggenbühl, Schneeberger, Pulver, & Stassen, 2002; Solovay, Shenton, & Holzman, 1987).

Depression. A depressive episode can be present in many disorders, including within the lifetime of someone with a diagnosis of schizophrenia or adjacent to a manic episode in bipolar disorder. A depressive episode is defined as either a depressed mood or loss of interest and pleasure during a two-week period accompanied by five or more of the following symptoms: a significant change in weight or appetite (i.e., increased or decreased), significant changes in sleep (i.e. hyper-or hypo-somnia), significant changes in psychomotor expression or experience (i.e., agitation or retardation), fatigue, feelings of worthlessness and/or excessive guilt, decreased

ability to concentrate and/or make decisions, and/or recurrent thoughts of death (American Psychiatric Association, 2013).

As depressive episodes occur in the context of many diagnoses, including schizophrenia, some research has been done to investigate the presence of incoherence in individuals with unipolar depression. In one study, incoherent speech was measured in both individuals with depression and individuals experiencing a manic episode (Andreasen & Pfohl, 1976). In another study, incoherence was found in both individuals with a diagnosis of depression and healthy controls (Oltmanns et al., 1985). In contrast, though, some research has used speech from individuals experiencing depression as a control group, under the assumption that their speech is coherent (Ragin & Oltmanns, 1987). Because of the little research in this area and the mixed results found, it is unclear to what extent incoherence is present in individuals experiencing a depressive episode.

Autism Spectrum Disorder. Individuals with an Autism Spectrum Disorder (ASD) diagnosis are another group sometimes studied in relation to individuals with schizophrenia. ASD is characterized by longstanding and pervasive difficulties with social relationships (i.e., communication and interactions) along with at least two restrictive and repetitive patterns of behavior or interests (e.g., repetitive movements or speech, inflexible routines or patterns of behavior, fixated interests, or abnormal reactivity to sensory input).

Despite the differences between these diagnoses, some overlap in symptoms and co-occurrence occurs. Furthermore, some researchers speculate that schizophrenia should be categorized as a neurodevelopmental disorder along with ASD (Insel, 2010; Owen, O'Donovan, Thapar, & Craddock, 2011). Some even postulate that schizophrenia may be a part of the autism spectrum (King & Lord, 2011) because of similar causes (i.e., inflammation and immune

functioning; Feigenson et al., 2014; Richard & Brahm, 2012) and symptoms (e.g., social and communication deficits; American Psychiatric Association, 2013; Pilowsky, Yirmiya, Arbelle, & Mozes, 2000). Because of this overlap, some work has been done to examine incoherence in individuals with an ASD diagnosis (Solomon, Ozonoff, Carter, & Caplan, 2008; Van Der Gaag, Caplan, Van Engeland, Loman, & Buitelaar, 2005). In both studies, individuals with a diagnosis of ASD had significant levels of incoherence, although this was admittedly not in comparison to incoherence in schizophrenia. One interesting takeaway, though, is that both studies interpret the presence of incoherence as a result of a deficit in communication skills, rather than a symptom of psychosis. Overall, not much research has been done comparing incoherence in both populations. More work could perhaps elucidate the reason behind this symptom overlap to determine if the same underlying mechanism is present.

Substance Use. Although coherence is not explicitly mentioned in any substance use disorder criteria, speech and language output aberrations are mentioned. For example, slurred speech is a symptom of substance intoxication. Also, because substances have the potential to cause symptoms that appear to be another psychiatric disorder (i.e., substance-induced disorders), use of any substance must be examined during a diagnostic evaluation. For example, it is not uncommon to have psychotic-like experiences under the influence of some substances. Therefore, substance/medication-induced psychotic disorder must be ruled out when making a diagnosis of schizophrenia (American Psychiatric Association, 2013). Luckily, though, most individuals that experience psychotic symptoms while under the influence of a substance have those experiences remit after use (Schuckit, 2006). For some individuals, though, comorbid diagnoses of a substance use disorder and a psychotic disorder does occur (Hartz et al., 2014; Kavanagh et al., 2004).

Even with an overlap in symptomology, only a few studies have analyzed the language output of individuals with a substance use disorder and none used individuals with schizophrenia as a comparison group. In one investigation, some individuals in the sample qualified for a substance use disorder and semantic content was analyzed pre- and post-treatment to understand treatment effects (Arvidsson, Sikström, & Werbart, 2011). Although coherence was not specifically examined and it is unclear how this finding applies to individuals with schizophrenia, this study served as a demonstration that linguistic analysis can be done in different clinical populations. In another experiment, Bedi and colleagues aimed to see whether linguistic analysis could differentiate the speech from individuals intoxicated by different substances (i.e., ecstasy or meth). Results showed that the linguistic analysis discriminated speech from the two substances accurately (Bedi et al., 2014). Although this was not a clinical sample (i.e., the individuals did not have a history of a substance use disorder), the conclusions drawn may still be relevant for detecting clinical states and may be able to generalize to clinical populations. If nothing else, this study demonstrated that linguistic analysis of speech is sensitive enough to differentiate subtle differences in speech.

Neuropsychology. Neuropsychology, a subfield of clinical psychology, has been examining speech patterns independently of clinical psychology's examination of incoherence in schizophrenia. Language, and by extension coherence, is an important phenomenon studied in the field of neuropsychology, particularly in individuals with aphasia. Aphasia is characterized by deficits in receptive or expressive language following brain damage (McNeil & Pratt, 2001). It is commonly assessed using standardized, performance-based measures including, most notably, an extensive battery called the Boston Diagnostic Aphasia Examination (BDAE). The entire battery determines the presence and severity of aphasia (Goodglass, Kaplan, Barresi,

Weintraub, & Segal, 2000), but tasks within the BDAE, such as the Cookie Theft picture task, can be isolated as useful and ecologically valid measures of deficits in language production as well (Giles, Patterson, & Hodges, 1996). Because individuals with aphasia demonstrate incoherence at times, measures from neuropsychology are useful tools when quantifying coherence in any population. For example, the BDAE battery, and in particular, the Cookie Theft picture task have been used in research with individuals with schizophrenia (Saykin et al., 1991). Both of these populations display some of the most extreme examples of incoherence, particularly disorganization within clauses or between words (American Psychiatric Association, 2013). Furthermore, schizophrenia was originally termed dementia praecox, or an early onset dementia, and is associated with cognitive deficits (Bleuler, 1950; Insel, 2010). With this said, the DSM-5 makes a distinction between incoherence in a neurocognitive disorder, aphasia, and schizophrenia, formal thought disorder (American Psychiatric Association, 2013). In practice, though, the difference between these two types of incoherence is not always evident. One study illustrated that two out of five psycholinguistic experts performed no different than chance, with only one expert performing highly accurately when distinguishing between the speech from an individual with aphasia or schizophrenia (Faber et al., 1983). On the other hand, with a more fine-grained analysis, a few structural and qualitative differences between speech in aphasia and schizophrenia have been documented (Gerson, Benson, & Frazier, 1977).

Psycholinguistics. Psycholinguistics is another field in which coherence is studied. One psycholinguistic measure that examines coherence similarly to clinical psychology is The Hoffman Sledge Linguistic deviance scale. This scale focuses on the syntax of language production and illustrates language output in tree formats to detect and quantify the severity of aberrations in syntax (Hoffman, Kirstein, Stopek, & Cicchetti, 1982). It visualizes language

output in a way that documents if, and when, grammatical rules are broken. Of note, this method disregards content of speech and solely focuses on form (Hoffman & Sledge, 1988; Hoffman, Stopek, & Andreasen, 1986). Another measure used in the field of psycholinguistics is the Cloze test (Salzinger, Portnoy, & Feldman, 1964; Taylor, 1953). All previously mentioned measures have primarily conceptualized incoherence as dysfunctional language output. For the Cloze test, though, incoherence is conceptualized as the amount of effort needed to be put in by the listener to understand the logic of a statement. This measure requires transcripts, particularly from the speech of an individual with schizophrenia, in which part of the discourse is removed, either a word or sentence. Once this is done, a second individual reads the discourse and has to fill in the missing information, heavily relying on context (Manschreck, Maher, & Ader, 1981). This type of test is used as a more objective way to test if language produced is consistent with context and more broadly if language is comprehensible. The Cloze test determines whether an increase in incoherence increases the effort or success of filling in that blank space (Rutter, Wishner, Kopytynska, & Button, 1978). These types of measures have been used for decades and have a literature that details not only the coherence of language produced, but also conversely measures the comprehensibility of that language.

Computational Linguistics. All previously mentioned methods of measuring coherence have led to an abundance of findings that have informed the coherence literature as it applies to individuals with schizophrenia as well as other clinical populations. The drawback to those methodologies is that they were developed prior to the recent technological revolution and therefore, they require extensive training and work to implement and interpret. With recent technological and computational advances, it is now possible to analyze large amounts of data more easily with linguistic analysis (Cabana, Valle-Lisboa, Elvevåg, & Mizraji, 2011). A newer

methodology, Latent Semantic Analysis (LSA), has a small, but rapidly growing literature documenting incoherence in individuals with psychosis (i.e., positive symptoms of schizophrenia). The methodology of LSA quantifies the relationship between words, utterances, or ideas (i.e., language components) by creating a 3D semantic space in which the component resides. Cosine is then calculated between the two language components selected, which quantifies the semantic relationship (Foltz, 2007).

LSA has been used in individuals with schizophrenia and shown promising results, indicating there is additional information gleaned from this methodology (Holshausen, Harvey, Elvevag, Foltz, & Bowie, 2014; Nicodemus et al., 2014). This has been demonstrated in individuals with a diagnosis of schizophrenia (Elvevåg, Fisher, Gurd, & Goldberg, 2002; Holshausen et al., 2014; Nicodemus et al., 2014), relatives of individuals with schizophrenia (Elvevåg, Foltz, Rosenstein, & DeLisi, 2010), and individuals at risk for schizophrenia (Bedi et al., 2015; Gupta, Hespos, Horton, & Mittal, 2017).

One advantage of LSA is decreased subjectivity compared to previously relied upon ratings (e.g., expert or clinician ratings) (Elvevåg et al., 2007). Computational linguistics, and LSA, is also data-driven and has the advantage of identifying patterns without limitations from theories (Evangelopoulos, Zhang, & Prybutok, 2012). Because coherence spans multiple disciplines, having a methodology agnostic to theories is an advantage when they could be in conflict. This lends LSA to be a useful tool in finding universalities in language production that could be overlooked if collaboration between fields did not occur. As this methodology is fairly new, validation comparing it to previous methodologies is necessary, particularly in multiple populations and contexts (Cohen, Blatter, & Patel, 2008; Elvevåg et al., 2007). With the advent of new technology, the study of coherence can continue and inform separate fields of study.

Gap in Literature

Psychometrics of Using LSA to Measure Coherence. Research applying the specific linguistic analysis LSA to clinical populations is just beginning and still presents many unknowns. The fine-grained analysis accomplished by computational linguistics and LSA captures nuances, which is why it is a potentially excellent choice for analyzing the complex phenomenon of coherence. With this level of detail, though, changes in seemingly benign and inconsequential parameters can have large changes in the interpretation of data. Because LSA has not been studied extensively in clinical populations, many of its parameters need to be determined and validated. For example, the unit of language selected (e.g., words, utterances, ideas) needs to be validated using psychometrics before being employed on a wider scale. In the past, incoherence has been studied by determining the presence of a relationship between words using performance-based measures (N M Docherty et al., 1996) or the relationship between ideas using clinical ratings (Andreasen, 1979a, 1986). Elvevåg and colleagues used LSA to quantify coherence by measuring the relationship between both words and ideas when answering conceptually different questions (Elvevåg et al., 2007). Depending on the research question, different language components may be appropriate to use in an LSA analysis, yet more evidence is required to determine how to distinguish what is appropriate, especially in clinical populations. With all of this in mind, the added benefit of introducing computational linguistics to psychological constructs also brings on added challenges and unknowns that need to be addressed systematically, particularly with the psychometrics of LSA.

One particular parameter of LSA, only quantified in one study to date in individuals with psychopathology, is window size (Elvevåg et al., 2007). A window size specifies the number of components that comprise one vector in your 3D semantic space. A vector is a quantity having

both magnitude and direction. To quantify the relationship, the cosine, or ratio between the magnitudes of the two vectors, is calculated (Foltz, 2007). In the study mentioned above, a window size of 8 words was found to be the boundary where coherence begins to break down in the speech responses of individuals with schizophrenia. This result means a window size of 8 words was able to distinguish between the high and low incoherence groups. In LSA terms, the cosine was significantly different when using a window size of 8 words to compare the two groups. Window sizes smaller than eight contain coherence and a stronger semantic relationship, while window sizes larger than eight have a weaker relationship (Elvevåg et al., 2007). The number eight falls within the range of items that can be stored in working memory, 7 ± 2 (Miller, 1956), therefore, the boundary of 8 words is postulated to be related to this capacity. Window size is an LSA parameter used to measure coherence in individuals with psychopathology that has some evidence to support its validity. Therefore, window size is an important parameter to validate.

Reliability of LSA versus Clinical Ratings Over Time. One aspect of coherence that is unknown is its relative stability over time. As many factors change on a moment by moment basis, coherence may change as well, particularly in clinical populations. One key factor to understanding coherence is to measure it repeatedly over time, within an individual. As language is highly variable between individuals, tracking it over time decreases noise and can help identify meaningful changes.

In general, clinical psychology has found evidence that coherence is trait-like in nature and stable over time (Nancy M Docherty, Cohen, Nienow, Dinzeo, & Dangelmaier, 2003; Holzman, Shenton, & Solovay, 1986). These results may be a product of the operationalizations and methodologies, such as Likert scales, that capture broader constructs than those measured

with computational linguistics. This assertion, though, remains to be tested. Luckily, recent technological advances have led to the development of methods that have increased sensitivity to detect and quantify subtleties in coherence to determine if they do in fact exist. Many unknowns, specifically in regards to coherence over time, still remain, but the introduction of computational linguistics into the study of clinical populations may aide in understanding the phenomenon of coherence.

Validity of LSA versus Clinical Ratings in the Context of Treatment. With current evidence, it is also unclear how coherence is affected across the duration of a disorder, including in relation to treatment. Some evidence suggests incoherence may be present at multiple stages of schizophrenia and serve as an endophenotype (i.e., a behavior with a clear genetic connection) (Gupta et al., 2017), which may imply that incoherence has trait-like qualities (Harvey, Earle-Boyer, & Wielgus, 1984). Other evidence suggests that aspects of this phenomenon are most prominent in acute episodes (Gouzoulis-Mayfrank et al., 2003; Marengo & Harrow, 1985). While this finding may reflect that incoherence is exclusively present in acute episodes, it could also mean that incoherence is simply the most severe in acute episodes, and therefore, the most easily detected with current measures. This is one unknown that could be addressed with measures that can quantify language production in smaller time scales and with more fine-grained detail, such as LSA. Furthermore, other studies suggest components of coherence fluctuate and reflect state changes (Andreasen & Grove, 1986; Barch & Berenbaum, 1997), including contextual factors like changes in state negative affect and cognitive load (Nancy M Docherty & Hebert, 1997; Le, Najolia, Minor, & Cohen, 2017). Further inquiry into the qualitative differences in coherence over time will be important to understand the complexities

that likely exist in this phenomenon. Moreover, if coherence is found to be dynamic, it could be a valid marker of treatment progress and have clinical significance.

Aims of the Current Study

This study aimed to extend the findings of previous work (Elvevåg et al., 2007) by examining coherence in a transdiagnostic sample (e.g., within an inpatient substance use facility). Ambulatory assessment was employed to collect data from individuals using an electronic device (i.e., an iPod). Data was collected frequently and involved active participation from each individual. Moreover, because ambulatory assessment is relatively unobtrusive and can be used outside of a lab setting, it is a potentially more ecologically valid measurement of moment by moment fluctuations in individuals compared to some laboratory procedures. In conjunction with computational linguistics, this method of data collection has the potential to add incremental validity by more easily and effectively gathering data that can be analyzed at a more fine-grained level.

Aim 1 Psychometrics of Using LSA to Measure Coherence in a Clinical Sample. To get a better understanding of this relatively new methodology (i.e., LSA), the first aim consisted of exploring how LSA-measured coherence related to traditional clinical-ratings. As both clinician-rated and LSA-measured coherence are supposed to measure the same theoretical construct, the study aimed to test the hypothesis that clinician-rated and LSA-measured coherence were related. Moreover, another intention was to replicate the prior findings that a window size of 8 words is optimal to detect the boundary between coherence and incoherence (Elvevåg et al., 2007). In accordance with this previous finding, the anticipated ideal window size to detect the boundary between coherence and incoherence is 7 ± 2, or between 5 and 9 words.

Aim 2 Reliability of LSA versus Clinical Ratings Over Time. In line with some previous findings indicating that language output in individuals with schizophrenia is dynamic (Le et al., 2017; Melinder & Barch, 2003), the second aim consisted of quantifying coherence over time. Using the previously found ideal window size to measure coherence in our sample (Aim 1), the change in coherence was analyzed at the group level to determine whether coherence increased each subsequent app session. Because change, specifically improvement, is anticipated with treatment adherence, the study aimed to test whether coherence increased as the individual progress through treatment. Furthermore, because LSA-measured coherence is more fine-grained, it is anticipated that LSA-measured coherence increased more than clinician-rated coherence over time.

Aim 3 Validity of LSA versus Clinical Ratings in the Context of Treatment. In order to determine whether computational linguistics, and LSA in particular, offered incremental validity (i.e., added additional information above and beyond existing clinical measures), the validity of LSA versus clinical-ratings in the context of treatment was studied. The study aimed to test the hypothesis that LSA added incremental validity when predicting treatment outcome, and provided information above and beyond what clinical ratings would provide. It was expected that coherence would naturally improve as a function of treatment, so this was a potential criterion for validating and understanding LSA and how it changes over time (Goldberg et al., 2000; Remberk et al., 2012).

Methods

Participants

Participants ($N = 99$) were males involved in a 28-day inpatient substance use program. As outlined below, some limitations existed in the dataset (e.g., silent recordings, unavailable demographic information), which resulted in the exclusion of some participants and the final sample included 84 individual participants (See Figure 1 for an outline of participant exclusions). The decision to exclude individuals who identified their ethnicity as "other" was made because their data created a group with skewed data. All other excluded data was based on recall-level information that is described below.

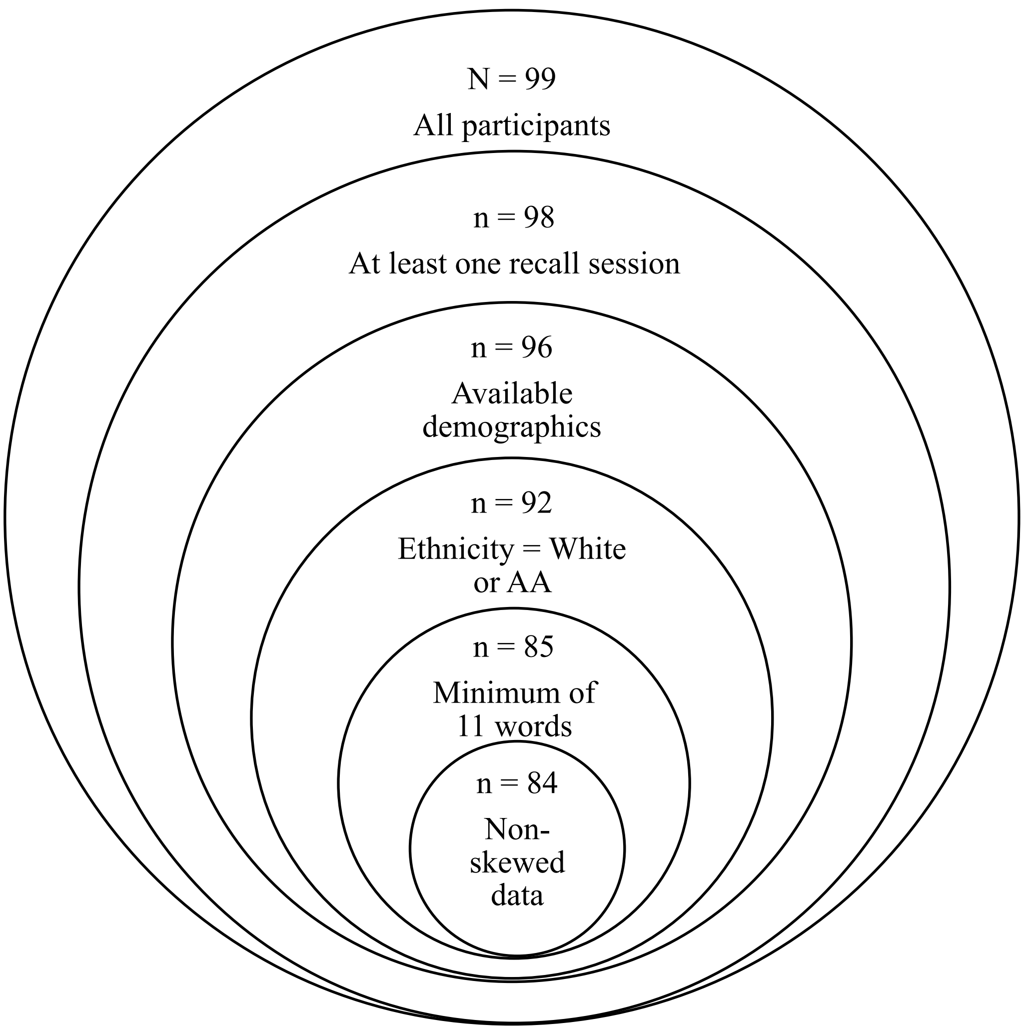


Figure 1. Participant Exclusions.

Participants were a mix of court-ordered and voluntary admissions who could leave the treatment program at will. The program, for some, occurred after detoxing in a medical setting. At the time of testing, all individuals were clinically stable and receiving treatment under the supervision of a multi-disciplinary team including doctors, nurses, social workers, and trained staff. Every individual met criteria for a substance use disorder and most individuals had comorbid diagnoses (See Table 1 for demographic and clinical descriptive statistics of the final sample used for Aims 1 and 3). As described below, some participants were excluded for Aim 2, which required multiple sessions to examine change over time. All diagnoses were made by the multi-disciplinary team and accessed through medical records. No exclusion criteria based on clinical information were utilized in this study. The test battery included many measures for research purposes, some of which were utilized and are described in the sections that follow. Data was collected between June and September 2017.

Medical Records. Demographic and clinical information was obtained from a medical record search to which each participant consented. Demographic information used in this study included age, ethnicity, and number of years of education. Clinical information obtained from records included treatment start date, treatment end date, end event (i.e., graduation or premature departure), treatment duration, primary psychiatric diagnosis, and primary substance use diagnosis. Demographic information was not available for a subset of participants (n = 2) who were excluded from all analyses.

Table 1. Descriptive statistics per participant (i.e., demographic and clinical) for the used sample (n = 84).

Demographic Info	Range	Mean (M)	Standard Deviation (S.D.)
Age (years)	19.00 – 69.00	36.96	11.13
Education (years)	3.00 – 16.00	11.38	2.32
Ethnicity	54.76% (n = 46) African-American, 45.24% (n = 38) Caucasian		
Diagnoses			
Primary Psychiatric Disorder	41.67% (n = 35) (Substance Use Disorder Only) 29.76% (n = 25) (Depression) 14.29% (n = 12) (Anxiety) 5.95% (n = 5) (Bipolar Disorder) 3.57% (n = 3) (Schizophrenia Spectrum) 2.38% (n = 2) (Trauma) 1.19% (n = 1) (Sleep) 1.19% (n = 1) (Adjustment)		
Primary Substance Use Disorder	28.57% (n = 24) (Alcohol) 26.19% (n = 22) (Opioid) 19.05% (n = 16) (Cocaine) 10.71% (n = 9) (Amphetamine) 9.52% (n = 8) (Cannabis) 3.57% (n =3) (None/Unknown) 2.38% (n = 2) (Sedative, hypnotic, or anxiolytic)		
Clinical Outcomes	Range	Mean (M)	Standard Deviation (S.D.)
Treatment Length (days)	4.00 – 34.00	25.43	6.35
Total App Sessions	1.00 – 13.00	8.14	3.01
Outcome Event	84.52% (n = 71) graduate treatment program		

Measures

Ambulatory Assessment App. Ambulatory assessment was used to monitor coherence during the period in which the individual was in the inpatient facility. Participants completed the delta Mental Status Examination (dMSE) application on an iPod provided by the study team. The dMSE app was developed to provide subjective, objective, and psychiatric information about patients regarding psychiatric risk. The research team visited the program with the iPods one to four times per week. During each session of the app, participants completed self-report assessments of in the moment experiences, digitally recorded speaking tasks, and measures

known to capture subtle fluctuations in cognitive performance. Each participant was compensated $2.50 for each session. In all, 747 recalls were successfully collected before any exclusions were applied and only 4 sessions were declined because of disinterest in completing the app. In total, only 7 recalls were excluded because of missing information (e.g., demographic information) and 30 more recalls were excluded from individuals who identified their ethnicity as "other." Some individuals did not respond to the task, which resulted in a silent recording. These recalls were excluded, along with recalls containing 10 words or less, when an 11-word minimum was applied (described in more detail below). Ultimately, the final number of recalls analyzed for this study was 512 (See Figure 2 for an illustration of these exclusions). The number of sessions completing the app task by an individual ranged from 1 to 13 (M = 8.14, SD = 3.01).

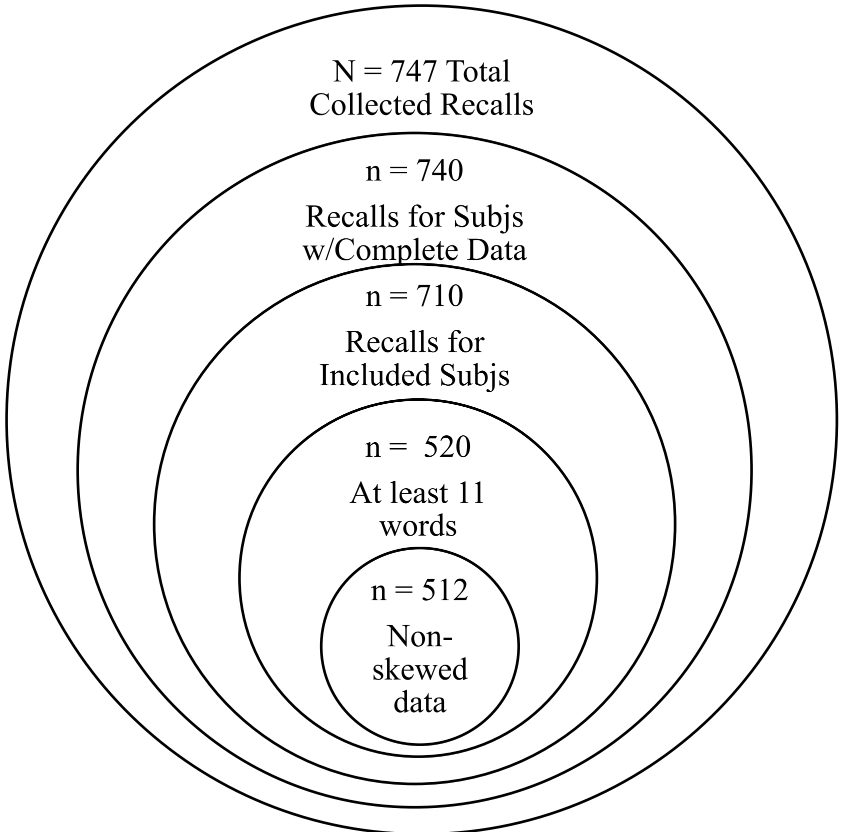


Figure 2. Exclusion of recall-level data.

Speaking Task. Participants were instructed to listen to and remember one of twenty-four stories introduced with a short title. Each participant was instructed to recall the story after hearing it for the first time (immediate recall; See Appendix 1). The length and content of these stories were inspired by the logical memory subtest on the Wechsler Memory Scale – 4th Edition (WMS-IV; Wechsler, 2009).

Clinical Ratings. Coherence of each recall was rated by clinicians on a scale from 0 to 6. This scale was inspired by the four gold standard clinical ratings scales used to quantify incoherence (i.e., BPRS, SANS, SAPS, and TLC). Response options included 0 (i.e., unintelligible/silent), 1 (i.e., structureless), 2 (i.e., limited), 3 (i.e., partial), 4 (i.e., skeletal), 5 (i.e., organized), 6 (i.e., complete). See Table 2 below for the complete rating scale used. The rating team included five individuals, including four graduate students and one clinical

psychologist, trained on the protocol. Each recall was first transcribed by a team of undergraduates and then rated using that transcript.

Table 2. Clinician-rated coherence scale.

Narrative Organization (Accurate Content & In Correct Order)		
Themes contributing to narrative arcs: Plans, causes, situations, actions, motivations which link the order of narrative concepts.		
Value	Label	Description
0	Silent or unintelligible	
1	Structureless	Any major or minor concepts/themes are not presented in any narrative arc; disorganized or incoherent.
2	Limited	Some limited concept sequences correspond to their order in the original, but largest arc is missing.
3	Partial	Several concept sequences are there, but the logic of the larger narrative is not fully expressed.
4	Skeletal	Most major concepts are in original order, and the logic of the narrative can be found and tied to story themes.
5	Organized	The narrative arc and causal logic is fully expressed with most major concepts included and tied to story themes.
6	Complete	All major concepts and themes are in logical order supporting the narrative with close fidelity to the original.

Latent Semantic Analysis. Natural language processing (NLP) using computerized linguistic analysis was used to create a coherence score for each response to the speaking task. Transcripts of the recall were proofread by the author of this study to correct for errors (i.e., spelling) and format the text appropriately for the software and the semantic space used for the analysis (e.g., eliminated all apostrophes). LSA-measured coherence is a cosine which represents the relationship between ideas present in the original story and ideas in each individual recall. Practically speaking, the words present were each separate vectors that were summed to create one vector for each component analyzed (i.e., story prompt, entire recall, or window). The vectors have context and meaning because they are placed in a predetermined 3D semantic space. In this way, LSA is more than simply analyzing the overlap in word content. The vectors created from sums of word vectors represent ideas. The relationship between ideas, or summed

vectors, is quantified by finding the cosine between the two vectors. Two separate cosines were calculated for each recall: one compared the relationship between the entire recall and the entire story prompt and the other compared a moving window, as previously described in Elvevag et al. (2007), to the entire story prompt. For the moving window cosines, a value was calculated between the entire story and the first n words of the immediate recall response, or the first window. Then the window moved over one word in the recall and the cosine was calculated for the relationship between the entire story and those n words, or the next window. This continued until the end of the recall was reached. Ultimately, all cosines assigned to each window were averaged over the entire recall to create one value. The 11 word minimum for recalls was applied to ensure that all analyses included the exact same recalls and at least two windows were available to create an average cosine. Cosines were calculated using the package LSAfun (Günther, Dudschig, & Kaup, 2015) in R and were quantified in the “TASA” semantic space. See Figure 3 and 4 below for a visual depiction of these processes.

Story Prompt		Entire Recall
a group of friends went to the beach for the day in the morning they went swimming and then lie on their towels sunbathing around noon a fisherman walked by carrying a very large fish i guess that means it is lunchtime said one of the friends they opened up their picnic basket and had sandwiches apples and cool sodas the sky became cloudy and they decided to leave for home before it started raining	1 cosine calculated comparing entire story prompt to entire recall →	a group of friends went to the beach for a picnic a fisherman passed by they brought towels lunch

Figure 3. A visual depiction of cosine calculation between entire story prompt and entire recall. One cosine is calculated per recall.

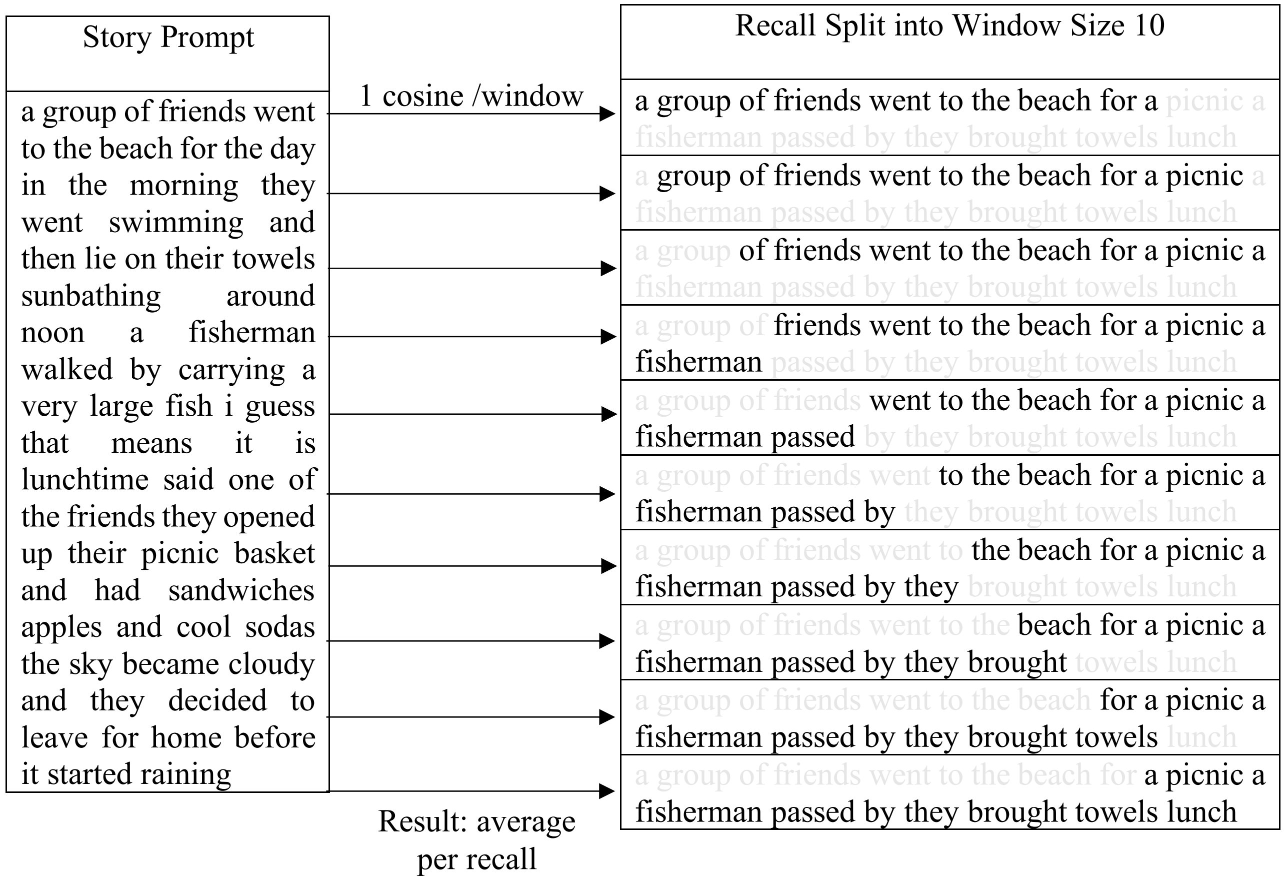


Figure 4. A visual depiction of cosine calculation between using a moving window. Multiple cosines are calculated comparing the entire story prompt to each subsequent window. Multiple cosines are calculated per recall and then averaged to create one cosine per recall.

Statistical Analyses

Preliminary Analysis Data Distribution. Normality of the distribution for each continuous variable was examined. Skew scores between -2.5 and +2.5 and kurtosis scores between -2.0 and +7.5 were used as the guideline because those are the ranges commonly found in published psychological science studies (Blanca, Arnau, López-Montiel, Bono, & Bendayan, 2013).

Preliminary Analyses Potential Covariates. Preliminary analyses meant to explore trait-like covariates and group differences in the data were conducted. The variables examined as potential covariates included demographic differences (i.e., age, ethnicity, education), clinical differences (i.e., primary psychiatric diagnosis, primary substance use diagnosis, treatment duration, treatment outcome, number of sessions), and a linguistic difference (i.e., average number of words in recall per participant). These analyses, using ANOVAs, evaluated whether both clinician-rated (i.e., the average clinician-rated coherence score per participant) and LSA-measured (i.e., average cosine comparison of the entire recall to the story prompt per participant) coherence differed on the aforementioned potential covariates. Because treatment duration was also examined as an outcome, covariates affecting this variable were examined. No significant differences based on these covariates were hypothesized.

Aim 1 Psychometrics of Using LSA to Measure Coherence. For each window size (i.e., 2 to 10 words and entire recall), the cosine between the prompt story and the entire immediate recall was calculated using the LSA protocol described above. These cosines were correlated with the clinician-rating to determine, first and foremost, if the two measures of coherence (i.e., clinician-rated and LSA-measured) were related. To address the second part of Aim 1, the median score of all recalls was first calculated. Then the recalls were split into a high

and low coherence group based on the clinician-rated score. Any clinician-rated recall below 3.0 was labelled as having “low” coherence and anything above 3.0 was labelled as having “high” coherence. Using a Receiver Operator Characteristic (ROC) curve, the area under the curve (AUC) of each window size (i.e., 2 to 10 words and the entire recall) was calculated. On the ROC curve, the window size with the best balance of sensitivity and specificity (i.e., largest AUC) represents the ideal window size for measuring coherence in this sample.

Aim 2 Reliability of LSA versus Clinical Ratings Over Time. Intraclass correlations (ICC) of coherence measures for each sequential session for that participant were calculated for every individual with at least two app sessions. As a result of the criteria for this aim, 7 participants were excluded from this analysis, compared to Aim 1. This means Aim 2 was conducted on 77 participants. This analysis, comparing session to each coherence measure, was done for both LSA-measured and clinician-rated coherence. See Table 3 below for the interpretation guidelines used (Koo & Li, 2016). Because not all participants had the same amount of sessions, only session numbers 1 through 10 were used in this analysis.

Table 3. Intra-Class Correlation Value interpretations.

Intra-Class Correlation Value	Description
< 0.50	Poor
0.50 – 0.75	Moderate
0.75 – 0.90	Good
> 0.90	Excellent

Aim 3 Validity of LSA versus Clinical Ratings in the Context of Treatment. Multi-level modelling was used to determine if LSA-measured coherence added incremental validity and predictive power to a treatment outcome. The first step in the model included the grouping variable of participant to account for the fact that recalls were nested within an individual. Next,

the amount of words in each recall was added. Then, clinician-rated coherence was entered. The final step included adding LSA-measured coherence into the model.

Power Analysis. A power analysis was previously conducted using easyROC (Goksuluk, Korkmaz, Zararsiz, & Karaagaoglu, 2016) to determine the minimum number of sessions required to detect the expected effect with power (1- β) of 0.80, two-tailed tests, and α = 0.05. Elvevåg and colleagues found a large effect size (*d*=0.98) when correlating clinician-rated with LSA rated coherence (Elvevåg et al., 2007). A priori power analyses suggest a sample size with a minimum of 28 recalls are needed to achieve the minimum effect size.

Results

Preliminary Analysis Data Distribution. LSA-measured coherence had a skew of -2.01. To address this issue, standard scores of the cosine of the entire recall compared to the story prompt were calculated and all sessions outside 3.5 standard deviations were excluded. This resulted in an additional 8 recalls being excluded from all analyses. After these sessions were removed, no recall-level variable had a skew outside of -2.5 to +2.5. No variables had a kurtosis score outside of -2.0 and +7.5.

Preliminary Analyses Potential Covariates. Both coherence measures were examined for differences based on the aforementioned potential covariates (i.e., demographic, clinical, and linguistic variables). Coherence measures significantly differed based on ethnicity (LSA-measured ($F(1,82) = 4.72$, $p = 0.03$) and clinician-rated ($F(1,82) = 25.38$, $p < 0.001$)). In both cases, Caucasian participants had higher average coherence scores compared to African-Americans. Because of this difference, the effect of ethnicity on treatment outcome (i.e., treatment duration) was examined to determine its relevance to all aims. A difference was not found in treatment duration based on ethnicity ($F(1,82) = 0.372$, $p = 0.54$) and therefore ethnicity was not included in analyses involving treatment outcome (Aims 2 and 3). Additionally, significant differences in both cosine (whole recall to story prompt) ($F(1,82) = 16.00$, $p < 0.001$, Figure 5) and clinician-rated coherence ($F(1,82) = 57.61$, $p < 0.001$, Figure 6) existed for the number of words in each recall averaged for each participant. In both cases, as the average number of words spoken per participant increased, both average clinician-rated and average LSA-measured coherence per participant also increased.

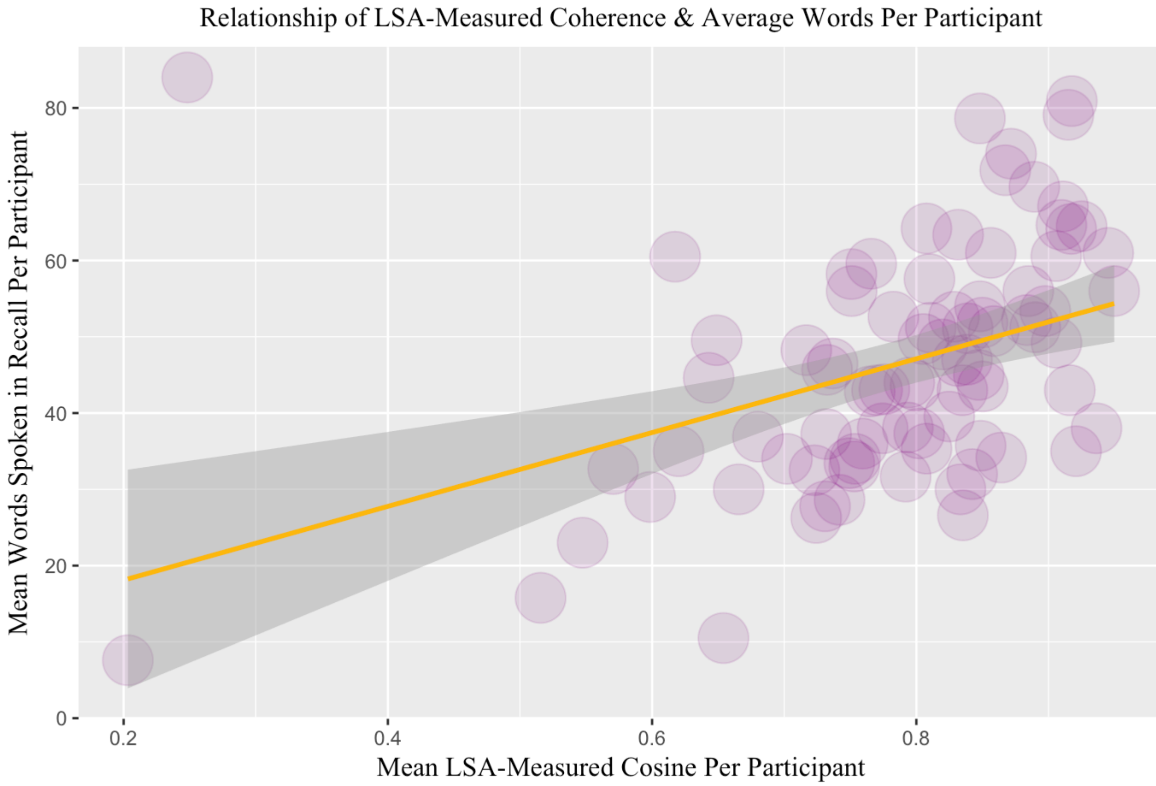


Figure 5. Data at the level of each participant (n = 84) was examined to determine the relationship between average LSA-measured coherence per participant (cosine between the entire recall and the whole story averaged within each individual) and average number of words (number of words per recall averaged by participant).

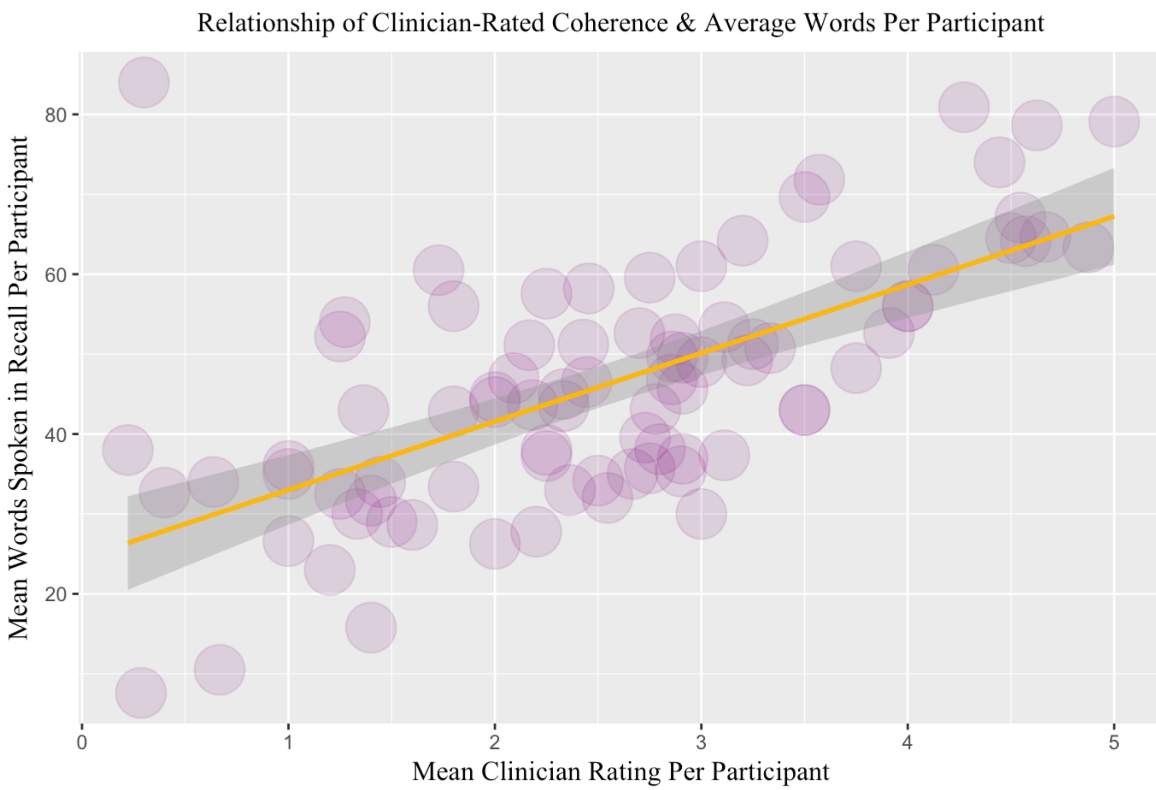


Figure 6. Data at the level of each participant (n = 84) was examined to determine the relationship between average clinician-rated coherence per participant (organization rated by each rater averaged by participant) and number of words (number of words per recall averaged by participant).

All other variables examined (i.e., age, education, primary psychiatric diagnosis, primary substance use diagnosis, total number of sessions, end event, and treatment duration), did not cause significant differences in either average measure of coherence per participant. Additionally, no significant differences in treatment duration were found based on the covariates examined (i.e., age, education, diagnoses, average words per participant).

Aim 1 Psychometrics of Using LSA to Measure Coherence. Recall-level data (n = 512) had a large correlation between clinician-rated and LSA-measured coherence (r = 0.51, p < 0.001, Figure 7). The optimal window size was the entire recall (AUC = 0.7305, Figure 8). (See Appendix 2 for all other ROC Curves). Descriptive statistics for the linguistic variables are in Table 4.

Table 4. Descriptive statistics per participant (i.e., linguistic variables) for the used sample (n = 84) and per recall (n = 512).

Linguistic Variables	Range	Mean (M)	Standard Deviation (S.D.)
Average Per Participant (n = 84)			
Clinician-Rated Coherence (Organization)	0.22 – 5.00	2.56	1.16
LSA Coherence (entire recall)	0.20 – 0.95	0.78	0.13
Words per Recall	7.60 – 84.00	46.38	15.52
Per Recall (n = 512)			
Clinician-Rated Coherence (Organization)	0.00 – 6.00	3.32	1.40
LSA Coherence (entire recall)	0.38 – 0.98	0.82	0.12
Words in Recall	11.00 – 109.00	49.93	20.39

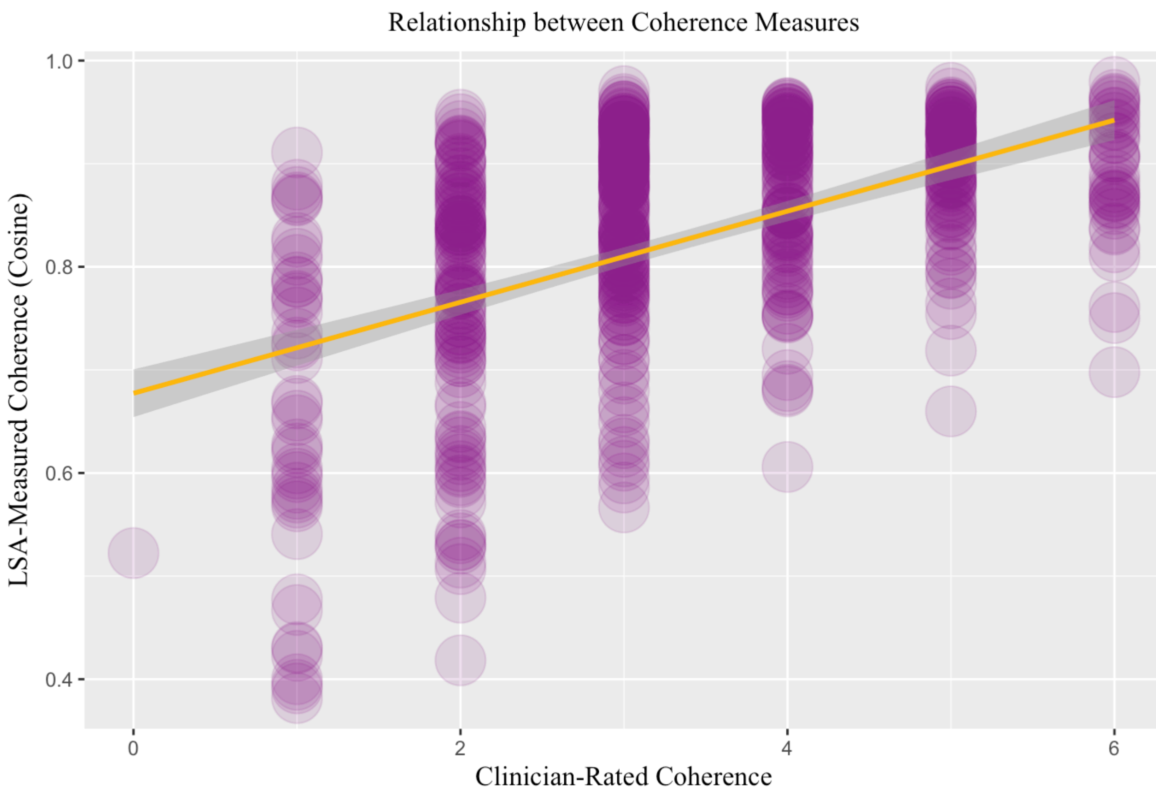


Figure 7. Recall-level data (n = 512) demonstrating the relationship between clinician-rated coherence (organization rated by each rater averaged by participant) and LSA-measured coherence (cosine between the entire recall and the whole story).

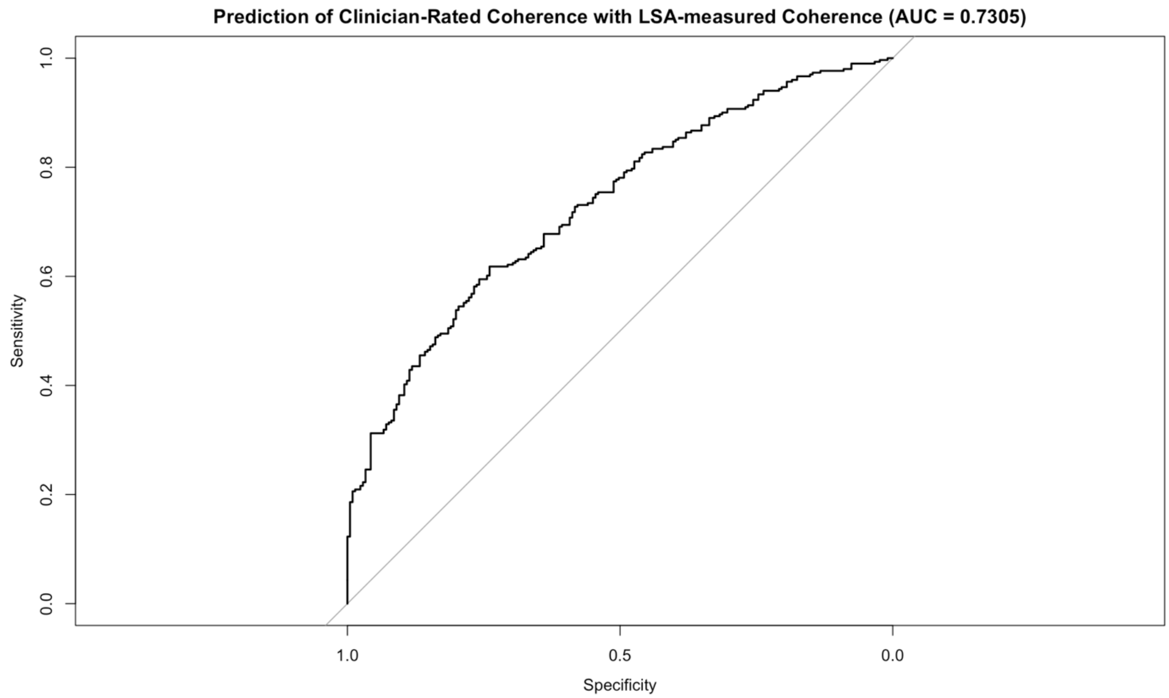


Figure 8. Receiver Operating Characteristic (ROC) Curve of recall-level data (n = 512) demonstrating the ability of LSA-measured coherence (cosine between the entire recall and the whole story) to predict the dichotomized clinician-rated coherence (high versus low rating).

Aim 2 Reliability of LSA versus Clinical Ratings Over Time. The Intraclass correlation between session and LSA-measured coherence (α =0.72, Figure 9) fell in the moderate range, and was slightly lower as compared to clinician-rated coherence (α =0.79, Figure 10), which fell in the good range. See Table 5 below for all ICC values computed using the differing window sizes. This table demonstrates the instability of moving windows as a measure of coherence.

Table 5. A table with all ICC values as a function of window size.

Window Size	ICC Value (α)
2 words	-0.03
3 words	-0.24
4 words	-0.25
5 words	-0.28
6 words	-0.23
7 words	-0.13
8 words	-0.20
9 words	-0.48
10 words	-0.25

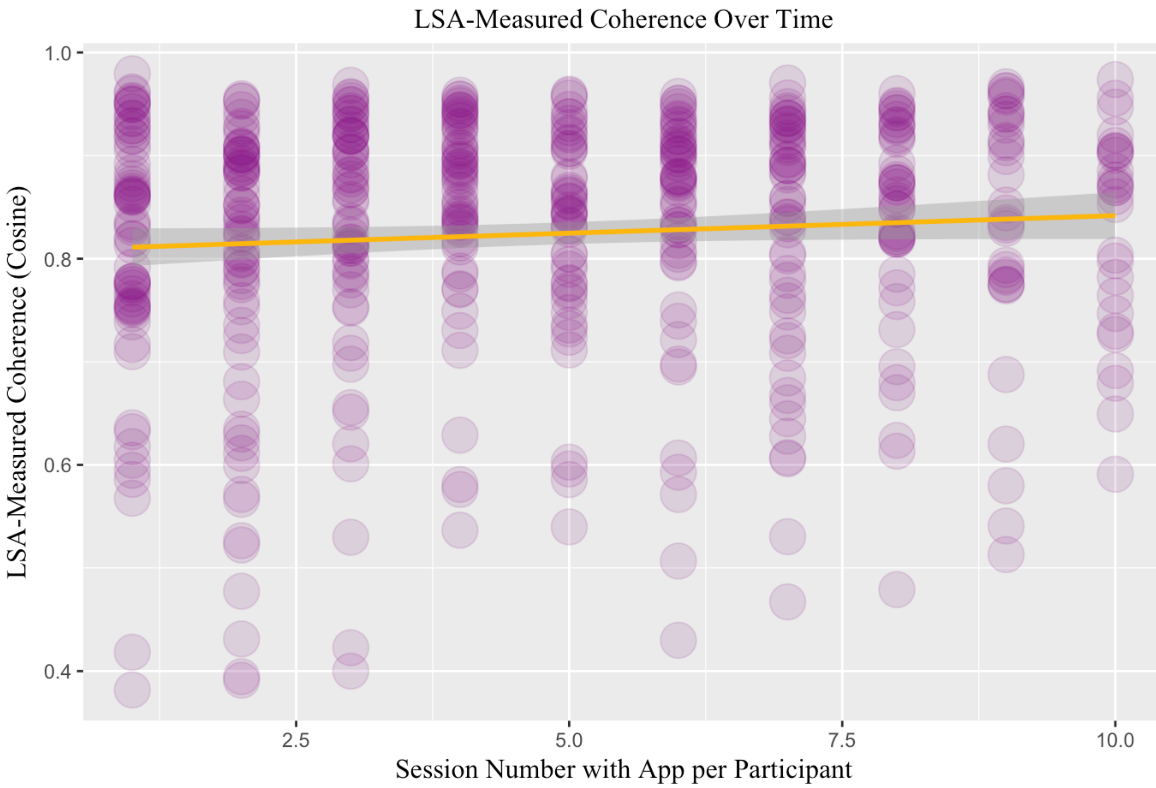


Figure 9. Participant-level data (n = 77) demonstrating LSA-measured coherence (cosine between the entire recall and the whole story) as it changes over subsequent session numbers.

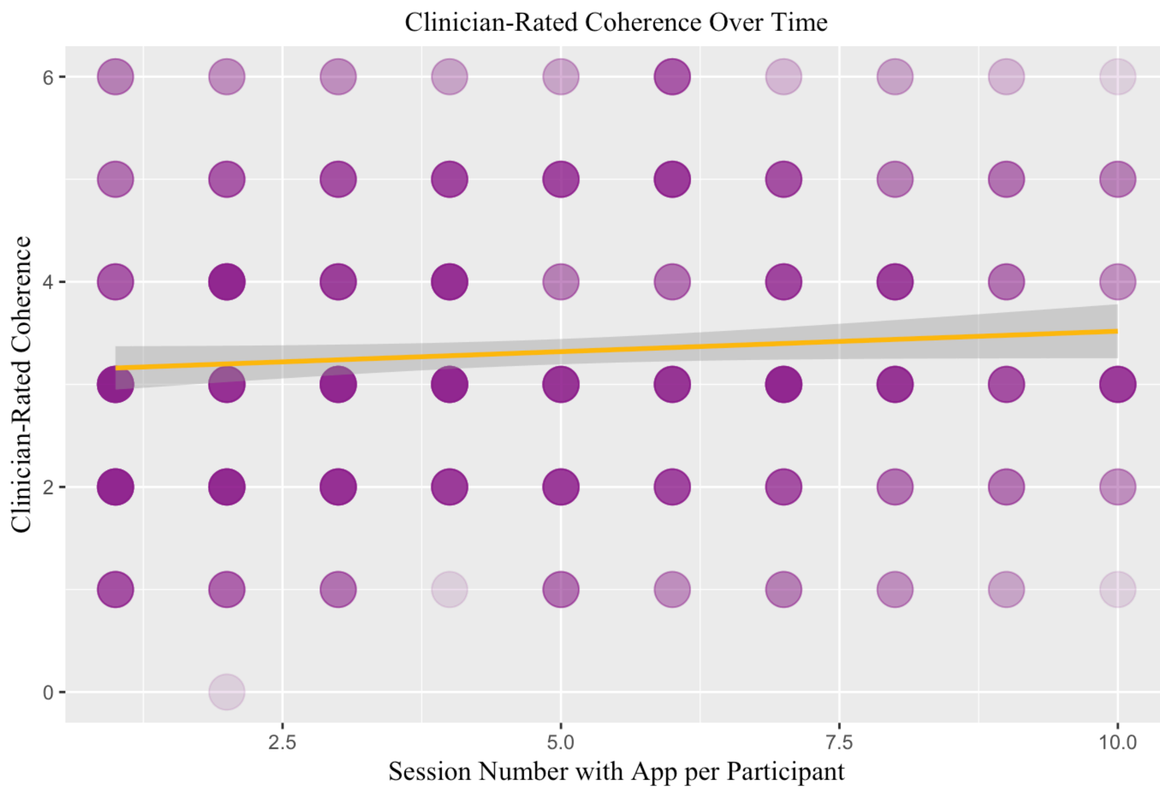


Figure 10. Participant-level data (n = 77) demonstrating clinician-rated coherence (organization rated by each rater averaged by participant) as it changes over subsequent session numbers.

Aim 3 Validity of LSA versus Clinical Ratings in the Context of Treatment. Both measures of coherence (clinician-rated and LSA-measured) did not add significant variance to the model when predicting days left in treatment (See Table 6). Although number of words per recall added significant information to the model ($\chi^2 = 9.02$, $p = 0.002$), no added variable had a significant weight.

Table 6. Multi-level modelling results with centered variables.

Model: Outcome Days Left in Treatment	Df	AIC	BIC	logLik	Deviance	Chisq	Df	Pr (>Chisq)
Participant (Intercept)	3	1452.9	1465.6	-723.44	1446.9			
Number of Words in Recall	4	1445.9	1462.8	-718.94	1437.9	9.02	1	0.003**
Clinician-Rated Coherence	5	1445.8	1467.0	-717.90	1435.8	2.07	1	0.15
LSA-Measured Coherence	6	1446.8	1472.2	-717.41	1434.8	0.98	1	0.32

** = $p < 0.01$

Random Effects	Variance	Std. Dev.	
Participant	0.09	0.29	
Residual	0.90	0.95	
Fixed Effects	Estimate (β)	Std. Error	t Value
Intercept	-0.03	0.05	-0.61
Number of Words	-0.08	0.06	-1.34
Clinician-Rating Coherence	-0.06	0.06	-1.06
LSA-Measured Coherence	-0.05	0.05	-0.99

Discussion

Overall, this study replicated some previous findings (e.g., found evidence of a moderately sized relationship between clinician-rated and LSA-measured coherence) and extended the literature on coherence with unique results. The eight-word moving window was not replicated in this sample, but the entire recall was found to have convergent validity when using LSA-measured to predict high versus low clinician-rated coherence. Also, both measures of coherence showed moderate to good stability, but also demonstrated change over time. The incremental validity of coherence was examined and found not to add any significant information to treatment outcome.

The first result of the study is that a modest relationship existed between the more traditional measure of coherence, clinician-rated, and the newer methodology, LSA-measured coherence. This replicates previous findings (Elvevåg et al., 2007) and suggests that both measures likely quantify a similar phenomenon, namely coherence. This relationship is not perfect, though. One explanation for divergence between the measures could be due to the methodological differences. Clinician-rated coherence could be affected by the specific raters who all have a slightly different semantic space in mind when determining logicality. On the other hand, LSA-measured coherence is a computerized analysis that used the same semantic space for all recalls. Another difference is the scope of the rating, or the scale of information. All clinical ratings reflected the organization of the entire recall as compared to each story prompt. Therefore, clinician-rated coherence is likely a global impression calculated by an implicit algorithm in each rater's head. LSA-measured coherence operates on words, which inform ideas and logicality. This modest relationship means the information added from each measure did not

perfectly overlap and perhaps the two coherence measures can be used in concert to gather different information.

The previous finding that the best LSA window for quantifying incoherence is eight words was not replicated in this sample. The entire speech sample, or recall, was found to be the best differentiator between high and low clinician-rated recalls. This could be interpreted as a result of the methodological differences between the previous study (Elvevåg et al., 2007) and the current one. In the previous study, for the one experiment that used a moving window analysis, participants responded to an abstract question (e.g., “What is democracy?), described a well-known story (e.g. “Tell me the story of Cinderella.”), or gave general steps to an activity (e.g., “What would someone have to do if they wanted to wash their hair?”). Responses likely relied primarily on long-term memory. In contrast, the current study involved the participant hearing a story and then repeating it back immediately, which likely relied primarily on working memory. The discrepant findings could be accounted for, at least partially, by the different kinds of memory used. Also, in the previous study, the participant chose the words and details rather than simply repeating back as many details of the twenty-four short stories as possible. Creating a narrative versus repeating back a story are two qualitatively different tasks. Further evidence to understand the difference in task structure and cognitive demands would need to be gathered to understand if and how the task had a differential effect on the speech samples. Another major difference that could account for the lack of replication of the moving window size of eight is that different clinical populations were used. Cognitive limitations are known to exist in individuals with schizophrenia and are not necessarily always associated with other disorders like anxiety, depression, and substance use, which were more common in this sample. Even with

discrepant findings, the overall results of the first aim show LSA's promise as a way to quantify coherence in clinical populations.

To examine if and how coherence changed over time, both LSA-measured and clinician-rated coherence were studied. Evidence for the stability of coherence was found, although clinician-rated coherence was slightly more stable as compared to LSA-measured coherence. Perhaps studying fluctuations within individuals over longer periods of time, such as months, could be helpful. It is possible that meaningful changes in coherence, as related to treatment, occur on the order of months or years, rather than days and weeks. Even though coherence is a dynamic phenomenon, meaningful increases in coherence as a result of psychiatric treatment may occur on a larger time scale than studied. For example, some psychiatric medications are thought to take weeks to reach full effectiveness. Because medication was a part of most of the participants' treatment plan, perhaps we needed to study coherence over a few months to capture meaningful change caused by medication. Also, measuring coherence fluctuations in response to environmental factors, such as cognitive or affective load, could be helpful to understand the mechanism behind this phenomenon. For example, previous research (Le et al., 2017) has shown that speech is influenced by changes in the cognitive demands placed on an individual. Because cognitive demands were not accounted for in this study, measuring that variable may account for some trends in the data.

Lastly, evidence supporting LSA's incremental validity for informing treatment outcome was not found. In other words, both measures of coherence were not predictive of treatment outcomes above and beyond other information present. However, this finding may not mean that LSA does not provide information about all treatment outcomes and clinical issues. This study chose a specific outcome to study (i.e., day in treatment), but other variables, such as changes in

symptom ratings (e.g., depression or anxiety), may be predicted by fluctuations in coherence. These variables could be tested in future research. Furthermore, the modest relationship between LSA-measured and clinician-rated coherence could partially explain LSA's lack of incremental validity in this study. Because these two methods are related, perhaps they provide similar information. This likely cannot account for all results, though, as the β weight for LSA-measured coherence was not significant in the model tested.

Limitations

One limitation of this study included the specific sample used. This sample consisted of all males who were residing in Louisiana receiving treatment at an inpatient substance use program. Because ethnicity impacted LSA-measured and clinician-rated coherence, differences in demographic information likely affect coherence. Different cultures and genders may show varying results. Although the one measure of treatment outcome used in this study did not differ based on ethnicity, language-based measures need to consider ethnicity more closely and further study of this phenomenon is warranted. Another important piece of information to consider about this sample's demographics is that Louisiana has a very unique culture and a rich linguistic history. Because of this, the language used in this sample was unique. To account for this, repeated measures from the same individual were grouped together in analyses when possible. This removed some intra-individual variability that likely occurs between this sample and others. Because of this specific sample, the generalizability of these results is limited.

Another limitation related to the sample used was that the information regarding whether or not an individual voluntarily participated in treatment, or was court-ordered to attend, was not available. This likely affected variables such as treatment outcome. Future studies using this type of sample would benefit from examining the difference between court-ordered and voluntary

treatment outcomes. These two groups likely have different motivations which may be reflective in their speech.

Lastly, a major limitation of any study using automated analysis of speech is that the process of transcribing involves decisions that could influence the outcome. Along the way, decisions were made about how to transcribe certain words, like slang. It is not always possible to capture exactly what an individual states in a transcription and some information is lost. That being said, a theoretically based protocol was followed, which enhanced the integrity of the data as much as possible.

Future Directions

Future directions include examining how coherence changes over the course of a recall in more detail. The previous research finding that eight words is the best unit to measure changes in coherence strongly suggests that coherence changes over small time courses, like sentences or conversations. With the current study, the relationship between the story prompt and the entire recall was used as it had the best predictive validity. This technique reduces data and potentially loses information about the temporal dynamics of coherence. Even using the moving window technique smoothed the data, potentially losing valuable information. It is unclear whether this is true, though, and future studies could examine this.

Another future direction is to examine the role of ethnicity. This was the only demographic variable for which group differences existed in both clinician-rated and LSA-measured coherence. Clearly ethnicity is a factor influencing speech and coherence. Luckily, a large literature exists examining the differences in language expression based on culture and ethnicity. Using the findings from other fields may be an important next step in the study of coherence. Because LSA-measured and clinician-rated coherence are meant to be agnostic to

ethnicity and culture, the fact that ethnicity is a covariate suggests future research is necessary to either decrease these effects or account for them more explicitly in ratings.

Lastly, because LSA semantic spaces are built primarily with text-based documents, rather than documents borne from speech, issues may emerge when using LSA on speech samples. The semantic space used in this study, TASA, was created primarily based on text documents (e.g., books, newspaper articles) from Touchstone Applied Science Associates, Inc. These heavily edited and curated text-based documents are different from spoken word, which is instantaneous and contains errors and slang not usually present in formal writing. Dialects like Black English Vernacular or the unique speech in Louisiana are likely not well represented in the documents used to create that semantic space. It will be important to utilize different semantic spaces as they may yield different, yet meaningful, results. Another option would be to create a semantic space just from the data gathered in this study and determine the coherence scores based on anomalies within this dataset. This may be advantageous with large datasets to begin to understand normative patterns of speech and begin to explicate when that becomes incoherence.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. *American Psychiatric Publishing*. Arlington, VA. https://doi.org/10.1176/appi.books.9780890425596.744053

Andreasen, N. C. (1979a). Thought, language, and communication disorders: I. Clinical assessment, definition of terms, and evaluation of their reliability. *Archives of General Psychiatry*, *36*(12), 1315–1321. https://doi.org/10.1001/archpsyc.1979.01780120045006

Andreasen, N. C. (1979b). Thought, language, and communication disorders: II. Diagnostic significance. *Archives of General Psychiatry*, *36*(12), 1325–1330. https://doi.org/10.1001/archpsyc.1979.01780120055007

Andreasen, N. C. (1983). *The Scale for the Assessment of Negative Symptoms (SANS)*. Iowa City, Iowa: The University of Iowa.

Andreasen, N. C. (1984). *Scale for the Assessment of Positive Symptoms (SAPS)*. Iowa City: University of Iowa.

Andreasen, N. C. (1986). Scale for the assessment of thought, language, and communication (TLC). *Schizophrenia Bulletin*, *12*(3), 473–482. https://doi.org/10.1093/schbul/12.3.473

Andreasen, N. C., & Grove, W. M. (1986). Thought, language, and communication in schizophrenia: Diagnosis and prognosis. *Schizophrenia Bulletin*, *12*(3), 348–359. https://doi.org/10.1093/schbul/12.3.348

Andreasen, N. C., & Pfohl, B. (1976). Linguistic analysis of speech in affective disorders. *Archives of General Psychiatry*, *33*(11), 1361–1367. https://doi.org/10.1001/archpsyc.1976.01770110089009

Arvidsson, D., Sikström, S., & Werbart, A. (2011). Changes in self and object representations following psychotherapy measured by a theory-free, computational, semantic space method. *Psychotherapy Research*, *21*(4), 430–446. https://doi.org/10.1080/10503307.2011.577824

Awad, A. G., & Voruganti, L. N. P. (2000). Intervention research in psychosis: Issues related to the assessment of quality of life 1. *Schizophrenia Bulletin*, *26*(3), 557–564. https://doi.org/10.1093/oxfordjournals.schbul.a033477

Barch, D. M., & Berenbaum, H. (1997). The effect of language production manipulations on negative thought disorder and discourse coherence disturbances in schizophrenia. *Psychiatry Research*, *71*(2), 115–127. https://doi.org/10.1016/S0165-1781(97)00045-0

Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman, M., Mota, N. B., … Corcoran, C. M. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophrenia*, *1*, 15030. https://doi.org/10.1038/npjschz.2015.30

Bedi, G., Cecchi, G. A., Slezak, D. F., Carrillo, F., Sigman, M., & de Wit, H. (2014). A window into the intoxicated mind? Speech as an index of psychoactive drug effects. *Neuropsychopharmacology*, *39*(10), 2340–2348. https://doi.org/10.1038/npp.2014.80

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology*, *9*(2), 78–84. https://doi.org/10.1027/1614-2241/a000057

Bleuler, E. (1950). *Dementia praecox or the group of schizophrenias*. Oxford, England: International Univerities Press. https://doi.org/10.1037/h0053126

Bowie, C. R., Depp, C. A., McGrath, J. A., Wolyniec, P., Mausbach, B. T., Thornquist, M. H., … Pulver, A. E. (2010). Prediction of real-world functional disability in chronic mental disorders: A comparison of schizophrenia and bipolar disorder. *American Journal of Psychiatry*, *167*(9), 1116–1124. https://doi.org/10.1176/appi.ajp.2010.09101406

Burlingame, G. M., Seaman, S., Johnson, J. E., Whipple, J., Richardson, E., Chen, S., … Lehman, A. (2006). Sensitivity to Change of the Brief Psychiatric Rating Scale-Extended (BPRS-E): An item and subscale analysis. *Psychological Services*, *3*(2), 77. Retrieved from https://login.qualifacts.org/carelogic/resources/refdocs/Markpaynehandout4.pdf

Cabana, Á., Valle-Lisboa, J. C., Elvevåg, B., & Mizraji, E. (2011). Detecting order-disorder transitions in discourse: Implications for schizophrenia. *Schizophrenia Research*, *131*(1–3), 157–164. https://doi.org/10.1016/j.schres.2011.04.026

Carpenter Jr., W. T., & Kirkpatrick, B. (1988). The heterogeneity of the long-term course of schizophrenia. *Schizophrenia Bulletin*, *14*(4), 645–652. https://doi.org/10.1093/schbul/14.4.645

Catts, S. V., & O'Toole, B. I. (2016). The treatment of schizophrenia: Can we raise the standard of care? *Australian & New Zealand Journal of Psychiatry*, *50*(12), 1128–1138. https://doi.org/10.1177/0004867416672725

Cohen, T., Blatter, B., & Patel, V. (2008). Simulating expert clinical comprehension: Adapting latent semantic analysis to accurately extract clinical concepts from psychiatric narrative. *Journal of Biomedical Informatics*, *41*(6), 1070–1087. https://doi.org/10.1016/j.jbi.2008.03.008

Covington, M. A., He, C., Brown, C., Naçi, L., McClain, J. T., Fjordbak, B. S., … Brown, J. (2005). Schizophrenia and the structure of language: The linguist's view. *Schizophrenia Research*, *77*(1), 85–98. https://doi.org/10.1016/j.schres.2005.01.016

Crow, T. J., MacMillan, J. F., Johnson, A. L., & Johnstone, E. C. (1986). A randomised controlled trial of prophylactic neuroleptic treatment. *The British Journal of Psychiatry*, *148*(2), 120–127. https://doi.org/10.1192/bjp.148.2.120

Cuesta, M. J., & Peralta, V. (1993). Does formal thought disorder differ among patients with schizophrenic, schizophreniform and manic schizoaffective disorders? *Schizophrenia Research*, *10*(2), 151–158. https://doi.org/10.1016/0920-9964(93)90050-S

Cuesta, M. J., & Peralta, V. (2011). Testing the hypothesis that formal thought disorders are severe mood disorders. *Schizophrenia Bulletin*, *37*(6), 1136–1146. https://doi.org/10.1093/schbul/sbr092

Dilsaver, S. C., Chen, Y.-W., Swann, A. C., Shoaib, A. M., Tsai-Dilsaver, Y., & Krajewski, K. J. (1997). Suicidality, panic disorder and psychosis in bipolar depression, depressive-mania and pure-mania. *Psychiatry Research*, *73*(1–2), 47–56. https://doi.org/10.1016/S0165-1781(97)00109-1

Docherty, N. M., Cohen, A. S., Nienow, T. M., Dinzeo, T. J., & Dangelmaier, R. E. (2003). Stability of formal thought disorder and referential communication disturbances in schizophrenia. *Journal of Abnormal Psychology*, *112*(3), 469–475. https://doi.org/10.1037/0021-843X.112.3.469

Docherty, N. M., DeRosa, M., & Andreasen, N. C. (1996). Communication disturbances in schizophrenia and mania. *Archives of Clinical Neuropsychology*, *53*(4), 358–364. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8634014

Docherty, N. M., & Hebert, A. S. (1997). Comparative affective reactivity of different types of communication disturbances in schizophrenia. *Journal of Abnormal Psychology*, *106*(2), 325–330. https://doi.org/10.1037//0021-843X.106.2.325

Dunayevich, E., & Keck Jr, P. E. (2000). Prevalence and description of psychotic features in bipolar mania. *Current Psychiatry Reports*, *2*(4), 286–290. https://doi.org/10.1007/s11920-000-0069-4

Elvevåg, B., Fisher, J. E., Gurd, J. M., & Goldberg, T. E. (2002). Semantic clustering in verbal fluency: Schizophrenic patients versus control participants. *Psychological Medicine*, *32*(5), 909–917. https://doi.org/10.1017/S0033291702005597

Elvevåg, B., Foltz, P. W., Rosenstein, M., & DeLisi, L. E. (2010). An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. *Journal of Neurolinguistics*, *23*(3), 270–284. https://doi.org/10.1016/j.jneuroling.2009.05.002

Elvevåg, B., Foltz, P. W., Weinberger, D. R., & Goldberg, T. E. (2007). Quantifying incoherence in speech: An automated methodology and novel application to schizophrenia. *Schizophrenia Research*, *93*(1–3), 304–316. https://doi.org/10.1016/j.schres.2007.03.001

Evangelopoulos, N., Zhang, X., & Prybutok, V. R. (2012). Latent semantic analysis: Five methodological recommendations. *European Journal of Information Systems*, *21*(1), 70–86. https://doi.org/10.1057/ejis.2010.61

Faber, R., Abrams, R., Taylor, M. A., Kasprison, A., Morris, C., & Weisz, R. (1983). Comparison of schizophrenic patients with formal thought disorder and neurologically impaired patients with aphasia. *The American Journal of Psychiatry*, *140*(10), 1348–1351. Retrieved from http://psycnet.apa.org/psycinfo/1984-04060-001

Feigenson, K. A., Kusnecov, A. W., & Silverstein, S. M. (2014). Inflammation and the two-hit hypothesis of schizophrenia. *Neuroscience and Biobehavioral Reviews*, *38*, 72–93. https://doi.org/10.1016/j.neubiorev.2013.11.006

Foltz, P. W. (2007). Discourse coherence and LSA. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 167–184). https://doi.org/10.4324/9780203936399.ch9

Gerson, S. N., Benson, D. F., & Frazier, S. H. (1977). Diagnosis: Schizophrenia versus posterior aphasia. *The American Journal of Psychiatry*, *134*(9), 966–969. https://doi.org/10.1176/ajp.134.9.966

Giles, E., Patterson, K., & Hodges, J. R. (1996). Performance on the Boston Cookie theft picture description task in patients with early dementia of the Alzheimer's type: Missing information. *Aphasiology*, *10*(4), 395–408. https://doi.org/10.1080/02687039608248419

Goksuluk, D., Korkmaz, S., Zararsiz, G., & Karaagaoglu, A. E. (2016). easyROC: An interactive web-tool for ROC curve analysis using R language environment. *The R Journal*, *8/2*, 213–230. Retrieved from https://journal.r-project.org/archive/2016/RJ-2016-042/RJ-2016-042.pdf

Goldberg, T. E., Dodge, M., Aloia, M., Egan, M. F., & Weinberger, D. R. (2000). Effects of neuroleptic medications on speech disorganization in schizophrenia: biasing associative networks towards meaning. *Psychological Medicine*, *30*(05), 1123–1130.

Goodglass, H., Kaplan, E., Barresi, B., Weintraub, S., & Segal, O. (2000). *Boston Diagnostic Aphasia Examination* (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins. Retrieved from https://scholar.google.com/scholar?hl=en&as_sdt=0%2C19&q=Goodglass+Barresi+2000 Boston+diagnostic+aphasia+examination&btnG=

Gouzoulis-Mayfrank, E., Voss, T., Mörth, D., Thelen, B., Spitzer, M., & Meincke, U. (2003). Semantic hyperpriming in thought-disordered patients with schizophrenia: State or trait?—A longitudinal investigation. *Schizophrenia Research*, *65*(2–3), 65–73. https://doi.org/10.1016/S0920-9964(03)00066-5

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun - An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, *47*, 930–944.

Gupta, T., Hespos, S. J., Horton, W. S., & Mittal, V. A. (2017). Automated analysis of written narratives reveals abnormalities in referential cohesion in youth at ultra high risk for psychosis. *Schizophrenia Research*. https://doi.org/10.1016/j.schres.2017.04.025

Harrow, M., Green, K. E., Sands, J. R., Jobe, T. H., Goldberg, J. F., Kaplan, K. J., & Martin, E. M. (2000). Thought disorder in schizophrenia and mania: Impaired context. *Schizophrenia Bulletin*, *26*(4), 879–891. https://doi.org/http://dx.doi.org/10.1093/oxfordjournals.schbul.a033502

Hartz, S. M., Pato, C. N., Medeiros, H., Cavazos-Rehg, P., Sobell, J. L., Knowles, J. A., … Pato, M. T. (2014). Comorbidity of severe psychotic disorders with measures of substance use. *JAMA Psychiatry*, *71*(3), 248–254. https://doi.org/10.1001/jamapsychiatry.2013.3726

Harvey, P. D., & Brault, J. (1986). Speech performance in mania and schizophrenia: The association of positive and negative thought disorders and reference failures. *Journal of Communication Disorders*, *19*(3), 161–173. https://doi.org/10.1016/0021-9924(86)90006-7

Harvey, P. D., Earle-Boyer, E. A., & Wielgus, M. S. (1984). The consistency of thought disorder in mania and schizophrenia: An assessment of acute psychotics. *The Journal of Nervous and Mental Disease*, *172*(8), 458–463. Retrieved from http://www.ncbi.nlm.nih.gov/sites/entrez?Db=pubmed&DbFrom=pubmed&Cmd=Link&LinkName=pubmed_pubmed&LinkReadableName=Related Articles&IdsFromResult=6747615&ordinalpos=3&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum%5Cnhttp://www.ncbi

Hasan, A., Falkai, P., Wobrock, T., Lieberman, J., Glenthøj, B., Gattaz, W. F., … Schizophrenia, W. T. F. on T. G. for. (2015). World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of schizophrenia part 3: Update 2015 management of special circumstances: Depression, suicidality, substance use disorders and pregnancy and lactation. *The World Journal of Biological Psychiatry*, *16*(3), 142–170. https://doi.org/10.3109/15622975.2015.1009163

Hasan, A., Falkai, P., Wobrock, T., Lieberman, J., Glenthoj, B., Gattaz, W. F., … Schizophrenia, W. T. force on T. G. for. (2013). World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of schizophrenia, part 2: Update 2012 on the long-term treatment of schizophrenia and management of antipsychotic-induced side effects. *The World Journal of Biological Psychiatry*, *14*(1), 2–44.

Hassan, L., Birmingham, L., Harty, M. A., Jarrett, M., Jones, P., King, C., … Shaw, J. (2011). Prospective cohort study of mental health during imprisonment. *The British Journal of Psychiatry*, *198*(1), 37–42. Retrieved from https://www.cambridge.org/core/journals/the-british-journal-of-psychiatry/article/prospective-cohort-study-of-mental-health-during-imprisonment/76CC2340DD5DB3D2723EE81D91276020

Hoffman, R. E., Kirstein, L., Stopek, S., & Cicchetti, D. V. (1982). Apprehending schizophrenic discourse: A structural analysis of the Listener's task. *Brain and Language*, *15*(2), 207–233. https://doi.org/10.1016/0093-934X(82)90057-8

Hoffman, R. E., & Sledge, W. (1988). An analysis of grammatical deviance occurring in spontaneous schizophrenic speech. *Journal of Neurolinguistics*, *3*(1), 89–101. https://doi.org/10.1016/0911-6044(88)90008-5

Hoffman, R. E., Stopek, S., & Andreasen, N. C. (1986). A comparative study of manic vs schizophrenic speech disorganization. *Archives of General Psychiatry*, *43*(9), 831–838. https://doi.org/10.1001/archpsyc.1986.01800090017003

Hollis, C. (2000). Adult outcomes of child- and adolescent-onset schizophrenia: Diagnostic stability and predictive validity. *The American Journal of Psychiatry*, *157*(10), 1652–1659. https://doi.org/10.1176/appi.ajp.157.10.1652

Holshausen, K., Harvey, P. D., Elvevag, B., Foltz, P. W., & Bowie, C. R. (2014). Latent semantic variables are associated with formal thought disorder and adaptive behavior in older inpatients with schizophrenia. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *55*, 88–96. https://doi.org/10.1016/j.cortex.2013.02.006

Holzman, P. S., Shenton, M. E., & Solovay, M. R. (1986). Quality of thought disorder in differential diagnosis. *Schizophrenia Bulletin*, *12*(3), 360–372. https://doi.org/10.1093/schbul/12.3.360

Insel, T. R. (2010). Rethinking schizophrenia. *Nature*, *468*(7321), 187–193. https://doi.org/10.1038/nature09552

Jobe, T. H., & Harrow, M. (2005). Long-term outcome of patients with schizophrenia: A review. *The Canadian Journal of Psychiatry*, *50*(14), 892–900. Retrieved from http://journals.sagepub.com/doi/pdf/10.1177/070674370505001403

Karam, Z. N., Provost, E. M., Singh, S., Montgomery, J., Archer, C., Harrington, G., & Mcinnis, M. G. (2014). Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 4858–4862). https://doi.org/10.1109/ICASSP.2014.6854525

Kavanagh, D. J., Waghorn, G., Jenner, L., Chant, D. C., Carr, V., Evans, M., … Mcgrath, J. J. (2004). Demographic and clinical correlates of comorbid substance use disorders in psychosis: Multivariate analyses from an epidemiological sample. *Schizophrenia Research*, *66*(2–3), 115–124. https://doi.org/10.1016/S0920-9964(03)00130-0

King, B. H., & Lord, C. (2011). Is schizophrenia on the autism spectrum? *Brain Research*, *1380*, 34–41. https://doi.org/10.1016/J.BRAINRES.2010.11.031

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/doi:10.1016/j.jcm.2016.02.012. PMC 4913118. PMID 27330520.

Kuperberg, G. R. (2010). Language in schizophrenia part 1: An introduction. *Language and Linguistics Compass*, *4*(8), 576–589. https://doi.org/10.1111/j.1749-818X.2010.00216.x

Le, T. P., Najolia, G. M., Minor, K. S., & Cohen, A. S. (2017). The effect of limited cognitive resources on communication disturbances in serious mental illness. *Psychiatry Research*, *248*, 98–104. https://doi.org/10.1016/j.psychres.2016.12.025

Loebel, A. D., Lieberman, J. A., Alvir, J. M., Mayerhoff, D. I., Geisler, S. H., & Szymanski, S. R. (1992). Duration of psychosis and outcome in first-episode schizophrenia. *The American Journal of Psychiatry*, *149*(9), 1183–1188. https://doi.org/10.1176/ajp.149.9.1183

Lott, P. R., Guggenbühl, S., Schneeberger, A., Pulver, A. E., & Stassen, H. H. (2002). Linguistic analysis of the speech output of schizophrenic, bipolar, and depressive patients. *Psychopathology*, *35*, 220–227. Retrieved from www.karger.com

Lukoff, D., Nuechterlein, H., & Ventura, J. (1986). Manual for the expanded brief psychiatric rating scale. *Schizophrenia Bulletin*, *12*, 594.

M.A., T., R., R., S., B., Taylor, M. A., Reed, R., & Berenbaum, S. (1994). Patterns of speech disorders in schizophrenia and mania. *Journal of Nervous and Mental Disease*, *182*(6), 319–326. Retrieved from http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L24227607%5Cnhttp://dx.doi.org/10.1097/00005053-199406000-00002%5Cnhttp://sfx.library.uu.nl/utrecht?sid=EMBASE&issn=00223018&id=doi:10.1097/00005053-199406000-00002&atitle=Patterns+o

Manschreck, T. C., Maher, B. A., & Ader, D. N. (1981). Formal thought disorder, the type-token ratio, and disturbed voluntary motor movement in schizophrenia. *The British Journal of Psychiatry*, *139*(1), 7–15. https://doi.org/10.1192/bjp.139.1.7

Marengo, J. T., & Harrow, M. (1985). Thought disorder: A function of schizophrenia, mania, or psychosis? *The Journal of Nervous and Mental Disease*, *173*(1), 35–41. Retrieved from http://journals.lww.com/jonmd/Abstract/1985/01000/Thought_Disorder__A_Function_of_Schizophrenia,.6.aspx

May, P. R., Tuma, A. H., Dixon, W. J., Thiele, D. A., & Kraude, W. H. (1981). Schizophrenia: A follow-up study of the results of five forms of treatment. *Archives of General Psychiatry*, *38*(7), 776. https://doi.org/10.1001/archpsyc.1981.01780320056006

McClellan, J. M., Werry, J. S., & Ham, M. (1993). A follow-up study of early onset psychosis: Comparison between outcome diagnoses of schizophrenia, mood disorders, and personality disorders. *Journal of Autism and Developmental Disorders*, *23*(2), 243–262. https://doi.org/10.1007/BF01046218

McNeil, M. R., & Pratt, S. R. (2001). Defining aphasia: Some theoretical and clinical implications of operating from a formal definition. *Aphasiology*, *15*(10–11), 901–911. https://doi.org/10.1080/02687040143000276

Melinder, M. R. D., & Barch, D. M. (2003). The influence of a working memory load manipulation on language production in schizophrenia. *Schizophrenia Bulletin*, *29*(3), 473–485. https://doi.org/10.1093/oxfordjournals.schbul.a007020

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. Retrieved from http://psycnet.apa.org/journals/rev/63/2/81/

Nicodemus, K. K., Elvevåg, B., Foltz, P. W., Rosenstein, M., Diaz-Asper, C., & Weinberger, D. R. (2014). Category fluency, latent semantic analysis and schizophrenia: A candidate gene approach. *Cortex*, *55*(1), 182–191. https://doi.org/10.1016/j.cortex.2013.12.004

Oltmanns, T. F., Murphy, R., Berenbaum, H., & Dunlop, S. R. (1985). Rating verbal communication impairment in schizophrenia and affective disorders. *Schizophrenia Bulletin*, *11*(2), 292–299. https://doi.org/10.1093/schbul/11.2.292

Osher, Y., & Bersudsky, Y. (2007). Thought disorder in euthymic bipolar patients: A possible endophenotype of bipolar affective disorder? *The Journal of Nervous and Mental Disease*, *195*(10), 857–860. https://doi.org/10.1097/NMD.0b013e318156832d

Owen, M. J., O'Donovan, M. C., Thapar, A., & Craddock, N. (2011). Neurodevelopmental hypothesis of schizophrenia. *The British Journal of Psychiatry*, *198*(3), 173–175. https://doi.org/10.1192/bjp.bp.110.084384

Oxman, T. E., Rosenberg, S. D., Schnurr, P. P., & Tucker, G. J. (1988). Diagnostic classification through content analysis of patients' speech. *The American Journal of Psychiatry*, *145*(4), 464–468. https://doi.org/10.1176/ajp.145.4.464

Patel, K. R., Cherian, J., Gohil, K., & Atkinson, D. (2014). Schizophrenia: Overview and treatment options. *Pharmacy and Therapeutics*, *39*(9), 638–645.

Perlini, C., Marini, A., Garzitto, M., Isola, M., Cerruti, S., Marinelli, V., … Brambilla, P. (2012). Linguistic production and syntactic comprehension in schizophrenia and bipolar disorder. *Acta Psychiatrica Scandinavica*, *126*(5), 363–376. https://doi.org/10.1111/j.1600-0447.2012.01864.x

Pilowsky, T., Yirmiya, N., Arbelle, S., & Mozes, T. (2000). Theory of mind abilities of children with schizophrenia, children with autism, and normally developing children. *Schizophrenia Research*, *42*(2), 145–155. https://doi.org/10.1016/S0920-9964(99)00101-2

Ragin, A. B., & Oltmanns, T. F. (1987). Communicability and thought disorder in schizophrenics and other diagnostic groups: A follow-up study. *The British Journal of Psychiatry*, *150*(4), 494–500. https://doi.org/10.1192/BJP.150.4.494

Raymond Lake, C. (2007). Disorders of thought are severe mood disorders: The selective attention defect in mania challenges the Kraepelinian dichotomy - A Review. *Schizophrenia Bulletin*, *34*(1), 109–117. https://doi.org/10.1093/schbul/sbm035

Remberk, B., Namysłowska, I., & Rybakowski, F. (2012). Cognition and communication dysfunctions in early-onset schizophrenia: Effect of risperidone. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *39*(2), 348–354. https://doi.org/10.1016/J.PNPBP.2012.07.007

Rice, D. P., Kelman, S., & Miller, L. S. (1992). The economic burden of mental illness. *Psychiatric Services*, *43*(12), 1227–1232. https://doi.org/10.1176/ps.43.12.1227

Richard, M. D., & Brahm, N. C. (2012). Schizophrenia and the immune system: Pathophysiology, prevention, and treatment. *American Journal of Health-System Pharmacy*, *69*(9), 757–766. https://doi.org/10.2146/ajhp110271

Roy, M.-A., Mérette, C., & Maziade, M. (2001). Subtyping schizophrenia according to outcome or severity: A search for homogeneous subgroups. *Schizophrenia Bulletin*, *27*(1), 115–138. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11215541

Rutter, D. R., Wishner, J., Kopytynska, H., & Button, M. (1978). The predictability of speech in schizophrenic patients. *The British Journal of Psychiatry*, *132*, 228–232. https://doi.org/10.1192/bjp.135.3.284a

Salzinger, K., Portnoy, S., & Feldman, R. S. (1964). Verbal behavior of schizophrenic and normal subjects. *Annals of the New York Academy of Sciences*, *105*, 845–860. https://doi.org/10.1111/j.1749-6632.1964.tb42967.x

Saykin, A. J., Gur, R. C., Gur, R. E., Mozley, P. D., Mozley, L. H., Resnick, S. M., … Stafiniak, P. (1991). Neuropsychological function in schizophrenia: Selective impairment in memory and learning. *Archives of General Psychiatry*, *48*(7), 618–624. https://doi.org/10.1001/archpsyc.1991.01810310036007

Schuckit, M. A. (2006). Comorbidity between substance use disorders and psychiatric conditions. *Addiction*, *101*(s1), 76–88. https://doi.org/10.1111/j.1360-0443.2006.01592.x

Solomon, M., Ozonoff, S., Carter, C., & Caplan, R. (2008). Formal thought disorder and the autism spectrum: Relationship with symptoms, executive control, and anxiety. *Journal of Autism and Developmental Disorders*, *38*(8), 1474–1484. https://doi.org/10.1007/s10803-007-0526-6

Solovay, M. R., Shenton, M. E., & Holzman, P. S. (1987). Comparative studies of thought disorders: I. Mania and schizophrenia. *Archives of General Psychiatry*, *44*(1), 13–20. https://doi.org/10.1001/archpsyc.1987.01800130015003

Taylor, W. L. (1953). “Cloze procedure”: A new tool for measuring readability. *Journalism Bulletin*, *30*(4), 415–433. https://doi.org/10.1177/107769905303000401

Tully, L. M., Niendam, T. A., Iosif, A.-M., Kumar, D., Nye, K. E., Denton, J. C., … Pierce, K. M. (2017). Enhancing early psychosis treatment using smartphone technology: A longitudinal feasibility and validity study. *Journal of Psychiatric Research*, *96*, 239–246. https://doi.org/10.1016/j.jpsychires.2017.10.017

Van Der Gaag, R. J., Caplan, R., Van Engeland, H., Loman, F., & Buitelaar, J. K. (2005). A controlled study of formal thought disorder in children with autism and multiple complex developmental disorders. *Journal of Child & Adolescent Psychopharmacology*, *15*(3), 465–476. https://doi.org/10.1089/cap.2005.15.465

Wang, P. S., Demler, O., & Kessler, R. C. (2002). Adequacy of treatment for serious mental illness in the United States. *American Journal of Public Health*, *92*(1), 92–98. https://doi.org/10.2105/AJPH.92.1.92

Wechsler, D. (2009). *Wechsler Memory Scale–Fourth Edition*. San Antonio, TX: Pearson.

Whitaker, R. (2004). The case against antipsychotic drugs: A 50-year record of doing more harm than good. *Medical Hypotheses*, *62*(1), 5–13. https://doi.org/10.1016/S0306-9877(03)00293-7

Appendix A. App Instructions & Story Text

Instructions “Listen to this talk about TITLE and remember it.”*
<App presented story (see below) orally in a male voice.>
<App prompted participant to recall the story with as many details as possible. The iPod device recorded the recall, for a maximum of two minutes, as well as presented the remaining time left on the screen of the app.>
Instructions “Remember this talk about TITLE so you can re-tell it again later.”
<At next session with the iPod device, between 2 to 5 days, the participant completed other tasks and was prompted to recall the original story.>

*The iPod device presented each participant with one story per session, which were counterbalanced using a rotating design. This means each participant was only presented with each story a maximum of one time and all stories were sampled across participants.

A. A Beach Picnic. A group of friends went to the beach for the day. In the morning, they went swimming and then lie on their towels sunbathing. Around noon a fisherman walked by carrying a very large fish. “I guess that means its lunch time” said one of the friends. They opened up their picnic basket and had sandwiches, apples and cool sodas. The sky became cloudy and they decided to leave for home before it started raining.
B. A Closed Street. One rainy Saturday morning, Stanley rode his bicycle into town. At 6thStreet, a policeman waved him over and told him that up ahead the street was closed. Stanley kept on riding. A few blocks on, he could see red barriers blocking the road and a crowd of people standing with umbrellas. He got off his bike and as he approached the crowd, he saw a band marching in a parade.
C. A Lost Horse. Jill’s daughter had left the gate open again. Their black horse wandered out of the stable and crossed the street. It jumped over a low fence and joined five of their neighbor’s horses in a pasture. Three of those horses were also black. Jill noticed that her horse was missing and frantically looked all around. At first she didn’t realize her horse was in the pasture since it looked similar to her neighbor’s horses. But it came when she called its name.
D. Eyeglasses. Simon checked his receipt at the supermarket. As he picked up his groceries he put his eyeglasses on the checkout counter instead of into his pocket. Simon walked all the way home, stopping to pick up the mail. Once inside he realized that he could not read the envelopes. He called the supermarket and asked if they had found any eyeglasses. The store manager said the glasses could be picked up at lost and found.
E. Paper Airplanes. Cindy and her brother Alex were throwing paper airplanes out of the upstairs window to see which would go farthest. Cindy’s airplane made two circles and then landed on the grass across the street. Alex’s airplane dove straight down and landed on the street. A truck drove by and the wind from the truck blew Alex’s airplane on to the grass beyond Cindy’s airplane. “I win”, said Alex. Cindy cried, “That’s not fair.”
F. Babysitting. Jasmine was babysitting her sister’s two children. She took them to the park. Her niece joined a group of kids playing in the yard with a ball, while her nephew joined a crowd of kids playing in the sandbox. The day was sunny and hot and the children were having a great time running around. Luckily, when it came time for a snack, Jasmine had brought juice and cookies for the children.

G. A Mountain Dog. Brittany and Emma visited Switzerland during their winter break from college. They visited small villages and toured a watch museum. In the mountains, they skied during the day and ate cheese and chocolate at night. In one town they saw cows and a dairy with a working Mountain dog. The dog was pulling a cart with big metal milk cans from the dairy to market. The owner took their picture with the dog and cart.

H. An Art Gallery. Chris slowly moved from one painting to the next at the art gallery. He studied each picture for a long time. A young woman came into the room, walked over and asked him what he thought of the paintings. As they talked about the painting, Chris began to realize that this woman knew a lot about painting. “Do you work for the gallery?” he asked. “Oh no”, she replied, “I’m the artist.”

I. A School Concert. Ms. Heather had worked hard to prepare her elementary school students for their concert. When onstage they were to sing, when offstage they were to sit and be quiet. The parents came into the theater and sat quietly. The first three songs went well and the parents applauded. Unfortunately, the children were now so excited that they started jumping up and dancing round and round. The parents' huge laughs and loud clapping indicated a successful concert.

J. A Giant Clock. Mrs. Winkler was visiting a new city with her daughter Tina. Usually they loved to visit new towns and explore new shops. Unfortunately, this city didn't seem to have fun places for kids. Entering the downtown area, they saw a giant clock that had little figures going around and a huge giraffe by the entrance. Inside they found a magical place with every kind of toys. Mrs. Winkler let Tina pick one to take home.

K. A Bakery Lunch. Betty and Helen had spent the morning shopping. They were now ready to sit down and have lunch. They spotted a bakery with display cases filled with beautiful cupcakes and pastries. Though they were tempted, they decided they didn't want to just eat desert for lunch. They were about to move on, when they noticed in the back of the bakery was a small coffee shop with sandwiches. This was the perfect place for lunch.

L. Horse Play. The Johnsons owned a stable out in the country. They allowed their friends to keep horses at the stable and offered lessons. They also had a play area to keep children busy when they weren't taking lessons. One afternoon, a child kicked a ball over the fence into a pasture with some of the horses. Two of the horses ran to the ball and it looked like they were playing soccer.

M. Borrowing A Pencil. Lindsey listened as the teacher assigned a project. She put her notebook on the desk, but couldn’t find a pencil in her backpack. Lindsey leaned over to ask Martin, who was sitting next to her, for a pencil. Instead Martin handed her a pen. Try as she might, Lindsey could not get the pen to write. Her friend Tiffany saw her shaking the pen in frustration and offered her a pencil.

N. Buying Candy. The day after Ben's 10th birthday he walked to town to buy candy. He looked in the store's window to decide which candy to buy. He finally decided on the large chocolate bar. Ben searched through his pockets, but he did not have enough money. Thinking what to do, he looked down and saw two shiny coins wedged in a crack in the sidewalk. He picked them up and went into the store and bought his candy bar.

O. Waiting For A Friend. Amelia sat in a café with a cup of coffee, reading the newspaper. Every so often she would glance out the window to see if her friend Lilly had arrived. Lilly had just been to Paris and Amelia wanted to hear all about her trip. Amelia noticed in the newspaper that a nearby store was having a half price sale. “I bet that is why Lilly is late,” she thought.

P. A Dining Room. Lucy and her husband loved eating in their dining room. Lucy thought that the color of the room's walls was too dark. She went to the store and brought back a book of sample colors. Though she looked at hundreds of colors, none seemed to be exactly right. Her husband suggested adding a light fixture and replacing their window blinds with ones that let in more light. With those changes the room was perfect.

Q. Buying Milk. One sunny afternoon, a woman drove to the store to buy some milk. As she got to the cash register, she realized that she had forgotten her purse in the car. Luckily, a guy she knew was also at the store and offered to pay for the milk. As they walked outside to the parking lot together, she thanked him, repaid him, and offered him a ride home.

R. Ice Cream. Two boys where walking down the street on a super-hot day when they heard the ice cream truck. Both boys wanted popsicles, but neither had enough money to buy one. One of the boys suggested that they combine their money and buy one double popsicle and split it in half. After counting all their money, they were both able to get a frozen treat on that sunny day.

S. Party Balloons. At the little girl's birthday party, there were many balloons. She loved them all, but loved the polka dot balloon best. She ran to the backyard to show her friends the colorful balloon. A neighbor boy asked if he could hold it. Right as she gave it to him, it flew away and got stuck in a tree. Luckily, her dad was able to get it back down for her.

T. Missing Homework. Susana ran down the hallway, excited to turn in the assignment that she had worked on all week. She stopped to read it one last time before school started. She looked for it in her backpack, but couldn't find it. She was walking into class frustrated and empty handed, when her teacher said that her assignment was amazing. Her mom had brought it in before class started. Susana was overcome with joy and relief.

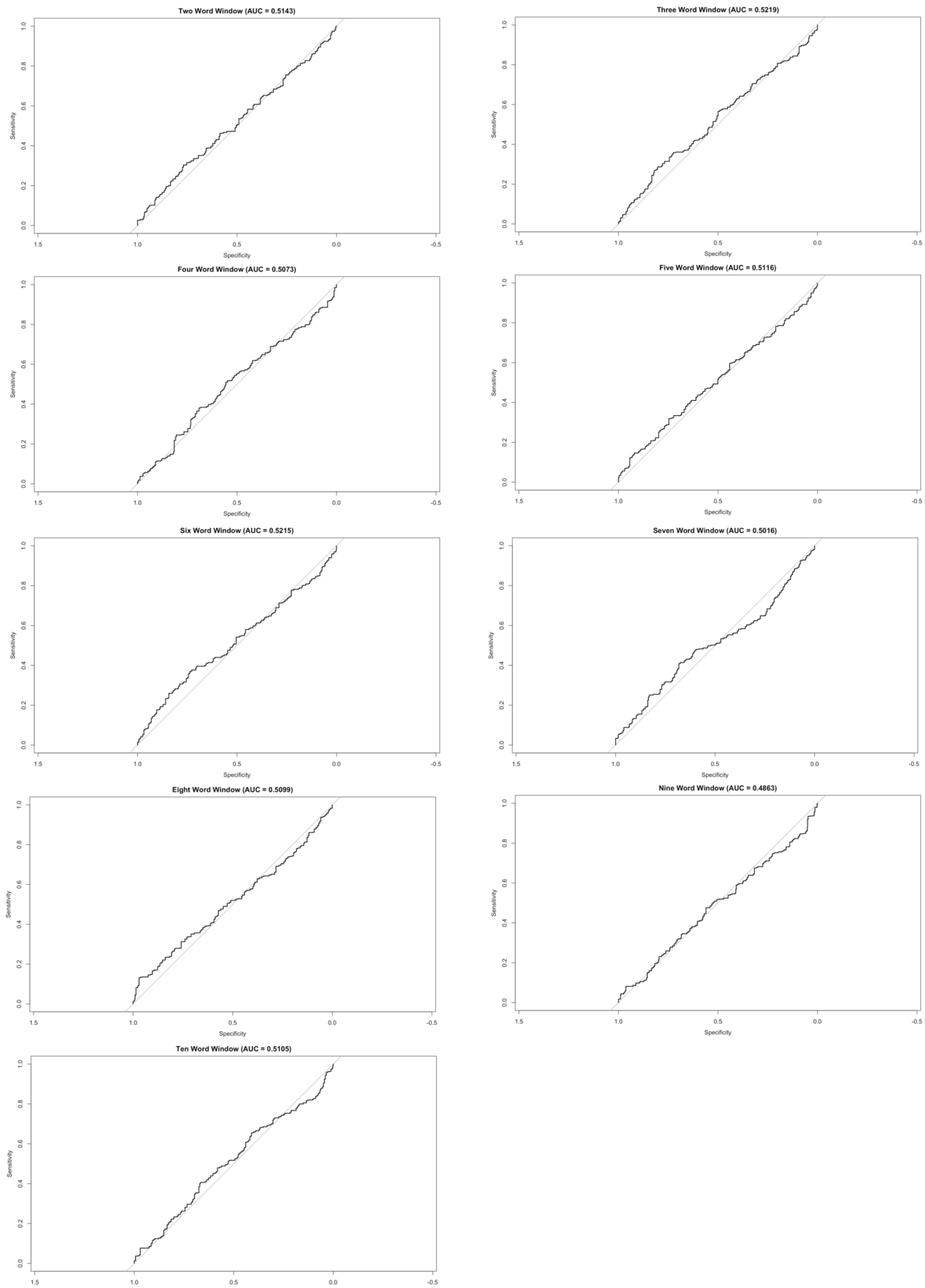
U. A Gas Station. Joshua was excited to get to work. He was going to get a promotion. As he started his car, he noticed it was way low on gas. He drove to the nearest gas station, but it was crowded with cars. He didn't want to be late on his big day, so he asked a lady if he could go ahead of her. She said she'd be OK with that and let him in ahead of her in line.

V. Car Keys. The woman's dog took her car keys and ran outside. She ran after the dog, but couldn't catch him. The dog thought it was a game, but the woman got angry. She was going to be late for a meeting. After chasing him around for five minutes, she took out some treats and told the dog to sit and drop the keys. The dog immediately obeyed and she got to her meeting on time.

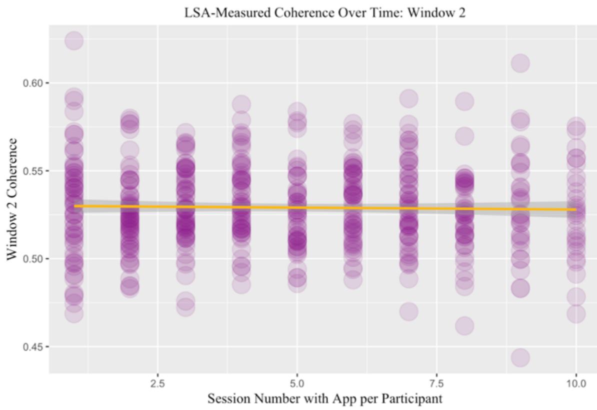
W. Over-Sleeping. On Monday morning, Susan woke up more tired than usual. When she walked downstairs to make herself a cup of coffee, she found her husband in the kitchen. She was surprised because he usually left an hour before she woke up. Her husband greeted her and reminded her that daylight savings time was over. Realizing the clocks were wrong, she happily ran upstairs and jumped back into bed.

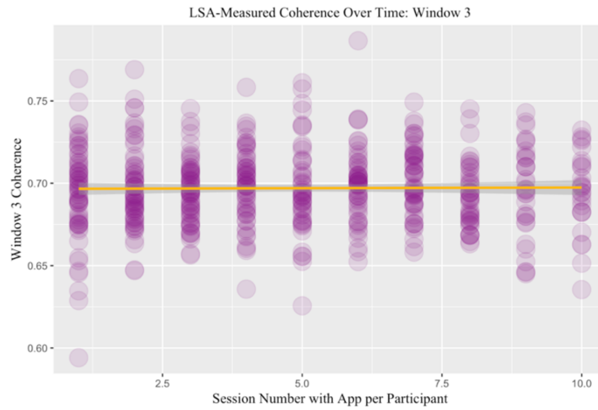
X. Learning To Drive. Learning how to drive a car was nerve wracking. Maria was doing well, but she felt like she was going to crash into everything. Her father was teaching her and one day he started singing a lullaby to her, which seemed to calm her down. Soothing music really relaxed her. On the day of her driver's test, she hummed all the way through the test and she passed with ease.

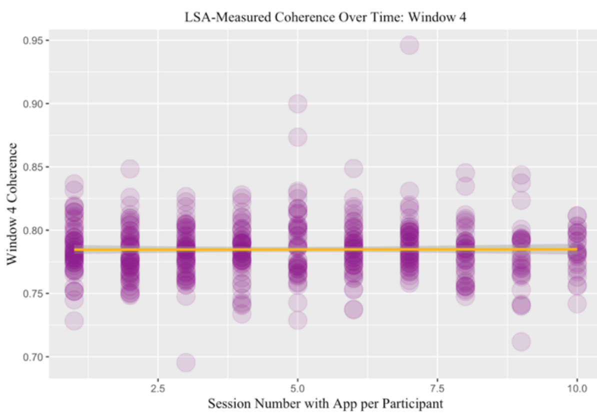
Appendix B. ROC Curves of Window Sizes Two to Ten Words

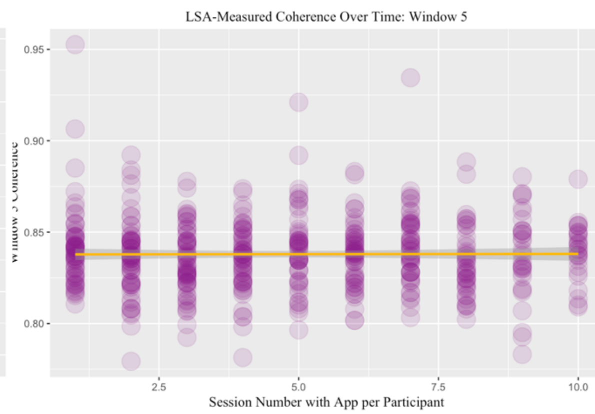


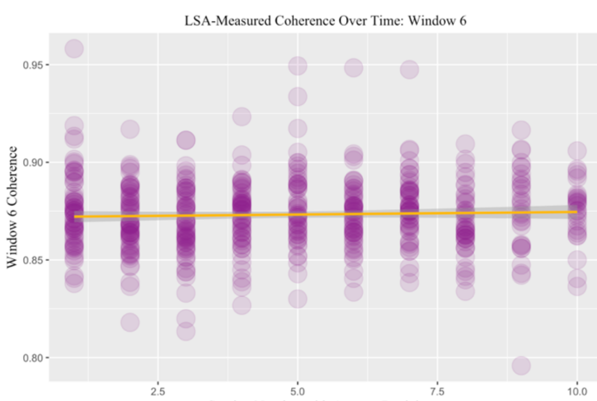
Appendix C. Graphs of ICC Values over Time Using all Window Sizes

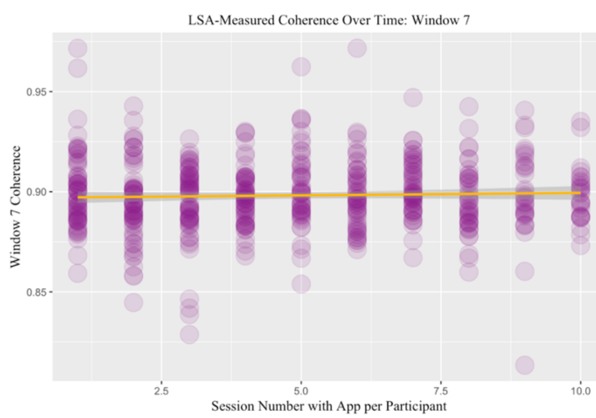


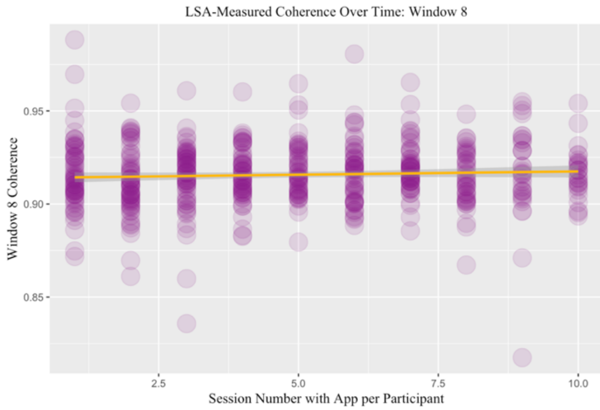


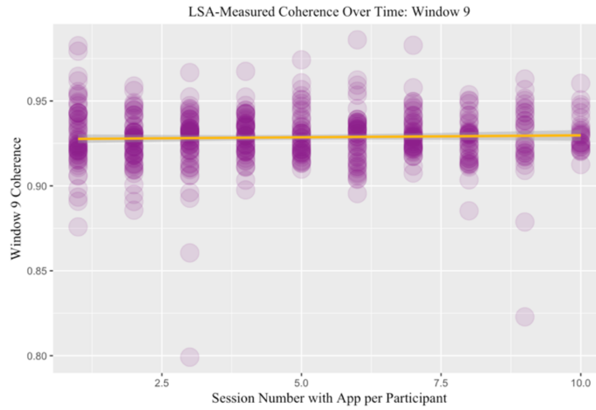


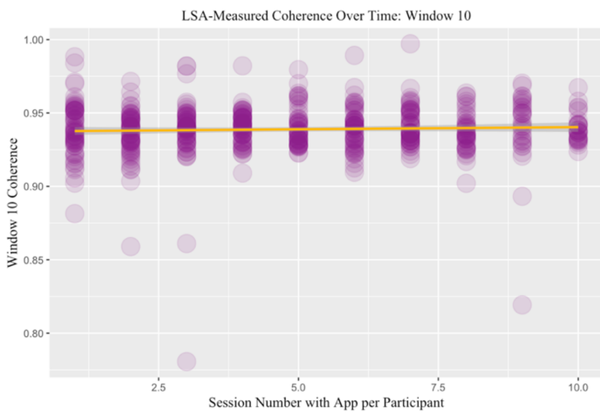







LSA-Measured Coherence Over Time: Window 8
Window 8 Coherence
0.95
0.90
0.85
2.5
5.0
7.5
10.0
Session Number with App per Participant


LSA-Measured Coherence Over Time: Window 9
Window 9 Coherence
0.95
0.90
0.85
0.80
2.5
5.0
7.5
10.0
Session Number with App per Participant


LSA-Measured Coherence Over Time: Window 10
Window 10 Coherence
1.00
0.95
0.90
0.85
0.80
2.5
5.0
7.5
10.0
Session Number with App per Participant

Vita

Taylor Leanne Fedechko, born in Mystic, Connecticut, worked as a research assistant for a few years in California after receiving her Bachelor's degree from the University of Maryland, College Park. In Fall 2016, she entered the Department of Psychology at Louisiana State University. Her primary research interests are serious mental illness and language analysis.