NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Evaluating measures of global coherence ability in stories in adults

Heather Harris Wright[†], **Gilson J. Capilouto**[‡], and **Anthony Koutsoftas**[§]

[†]Communication Sciences & Disorders, East Carolina University, Greenville, NC, USA

[‡]Rehabilitation Sciences, University of Kentucky, Lexington, KY, USA

[§]Seton Hall University, South Orange, NJ, USA

Abstract

Background—Discourse coherence is a reflection of the listener's ability to interpret the overall meaning conveyed by the speaker. Measuring global coherence (maintenance of thematic unity of the discourse) is useful for quantifying communication impairments at the discourse level in clinical populations and for measuring response to discourse-level treatments.

Aims—The aim was to determine feasibility of a four-point global coherence scale developed by the authors. Specifically, they were (1) to estimate measurement reliability of the four-point global coherence scale; and (2) to estimate construct validity for the four-point global coherence scale.

Method & Procedures—Fifty cognitively healthy adults aged between 28 and 58 years participated in the study. Participants viewed and then told the stories depicted in two wordless picture books. Participants' stories were orthographically transcribed and segmented into communication units (C-unit). Raters scored each participant's story for global coherence using two global coherence scales (four- and five-point scales). Each C-unit received an individual score, then the mean global coherence score was computed, resulting in two mean global coherence scores for each coherence scale, one for each story, for all participants.

Outcomes & Results—Results indicated high reliability estimates for the scale. In addition, construct validity, specifically face validity and convergent validity, was effectively estimated for using the four-point scale as a measure of maintenance of global coherence in stories told by cognitively healthy adults. Lastly, it was found that the wordless picture books elicited stories that are comparable and can be reliably interchanged as different forms to evaluate maintenance of global coherence.

Conclusions & Implications—The assumptions that the measure is feasible were achieved and face and convergent validity were adequately estimated. Future investigations should consider estimating predictive validity, concurrent validity and discriminant validity of the measure.

Keywords

discourse; coherence; narratives

Address correspondence to: Heather Harris Wright, Communication Sciences & Disorders, East Carolina University, 3310 P Allied Health Sciences, MS 668, Greenville, NC 27834, USA; wrighth@ecu.edu.

Introduction

Discourse coherence is a reflection of the listener's ability to interpret the overall meaning conveyed by the speaker. Several researchers have conceptualized different 'levels' of discourse coherence—*global* and *local* (Agar and Hobbs 1982, Glosser and Deser 1992, Kintsch and van Dijk 1978). Global coherence refers to how the measured units of discourse (i.e. utterance, proposition, verbalization and sentence) maintain the overall topic; whereas local coherence refers to how the content from one unit of discourse relates to the content of the preceding unit. Global coherence is the focus of the current study. Maintenance of global coherence ability has been explored in discourse produced by children, adults across the lifespan and adults with acquired neurogenic disorders (e.g. aphasia, dementia and traumatic brain injury). Discourse organization is realized through global coherence and measurement of it may be useful for quantifying age-related changes in macro-linguistic organization. Additionally, measuring global coherence is useful for quantifying communication impairments at the discourse level in clinical populations and for measuring response to discourse-level treatments.

Researchers have developed and applied different methods for measuring coherence ability. Glosser and Deser (1990, 1992) provided the seminal work and methodology for investigating coherence ability in cognitively healthy adults as well as in clinical populations. They developed a five-point rating scale of global coherence ability that paralleled their definition of global coherence (i.e. maintenance of overall topic). Alternatively, others have developed methods that include measures of frequency and type of coherence violations (e.g. Christiansen 1995) and degree of global coherence and global coherence errors (e.g. Marini *et al.* 2005). These methods align more closely with the conceptualization of global coherence as a measure of the completeness of the story gist and relating to or requiring knowledge and production of story structure rather than maintenance of thematic unity of the discourse.

Glosser and Deser (1992) developed and applied their five-point rating scale to measure global coherence ability in discourse samples obtained from two groups of cognitively healthy adults: middle aged (mean age = 51.9 years) and elderly (mean age = 76.2 years). The discourse elicitation tasks included describing family and work experiences. Language samples were transcribed and segmented into verbalizations. Each verbalization was scored and then a mean global coherence score for each sample was computed. A high global coherence score (i.e. 5) indicated that the verbalization included 'substantive information directly related to the designated topic' (Glosser and Deser 1992: 268). A low global coherence score (i.e. 1) indicated the verbalization was incoherent. Glosser and Deser found that the middle-aged group had a significantly better mean global coherence score compared with the older group (4.28 versus 3.69). Further, the middle-aged group had significantly fewer incoherent verbalizations (i.e. scores of 1) compared with the older group (5.1% versus 17.9%) suggesting that the older group abandoned the topic and became tangential with greater frequency than the middle-aged group, subsequently disrupting discourse organization (Glosser and Deser 1992).

Van Leer and Turkstra (1999) adapted the Glosser and Deser (1990) five-point rating scale and included more explicit procedures in their investigation of global coherence ability in discourse samples collected from adolescents with and without brain injury. During data analysis, they found that scores 2 and 4 were rarely assigned by trained raters. Rather than reporting mean scores, they collapsed the scores into three rating levels: low coherence (scores 1 and 2), medium coherence (score 3) and high coherence (scores 4 and 5) and computed the per cent of occurrence for each level. They reported no difference between the groups on the global coherence measure. Coelho and Flewellyn (2003) used van Leer and

Turkstra's (1999) adapted version and Hough and Barrow (2003) used Glosser and Deser's (1990) five-point rating scale with clinical populations (persons with aphasia and traumatic brain injury, respectively). Both followed Glosser and Deser's procedures of computing mean scores. Coelho and Flewellyn (2003) went a step further and converted the mean scores to *z*-scores to compare performance by their individual with aphasia with a control group. Both Coelho and Flewellyn and Hough and Barrow found that their control groups yielded higher global coherence scores compared with clinical participants.

Rogalski *et al.* (2010) examined the relationship between cognitive variables and discourse coherence in mobility-impaired stroke survivors. They also used van Leer and Turkstra's (1999) adapted version of Glosser and Deser's (1990) coherence rating scale but only the scores 5, 3 and 1. A mean global coherence score was computed across the discourse tasks for each study participant. Several measures of attention, working memory and processing speed were also administered. Rogalski *et al.* (2010) found a significant relationship among mean global coherence scores and performance on sustained attention and processing speed measures and concluded that maintaining global coherence is cognitively demanding.

Taken together, these studies demonstrate that Glosser and Deser's (1990, 1992) five-point rating scale and its adapted version (i.e. van Leer and Turkstra 1999) has been used in multiple forms across multiple studies. The scale was developed to parallel the definition of global coherence (i.e. maintenance of overall topic) suggesting that it has strong face validity. However, measurement reliability and validity of the scale have not been estimated and cannot be assumed across different populations without empirical evidence. Additionally, data extracted from individual studies and subjected to statistical analyses have differed because of the adaptations made to the scale. For example, van Leer and Turkstra (1999) reported the per cent occurrence of combined scale levels because two of the levels were rarely used by their raters; whereas Rogalski *et al.* (2010) only used three of the levels. These modifications to the scale suggest that having five levels should be reconsidered. A revised scale with fewer level choices, then, may be a better option.

Further, of particular importance to the current study is the observation that measurement reliability and validity have not been estimated empirically, which potentially limits interpretation and application of the study results. The goal of the current study was to estimate reliability and validity of a global coherence scale. The rating scale developed and used in the current study was based on a general concept of coherence—that it is a reflection of the listener's ability to interpret the overall meaning conveyed by the speaker; and, more specifically, how well each discourse unit maintains the overall topic. The authors' conceptualization of global coherence aligns with that of Glosser and Deser's (1992). Therefore, their measure is used, in part, to estimate if the global coherence scale is valid.

The purpose of current study was to determine the feasibility of a four-point global coherence scale developed by the authors. Specifically, the aims were (1) to determine measurement reliability of the four-point global coherence scale; and (2) to estimate construct validity for the four-point global coherence scale. In a pilot study, we investigated reliability of the four- and five-point scales and validity for the four-point scale with a small sample of participants with and without aphasia ($N = 15$ in each group) (Wright *et al.* 2010). For the discourse elicitation task, participants told stories depicted in two wordless picture books. The two global coherence scales significantly correlated across the stories, providing evidence for the four-point scale's convergent validity. Further, global coherence scores strongly correlated between the two stories for the four-point scale ($r = 0.955$) and the five-point scale ($r = 0.614$) suggesting they are reliable measures.

We are extending the present work from the pilot study by applying both global coherence measures (a five-point global coherence scale and a four-point global coherence scale) to storytelling discourse samples from a larger group of cognitively, healthy adults. Storytelling narratives, elicited from wordless picture books, were selected to address the study aims because storytelling tasks are considered more representative of spontaneous communication (Liles 1993) and produce longer narratives and more discourse for analyses than picture description tasks.

Discourse schema, cognitive demands and age may contribute to how well discourse is coherently produced. However, reliable and valid measures of coherence are necessary in order to examine these relationships systematically and evaluate and inform theories of discourse production empirically.

Methods

Participants

A subset of participants was randomly selected from a larger database. Fifty cognitively healthy adults aged between 28 and 58 years (mean age = 47.72 years, SD = 6.44 years) participated in the study. An equal number of males and females participated; the mean years of education completed for the group was 15.6 years (SD = 2.48 years). All participants met the following study inclusionary criteria: (1) aided or unaided visual acuity within normal limits, as indicated by passing a vision screening (Beukelman and Mirenda 1998); (2) aided or unaided hearing within functional limits as measured by the Central Institute for the Deaf (CID) List of Everyday Speech (Davis and Silverman 1978); (3) no presence of depression at the time of the study participation as measured by performance on the short form of the Geriatric Depression Scale (GDS; Sheikh and Yesavage 1986); (4) normal cognitive functioning as indicated by scaled score performance on the Mini Mental State Examination (MMSE; Folstein and Folstein 2002); (5) no history of stroke, head injury, or progressive or acquired neurogenic disorder per self report; and (6) English as their first language per self report. For participants' demographic data, see table 1.

Storytelling task

Participants viewed and told the stories depicted in two wordless picture books; *Picnic* (McCully 1984) and *Good Dog Carl* (Day 1985). *Picnic* is a story about a family of mice who drive to the forest for a picnic. The baby mouse falls out of the truck on the way to the picnic site; however, the family does not notice and continues on without her. The family eventually realizes the baby mouse has been lost, and the story concludes when the family finds the baby mouse back on the road and decides to have their picnic then and there. In *Good Dog Carl*, a mother asks the family dog, Carl, to look after the baby in his crib while she is gone. Carl and the baby get into mischief all over the house and make a mess. However, by the time the mother returns, Carl has bathed the baby, put him back into his crib and cleaned the mess. The mother tells Carl he is a good dog as she does not know what happened while she was gone.

Global coherence analyses

The purpose of the global coherence analyses was to measure participants' ability to maintain the overall topic/theme for the discourse elicitation task. Prior to completing the global coherence analyses, all discourse samples were audio or video recorded, then orthographically transcribed and segmented into C-units. A C-unit is a communication unit and includes an independent clause with its modifiers (Loban 1976); it is commonly used to segment oral discourse samples (Hughes *et al.* 1997). For both global coherence scales, each C-unit received an individual score then the mean global coherence score was computed

resulting in two mean global coherence scores for each coherence scale, one for each story, for all participants. An example of an utterance segmented into C-units is as follows:

- *Pre-C-unit segmented sample:* there's a family of mice who live in a house in the forest and one day they decide to pack everyone up a large family of mice into the truck and go out for a picnic with the whole family.

- *C-unit segmented:* (1) there's a family of mice who live in a house in the forest; and (2) one day they decide to pack everyone up a large family of mice into the truck and go out for a picnic with the whole family (Wright and Capilouto 2009: 1299).

Four-point scale

Using van Leer and Turkstra's (1999) adapted version of Glosser and Deser's (1990) global coherence scale as a base, the authors developed the four-point scale for scoring global coherence ability. A high global coherence score (4) was assigned to C-units that were overtly related to the stimulus and included details of significant importance to the main details of the stimulus or topic. A low global coherence score (1) was assigned to C-units that were entirely unrelated to the stimulus or topic. For the scoring criteria for the four-point global coherence rating scale, see table 2.

Scorers followed a multi-step training protocol for completing the global coherence analysis prior to scoring independently study participants' transcripts. The training protocol included first having the scorer review discourse task stimuli and scoring procedures. Next, the scorer reviewed two transcripts that had been marked up indicating global coherence scores for each C-unit. For each global coherence score, an explanation was provided indicating why the C-unit received the assigned score. For the final step, the scorer completed the global coherence analysis on two transcripts. The scorer compared their results with previously scored transcripts for the same discourse samples. Scorers tallied the number of agreements and disagreements. For any disagreements, the scorer was referred to the explanation provided on the previously scored transcript. Once the scorer was in 100% agreement with the previously scored transcript, training was considered complete. Scoring procedures and training protocol are available from the authors upon request.

Five-point scale

For comparison and to estimate the validity of the measure, we used the five-point scale that van Leer and Turkstra (1999) adapted from Glosser and Deser (1990). The purpose of the scale was to quantify global coherence as defined by how each utterance related to the overall meaning of the established topic (Glosser and Deser 1990, 1992, van Leer and Turkstra 1999). A score of 5 indicated that 'the utterance provides substantive information related to the general topic' (van Leer and Turkstra 1999: 344) and a score of 1 indicated that 'the utterance is unrelated to the general topic or is a comment on the discourse' (p. 345). Research assistants followed van Leer and Tursktra's procedures for scoring the discourse samples (for procedures, see van Leer and Turkstra 1999). No additional procedures or training protocols were created for scoring the samples.

Experimental procedures

All participants were tested individually and attended two sessions, each lasting no more than 2 h. In the initial session, participants gave consent, were screened to determine eligibility to complete the study, and provided medical and demographic data. This was followed by completion of discourse tasks or a cognitive test battery, referred to as the discourse session and the cognitive session, respectively. Session order was randomized across participants. The discourse session included 11 discourse elicitation tasks randomized

across participants. Only results from the discourse session, specifically, those of the storytelling task, are reported herein. Data obtained during the cognitive session are not reported. All discourse samples were either audio or video recorded.

For the storytelling task, the examiner read the following script:

> These are children's books without words—so that a person can make up their own story. First, I will look through the children's book and get an idea of the story. Then, I will start at the beginning and tell you the story that goes with the pictures.

Next, the examiner read the scripted storytelling of *The Great Ape* (Krahn 1978) to show the participant how the task was to be completed. The examiner then gave the participant one of the wordless picture books and said, 'Now, it is your turn. Look at this book and when you are ready tell me the story that goes with the pictures.' Participants were provided an unlimited amount of time to look through the book and they were also allowed to look at the pictures in the book during the storytelling. The order of the picture books was randomized across participants.

Transcription and rater reliability

Inter- and intra-rater reliability for word-by-word agreement and C-unit segmenting were determined for 10% of the total samples ($n = 5$) collected from the participants (i.e. including both stories). Agreements and disagreements were subjected to the following formula:

Total agreements/[total agreements + total disagreements] $\times$ 100

Word-by-word transcription inter-rater agreement was 97.5% and intra-rater agreement was 98.6%. C-unit segmentation inter-rater agreement was 90.5% and intra-rater agreement was 88.4%.

Inter-rater reliability for both global coherence ratings was calculated on a random selection of five transcripts, including both stories (10%), where a second research assistant applied the same scoring procedures to the transcripts. Inter-rater reliability for global coherence scales was 95.78% (range = 89.4–100%) for Glosser and Deser's (1992) scale and 98.19% (range = 95.1–100%) for the four-point scale. Intra-rater reliability was calculated by having the same rater score a random selection of five transcripts, including both stories (10%), which they previously scored for global coherence. Intra-rater reliability for global coherence scales was 97.91% (range = 92.9–100%) for Glosser and Deser's (1992) scale and 97.45% (range = 93.5–100%) for the four-point scale. The rater agreement results are comparable with results from previous studies with the scales. Koutsoftas *et al.* (2009) calculated rater reliability on a random selection of 10% of transcripts and reported 99.5% intra-rater agreement and 91.7% inter-rater agreement.

Results

Prior to performing the statistical analyses, data were examined for accuracy of data entry, missing values, univariate outliers, and fit between variables' distributions (in terms of skewness and kurtosis), and the assumptions of univariate analysis (i.e. gross violations of normality) using PASW Statistics 18.0.1 (SPSS, Inc., Chicago, IL) program. No outliers or cases with missing data were identified. Histograms were used to assess the shape of distributions and were found to be satisfactory.

To determine if gender was a significant factor that needed to be considered in subsequent analyses, paired sample *t*-tests were performed for both scales and each story. To control for Type I error, family-wise error rate across the *t*-tests was controlled using the Bonferroni

approach with alpha set to 0.0125. For the five-point scale, males and females did not significantly differ for *Good Dog Carl*, $t(24) = 2.45$, $p = 0.022$; but they did for *Picnic*, $t(24) = 2.84$, $p = 0.009$. No significant gender differences were found for the four-point scale for either story; $t(24) = 0.773$, $p = 0.45$ (*Good Dog Carl*) and $t(24) = 2.66$, $p = 0.014$ (*Picnic*). Because gender was not a significant factor for most of the comparisons, it was not considered in subsequent analyses. For mean scores for the entire sample, and by gender, for both scales, see table 3.

Reliability and validity estimates

One method for estimating reliability of the scales is to determine the repeatability of participants' scores across similar discourse elicitation tasks; in this case, stories elicited from wordless picture books. Based on our pilot work, we expected significant correlations between the two stories for both global coherence measures. Pearson correlation coefficients were calculated for both scales, between the two stories. For the five-point global coherence scale, a statistically significant correlation between *Picnic* mean scores and *Good Dog Carl* mean scores was found: $r = 0.66$, $p < 0.0001$. Similarly, a statistically significant correlation was found for the four-point scale: $r = 0.71$, $p < 0.0001$.

Convergent validity is one type of construct validity and is a measure's ability (i.e. four-point scale) to vary directly with a similar measure (i.e. five-point scale) of the same construct (i.e. global coherence). The five-point scale was used in previous studies, was developed based on Glosser and Deser's (1990, 1992) conceptualization of global coherence, and has good face validity. Convergent validity of the four-point scale may be determined by evaluating if it varies directly with the five-point scale across the same discourse elicitation tasks. To estimate convergent validity of the four-point scale, Pearson correlation coefficients were calculated for each story. Statistically significant correlations were found between the four- and five-point scales for *Picnic*, $r = 0.79$, $p < 0.0001$ and *Good Dog Carl*, $r = 0.66$, $p < 0.0001$, indicating a strong relationship between the two scales for measuring global coherence. For Pearson correlation coefficients among the global coherence scales, see table 4.

Discussion

The goal of this study was to establish the feasibility of a four-point scale of global coherence by determining if it is a reliable and valid measure. Global coherence was evaluated in story narrative discourse samples collected from cognitively healthy adults using the four-point scale and Glosser and Deser's (1992) five-point scale. Reliability estimates for both scales were high. The four-point scale is based on the concept that global coherence is a reflection of the listener's ability to interpret the overall meaning conveyed by the speaker and is measured by how well each discourse unit (i.e. C-unit) maintains the overall topic. The results suggest that the scale aligns well with how global coherence is operationalized. Further, the four-point scale was partly developed to address limitations with the five-point scale that have resulted in several modifications to it. These limitations have included scoring modifications (i.e. grouping scores together to compute the per cent of occurrence) and eliminating levels (i.e. removing 2 and 4) because they are rarely used. Despite the limitations, the five-point scale has been readily used in the literature and was developed based on a conceptualization of global coherence that closely aligns with ours; therefore, it was used as one estimate of construct validity of the four-point scale. Construct validity, specifically face validity and convergent validity were effectively estimated for using the four-point scale as a measure of maintenance of global coherence in stories told by cognitively healthy adults.

Reliability

Reliability is the repeatability or consistency of an observation. For the current study, the observation of interest was global coherence using the four- and five-point rating scales. The correlation between two observations is a reliability estimate. Trochim (2006) identified four general classes of reliability estimates: inter-rater; test–retest; parallel forms and internal consistency. Because of the study design employed, test-retest and internal consistency reliability estimates were not calculated; however, they should be considered in future investigations. Inter-rater reliability is when two observers are consistent in their observations. Relevant to the current study, then, inter-rater reliability was estimated when two raters were consistent with the global coherence rating scale scores they assigned to the C-units. We estimated inter-rater reliability for both scales and found that scores were consistent across raters. Possibly, contributing to the high inter-rater agreements may have been the comprehensive training protocols raters completed prior to scoring any samples using the scales. The per cent of agreement for scoring coherence for each C-unit was 98% with the four-point scale and 96% with the five-point scale. These findings support previous results found with the four-point scale (Fergadiotis and Wright 2011, Wright and Capilouto 2012, Wright *et al.* 2010) and the five-point scale (Coelho and Flewellyn 2003, Fergadiotis and Wright 2011, Hough and Barrow 2003, Rogalski *et al.* 2010, van Leer and Turkstra 1999).

For parallel-forms reliability, two forms were administered to the same study participants and the correlation between the two forms was the reliability estimate. This method relies on the assumption that the two forms are equivalent (Trochim 2006). In our study, each participant told two stories. In order to estimate parallel-forms reliability of the rating scales, the assumption that the two stories are equivalent must be met. Several factors were considered to meet this assumption. Both stories were elicited using wordless picture books and study participants were not familiar with the wordless picture books prior to their participation. Further, presumably the study participants produced story-structured discourse samples in response to the discourse elicitation task.

Heath (1986) described stories as a type of narrative that is fictionalized and has a highly structured form. Longacre (1996) characterized narrative discourse by two etic parameters: contingent temporal sequencing and agent orientation. Prior to completing the experimental task, the examiner provided an example of how to tell a story using a different wordless picture book. Further, in a previous study using the same discourse elicitation tasks, we found that participants' storytellings consistently included the two etic parameters (Fergadiotis *et al.* 2011). The stories conveyed by the participants consisted of narratives unfolding over time and involving relevant events and characters. Positive, significant correlations were found between the two stories for both scales. We demonstrated that the wordless picture books elicited stories that are comparable and can be reliably interchanged as different forms to evaluate maintenance of global coherence using the four- or five-point scales. Collectively, results of the different methods employed demonstrated that the rating scales are reliable.

Construct validity

Construct validity refers to how well a concept is translated to reflect its construct; that is, how well operationalization reflects the construct (Trochim 2006). The goal was to determine the feasibility of the measure, the four-point scale, as a valid measure of the construct—global coherence. There are several types of construct validity and it is beyond the scope of the study to consider all types. Rather, we took a systematic approach to evaluating validity of the measure by considering only face and convergent validity initially.

Subsequently, future investigations should consider estimating other validity types; specifically, predictive and concurrent validity of the four-point measure.

To estimate face validity, the operationalization is evaluated 'on its face' as to how well it reflects the construct (Trochim 2006). The four-point scale was developed to align closely with how global coherence is conceptualized—a reflection of the listener's ability to interpret the overall meaning conveyed by the speaker; thus demonstrating that face validity of the measure was adequately estimated. Face validity may be considered a weak method for demonstrating construct validity because it is largely based on subjective assessment; however, it is an appropriate initial step to evaluate the validity of a measure systematically.

Convergent validity entails evaluating how similar the measure (i.e. four-point scale) is to other measures (i.e. five-point scale) that, in theory, it should be. To estimate convergent validity of the four-point scale, all story narrative samples were also scored using the five-point scale. The five-point scale has been extensively used in the literature as a measure of global coherence (Coelho and Flewellyn 2003, Glosser and Deser 1990, 1992, Hough and Barrow 2003, Rogalski *et al.* 2010, van Leer and Turkstra 1999). For both stories, the four-point scale significantly correlated with the five-point scale. Wright *et al.* (2010) also found a strong, positive correlation between the four- and five-point scale for stories told by adults with aphasia. The strong correlations between the two scales and across studies provide evidence for the four-point scale's convergent validity when applied to stories told from wordless picture books. Future investigations should include different discourse elicitation tasks to ascertain that the relationship holds true with different discourse types (i.e. picture description, story retelling and recounts). Further, applying Marini *et al.*'s (2005) method for estimating global coherence to the samples and investigating the relationship among the measures would provide further evidence for the validity of the measure.

Conclusions

This paper achieved the assumption that the four-point global coherence scale is feasible by successfully estimating reliability and validity of the measure. Though only a cohort of middle-aged cognitively healthy participants were included, findings extend previous work (e.g. Fergadiotis and Wright 2011) demonstrating that the measure is a reliable and valid estimate of global coherence ability for stories.

Acknowledgments

References

Agar M, Hobbs JR. Interpreting discourse: coherence and the analysis of ethnographic interview. Discourse Processes. 1982; 5:1–32.

Beukelman, DR.; Mirenda, P. Augmentative and Alternative Communication: Management of Severe Communication Disorders in Children and Adults. 2nd edn. Baltimore, MD: Brookes; 1998.

Christiansen JA. Coherence violations and propositional usage in the narratives of fluent aphasics. Brain and Language. 1995; 51:291–317. [PubMed: 8564473]

Coelho C, Flewellyn L. Longitudinal assessment of coherence in an adult with fluent aphasia: a follow-up study. Aphasiology. 2003; 17:173–182.

Davis, H.; Silverman, S. Hearing and Deafness. New York, NY: Holt, Rinehart & Winston; 1978.

Day, A. Good Dog Carl. New York, NY: Simon & Schuster; 1985.

Fergadiotis, G.; Wright, HH. Global discourse coherence in normal and aphasic discourse. Paper presented at Trends in Experimental Psycholinguistics; Madrid, Spain. 2011.

Fergadiotis G, Wright HH, Capilouto GJ. Productive vocabulary across discourse types. Aphasiology. 2011; 25:1261–1278. [PubMed: 22904592]

Folstein, M.; Folstein, S. Mini-Mental State Examination. 2nd Edition. Lutz, FL: PAR; 2002.

Glosser G, Deser T. Patterns of discourse production among neurological patients with fluent language disorders. Brain and Language. 1990; 40:67–88. [PubMed: 2009448]

Glosser G, Deser T. A comparison of changes in macrolinguistic and microlinguistic aspects of discourse production in normal aging. Journal of Gerontology: Psychological Sciences. 1992; 47:266–272.

Heath SB. Taking a cross-cultural look at narratives. Topics in Language Disorders. 1986; 7:84–95.

Hough MS, Barrow I. Descriptive discourse abilities of traumatic brain-injured adults. Aphasiology. 2003; 17:183–191.

Hughes, D.; McGillivray, L.; Schmidek, M. Guide to Narrative Language: Procedures for Assessment. Eau Claire, WI: Thinking; 1997.

Kintsch W, van Dijk T. Toward a model of text comprehension and production. Psychological Review. 1978; 85:363–394.

Koutsoftas, A.; Wright, HH.; Capilouto, G. Discourse coherence in healthy younger & older adults; Poster presented at the ASHA Convention; New Orleans, LA, USA. 2009.

Krahn, F. The Great Ape. New York, NY: Viking; 1978.

Liles B. Narrative discourse in children with language disorders and children with normal language: a critical review of the literature. Journal of Speech and Hearing Research. 1993; 36:868–882. [PubMed: 8246476]

Loban, W. Language Development: Kindergarten Through Grade Twelve. Urbana, IL: National Council of Teachers of English; 1976. Report No. 18

Longacre, RE. The Grammar of Discourse. 2nd edn. New York, NY: Plenum; 1996.

Marini A, Boewe A, Caltagirone C, Carlomagno S. Age-related differences in the production of textual descriptions. Journal of Psycholinguistic Research. 2005; 34:439–463. [PubMed: 16177935]

McCully, EA. Picnic. New York, NY: HarperCollins; 1984.

Rogalski Y, Altmann LP, Plummer-D' Amato P, Behrman AL, Mariske M. Discourse coherence and cognition after stroke: a dual task study. Journal of Communication Disorders. 2010; 43:212–224. [PubMed: 20219209]

Sheikh, JI.; Yesavage, JA. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. In: Brink, TL., editor. Clinical Gerontology: A Guide to Assessment and Intervention. New York, NY: Haworth; 1986. p. 165-173.

Trochim WM. The Research Methods Knowledge Base (2nd Edition). 2006 (available at: http://www.socialresearchmethods.net/kb).

Van Leer E, Turkstra L. The effect of elicitation task on discourse coherence and cohesion on adolescents with brain injury. Journal of Communication Disorders. 1999; 32:327–349. [PubMed: 10498013]

Wright HH, Capilouto GJ. Manipulating task instructions to change narrative discourse performance. Aphasiology. 2009; 23:1295–1308.

Wright HH, Capilouto GJ. Considering a multilevel approach to understanding maintenance of global coherence in adults with aphasia. Aphasiology. 2012; 26:656–672. [PubMed: 22962509]

Wright H, Fergadiotis G, Koutsoftas A, Capilouto G. Coherence in stories told by adults with aphasia. Procedia Social and Behavioral Sciences. 2010; 6:111–112.

What this paper adds

Global coherence reflects the thematic unity of discourse. Several measures have been used in the literature to quantify maintenance of global coherence. Further, maintenance of global coherence may be impaired in clinical populations, thus warranting reliable and valid measures of global coherence ability. Our goal was to determine the feasibility of a global coherence scale by estimating its reliability and validity. We measured maintenance of global coherence in stories told by 50 cognitively healthy adults. Findings from the study indicate good measurement reliability and construct validity for the global coherence measure.

Table 1

Study participants' demographic data including means, standard deviations (SD) and ranges

	Age (years)	Education (years)	MMSE[a]
Mean	47.22	15.60	51.68
SD	6.44	2.48	6.19
Range	28–59	12–20	32–61

Note:

[a]Mini Mental Status Examination *t*-score; study inclusion criteria was an MMSE *t*-score of 30 or greater.

Table 2

Scoring criteria for a four-point global coherence rating scale

Score	Criteria
4	The utterance is overtly related to the stimulus as defined by the mention of actors, actions and/or objects present in the stimulus which are of significant importance to the *main details of the stimulus*. In the case of procedural descriptions and reactions when a designated topic acts as the stimulus, overt relation is defined by the provision of substantive information related to the topic so that no inference is required by the listener
3	The utterance is related to the stimulus or designated topic, but with some inclusion of suppositional or tangential information that is relevant to the *main details of the stimulus*; or substantive information is not provided so that the topic must be inferred from the statement. In recounts[a], appropriate elaborations that are not essential but are related to the main topic are given a score of 3
2	The utterance is only remotely related to the stimulus or topic, with possible inclusion of inappropriate egocentric information; it may include tangential information or reference some element of the stimulus that is regarded as non-critical
1	The utterance is entirely unrelated to the stimulus or topic; it may be a comment on the discourse or tangential information is solely used

Note:

[a]Recounts are verbal reiterations of an event, e.g. telling about a recent vacation.

Table 3

Mean (standard deviation) global coherence scores for four- and five-point scales across participants and grouped by gender

Coherence type	Entire sample (*N* = 50)	Females (*n* = 25)	Males (*n* = 25)
Five-point scale			
Picnic	4.81 (0.25)	4.91 (0.08)	4.72 (0.32)
Good Dog Carl	4.81 (0.30)	4.91 (0.10)	4.72 (0.39)
Four-point scale			
Picnic	3.93 (0.13)	3.97 (0.04)	3.88 (0.14)
Good Dog Carl	3.93 (0.09)	3.94 (0.06)	3.92 (0.12)

Table 4

Pearson correlation coefficients among the four- and five-point scales across stories

	Five-point *Good Dog Carl*	Five-point *Good Dog Carl*	Five-point *Picnic*	Five-point *Picnic*
Five-point *Good Dog Carl*	1.0			
Four-point *Good Dog Carl*	0.66[*]	1.0		
Five-point *Picnic*	0.66[*]	–	1.0	
Four-point *Picnic*	–	0.71[*]	0.79[*]	1.0

Note:

[*] Significant at $p < 0.0001$.