Spring 2020

Narrative Discourse in Traumatic Brain Injury: Examining the Validity and Clinical Utility of Main Concept, Sequencing, and Story Grammar (MSSG) Analyses

Shannon M. Duffy
University of New Hampshire, Durham

Narrative Discourse in Traumatic Brain Injury: Examining the Validity and Clinical Utility of Main Concept, Sequencing, and Story Grammar (MSSG) Analyses

By

Shannon Duffy

Communication Sciences and Disorders B.S., University of Rhode Island, 2018

THESIS

Submitted to the University of New Hampshire

in Partial Fulfillment of

the Requirement for the Degree of

Master of Science

in

Communication Sciences and Disorders

May, 2020

This thesis was examined and approved in partial fulfillment of the requirements for the degree of Master of Science in Communication Sciences and Disorders by:

_____

Kathryn J. Greenslade, Ph.D., CCC-SLP, Assistant Professor in Communication Sciences & Disorders

_____

Amy E. Ramage, Ph.D., CCC-SLP, Assistant Professor in Communication Sciences & Disorders

_____

Don Robin, CCC-SLP, Professor in Communication Sciences & Disorders

On April 23, 2020

Original approval signatures are on file with the University of New Hampshire Graduate School.

TABLE OF CONTENTS

LIST OF TABLES....iv

LIST OF FIGURES....v

ABSTRACT....vi

CHAPTER PAGE

INTRODUCTION....1

II. METHODS....11

III. RESULTS....17

IV. DISCUSSION....22

V. CONCLUSIONS....29

REFERENCES....30

LIST OF TABLES

1. Table 1: A comparison of existing narrative discourse analysis tools.

2. Table 2: Mean age and education level and sex distributions for the 78 transcripts selected from the TalkBank database.

3. Table 3: Frequency of age and post-traumatic amnesia length and sex distributions for the 21 consistently poor narratives.

LIST OF FIGURES

1. Figure 1: Paneled histograms of participants with TBI versus NHC on MC total scores, sequencing total scores, and MC + sequencing total scores.

2. Figure 2: Paneled histograms of participants with TBI versus NHC on the two story grammar measures: episodic complexity and total episodic components.

3. Figure 3: Scatterplot analysis comparing MC + sequencing total scores to total episodic components scores for the TBI and NHC groups.

ABSTRACT

Background & Aims: Narrative discourse, or storytelling, is critical to assess in adults with traumatic brain injury (TBI) as many of them present with deficits in accuracy, completeness, and logical sequencing of story content as well as in story grammar organization. Richardson and Dalton (2016) and Greenslade et al. (in submission) created Main Concept, Sequencing, and Story Grammar (MSSG) to analyze these variables in *Cinderella* narratives, with preliminary data revealing age-related declines in neurologically healthy control (NHC) performance. To extend these findings, the present study sought to evaluate MSSG's clinical utility in identifying narrative deficits in adults with TBI. Research questions asked whether 1) adults with TBI would receive poorer scores across analyses compared to NHCs, and 2) more adults with TBI would show consistently poor or discrepant performance across the accuracy, completeness, and sequencing of story content and story grammar organization.

Methods: Seventy-six *Cinderella* narrative transcripts were downloaded from the online database, TalkBank, with equal numbers of participants with TBI and NHCs. MSSG analyses were applied to examine five measures: 1) main concept (MC; presence, accuracy, and completeness of story content), 2) sequencing, 3) MC + sequencing, and 4 & 5) two story grammar measures: episodic complexity and total episodic components.

Outcomes/Results: MSSG analyses detected statistically significant between-group differences across all measures, documenting how adults with TBI told less accurate, complete, and logically sequenced story content while including fewer complex episodes and fewer episodic

components. More adults with TBI demonstrated consistently poor performance across MC + sequencing and total episodic components as compared to NHC.

Conclusions: The present study provides preliminary construct validity for using MSSG analyses to detect differences in the narrative discourse of adults with TBI. Results revealed that half of the TBI sample demonstrated consistently poor narrative discourse performance, producing less accurate, complete, well-sequenced, and complex narratives. These data provide initial evidence supporting the use of MSSG to quantify the narratives of adults with TBI in research and clinical settings.

INTRODUCTION

Life is a series of organic and breathing moments; moments that are stored in an individual's memory for years. When someone wants to share the plot of his or her favorite movie, a funny childhood tale, or a most miserable day at work, these memories become living stories. Storytelling allows people to actively participate in life by partaking in conversations, establishing meaningful relationships, and expressing feelings and emotions. The ability to successfully convey a story is a primary part of communication and living a fulfilling life. Unfortunately, this ability may be limited in some individuals.

Individuals with cognitive communication disorders, such as those resulting from a traumatic brain injury (TBI), have difficulty telling a story, or narrative discourse. These difficulties are characterized as macrostructural deficits. That is, the overall narrative structure and meaning conveyed in stories is less streamlined or cohesive, lacking clear cause-effect relationships – making it difficult for a listener to understand the main gist of the story (Coelho, Biles, & Duffy, 1995; Coelho, Ylvisaker, & Turkstra, 2005; Marini, Galetto, et al., 2011; Peach, 2013; Power et al., 2020). Narratives of adults with TBI have been characterized as disorganized or lacking a logical flow or sequence of events (Coelho, Biles, & Duffy, 1995; Marini, Galetto, et. al, 2011; Matsuoka & Yamasoto, 2012; Peach, 2013). Specifically, deficits have been found in *story grammar*, a common framework for building a cohesive narrative which involves telling complete story episodes (Coelho, Biles, & Duffy, 1995; Coelho, Ylvisaker, & Turkstra, 2005; Power et. al, 2020). These deficits in narrative discourse depend heavily upon the area of damage and the effects of that damage on cognition, memory and executive function, which then

can affect the way these individuals communicate, connect with others, and feel fulfilled (Jorgensen & Togher, 2009).

Further, the stories of adults with TBI have been found to have significant microstructural (sentence level) deficits, including the use of less specific or inaccurate language, assuming the listener knows what is meant or being referred to without it being directly stated (Biddle et al., 1996; Carlomagno et al., 2011; Ghayoumi, 2014; Marini, Galetto, et al., 2011). These narratives have been described as ambiguous or inefficient, and lacking complete or accurate story content (Carlomagno et al., 2011; Davis & Coelho, 2004; Elbourn et al., 2019; Hartley & Jensen., 1991; Jorgensen & Togher, 2009; Lê et al., 2011). Additionally, Biddle, McCabe, and Bliss (1996) found that adults with TBI may also produce more incomplete propositions, or statements, resulting in absent story content. These findings suggest that the stories are lacking in content accuracy and completeness, two measures that should aid in listener tracking and understanding of a story (Davis & Coelho, 2004, Richardson & Dalton, 2019).

A number of tools exist to measure narrative discourse in ways that account for these deficits. However, there is not one measure that accurately accounts for the lack of logical sequencing, accuracy, and completeness of story content as well as lack of adherence to story grammar commonly seen in adults with TBI. Greenslade, Stuart, Richardson, Dalton, and Ramage (in submission) created Main Concept, Sequencing, and Story Grammar (MSSG) analyses, a new narrative analysis tool to quantify these deficits. In this study, we will use MSSG to compare the content, sequencing, and story grammar of complex *Cinderella* narratives told by adults with TBI and neurologically healthy controls (NHC) to provide evidence of the tool's clinical utility and efficiency.

Discourse Analysis Tools

Numerous tools aim to analyze narratives in an objective and exhaustive manner. These narrative analysis methods include assessments of story grammar and the sequencing, accuracy, and completeness of story content.

Stein and Glenn's (1979) story grammar framework proposed that narratives should follow key rules and contain certain components to successfully convey a story. Stein and Glenn's (1975) schema was based upon a setting, one or more episodes, and a conclusion or coda. The setting serves a twofold purpose: describing the time, society, and physical context and introducing the main characters (Stein & Glenn, 1975). These characteristics and states describe long-term attributes of the story's places and characters, establishing a status quo that may be disrupted by the story's unfolding events. For example, in Cinderella, the setting describes how a man whose wife died remarries a woman with two daughters. This setting creates a context which allows for the rest of the story to unfold in a system of episodes.

An episode is the fundamental building block of a story which includes events that affect a character, how that character responds internally and externally, and the results of those responses (Roth & Spekman, 1986; Stein & Glenn, 1975). A story may include only a single episode, or it may require many more depending on the story's complexity. At least two of the following three components must be included within an episode for it to be considered complex: 1) an initiating event (an occurrence that spurs the protagonist into action to accomplish a goal), 2) an attempt (the action(s) the protagonist takes to accomplish that goal), and 3) a direct consequence (the direct result of that attempt). For example, in one episode, Cinderella arrives at the ball (initiating event). Spurred on by her arrival, Cinderella dances with the prince (attempt). As a result of this dance, the prince then falls in love with her (direct consequence). This is a

logically sequenced, complex episode. Finally, stories end with a conclusion or coda, which provides the resolution of the story. For example, Cinderella and the prince get married and live happily ever after.

Stein and Glenn (1975) found that story grammar was effective in identifying types of information included in stories and parsing that information into meaningful units rather than using propositions (rough equivalents to a basic sentence), as prior organizational analyses had done. However, they failed to provide validity evidence for its use. Using a modified version of Stein and Glenn's story grammar, Roth and Spekman (1986) found their analysis at the episode level was more representative of the quality of a story's organization than analysis at the proposition level. Present analyses will be conducted at the episode level, following Stein and Glenn (1975).

The Story Goodness Index (SGI; Lê et al., 2011) is another variation of narrative analysis. SGI combines and quantifies story completeness and story grammar to measure narrative macrostructure with the intention of differentiating discourse deficits in those with TBI vs. NHCs (e.g. lack of essential content). SGI's story grammar measure is a ratio of t-units within episodes to total t-units in the narrative, where a t-unit refers to a main clause + any attached subordinate clauses (Coelho, 2002; Hunt, 1965). SGI accounts for completeness based on an exhaustive list of vital elements (events and characters) that were mentioned in 80% or more of NHC narratives. Lê and colleagues (2011) initially created this list with the story of *Old McDonald Had an Apartment* (Barrett, 1998), finding five critical components based on the stories of forty-six NHCs. The completeness score was determined by adding up the total number of critical components each participant included in their story.

Research shows that SGI presents a more accurate picture of narrative discourse

performance than measures of story grammar or content completeness alone. Lê and colleagues' (2011) story grammar measure showed that adults with TBI produced a significantly larger percentage of t-units outside of episodes than NHCs, indicating that adults with TBI provided more unnecessary information thus telling their stories less efficiently. Additionally, adults with TBI produced significantly fewer vital elements on the completeness measure as compared to NHCs. Overall, Lê et al. (2011) found that SGI was a reliable and sensitive tool for identifying narrative discourse deficits in adults with TBI. Further, Lindsey et al. (2018) presented evidence of SGI's validity in identifying differences in the narrative discourse performance of adults with TBI and NHCs. However, their findings also showed that there was a significant group difference for story grammar but not story completeness, attributed to the use of a less complex story and thus limiting SGI's clinical utility. Research on SGI suggests that more complicated stories (i.e., ones that require storytellers to make inferences and understand figurative language) may detect group differences between TBI and NHC groups more consistently (Lê et al., 2011; Lindsey et al., 2018). Thus, the present study selected the story *Cinderella* given the complexity of its character motivations, internal responses, and reactions, as well as its requirement for multiple complex episodes.

SGI findings emphasize that to efficiently measure narrative discourse in adults with TBI, a multifaceted tool must be used to capture an accurate and exhaustive image of each individual's skills. While SGI proves to be a more exhaustive tool in measuring both completeness and story grammar, evidence for the construct validity of both measures is inconsistent. Further, SGI does not account for the accuracy of included story content.

Finally, Nicholas and Brookshire (1995) created main concept (MC) analysis to capture the presence, accuracy, and completeness of a story's main content, thus identifying an

individual's ability to communicate the foremost gist of the story. *Accurate* is defined as containing true and correct content, while *complete* is defined as containing all of the essential elements of an MC. An MC is a proposition that conveys a central idea to a story and usually, but not always, includes 1) a main verb 2) a subject and 3) an object when necessary (Greenslade et al., in submission). Main concept checklists for different elicitation tasks are created based on the essential information included by NHCs.

Based on 92 healthy control transcripts from the online database AphasiaBank, Richardson and Dalton (2016) developed an MC checklist for *Cinderella*, a complex tale both familiar to the general population and commonly used in evaluating the language of individuals with aphasia, TBI, and/or other acquired communication disorders. To create the checklist, Richardson and Dalton (2016) identified relevant concepts (i.e., candidate main concepts) from the sample of 92 healthy control transcripts. They applied a 33% cut-off threshold, determined by how frequently a concept was produced across transcripts. A relevant concept that was produced in the sample but did not make the 33% cut-off threshold was eliminated from subsequent main concept analyses.

Thirty-four main concepts were established for *Cinderella*. Each MC had two to four essential elements, and each MC was represented by a generalized production (e.g., "[1]The fairy godmother [2]makes [3]{items} turn into {items}", where numbering indicates essential elements) and a list of some potential alternative productions (e.g., "[1]The fairy godmother [2]changes [3]mice into horses"). To analyze a story for MCs, the participant's utterances are first matched with the corresponding MC, and then those utterances are judged based on whether each essential element is communicated inaccurately (e.g., "fairy stepmother" instead of "fairy godmother) or incompletely (e.g., "the fairy godmother makes horses" – omits the item that turns into "horses").

Main concepts are scored following a coding system: AC for accurate and complete, AI for accurate but incomplete, IC for inaccurate but complete, II for inaccurate and incomplete, and AB for absent (Richardson & Dalton, 2016). Richardson and Dalton (2016) explain that an individual with TBI may have just as many ABs or ACs as a NHC but may greatly differ in the number and combinations of accuracy and completeness. Quantifying the MCA codes aids in this discrimination.

Evidence supports the reliability and validity of MC analysis for identifying discourse deficits in individuals who are aging and in those with aphasia and cognitive communication disorders. First, acceptable intra- and inter-rater reliability has been found for main concept analysis (Elbourn et al., 2019; Nicholas & Brookshire, 1995; Richardson & Dalton, 2016). As evidence of the tool's validity in identifying discourse deficits, Richardson and Dalton (2016) identified that younger participants produced more present and accurate/complete content. Nicholas and Brookshire (1995) and Richardson and Dalton (2019) found that connected speech of individuals with aphasia had more absent, incomplete, and inaccurate main concepts than NHCs. Finally, Elbourn and colleagues (2019) found that 57 participants with TBI of varying severity produced more inaccurate and incomplete main concepts and omitted more main concepts than NHCs. Interestingly, a subset of individuals with TBI performed within normal limits according to MC analysis control data but still showed small differences in accuracy and completeness from their matched controls. This suggests that although people with TBI may perform within normal limits, their communication may lack the clarity and effectiveness of neurologically healthy individuals. Thus, MC analysis appears to be a clinically useful tool for identifying discourse deficits in aging and clinical populations.

Similar to Richardson and Dalton's (2016) process, Stark (2010) analyzed oral retellings

of *Cinderella* told by healthy controls by reducing the story structure to 41 potential propositions, comparable to main concepts. Once a participant's utterances were matched to the intended meaning of the propositions, they were scored as being explicitly produced, implicitly produced, or omitted. These propositions were then placed into a story superstructure composed of a setting, episodes, and a conclusion similar to that of story grammar. However, Stark's analyses primarily focused on the presence and type of propositions (explicit versus implicit) and did not break the episodes down further into initiating events, attempts, and direct consequences. Thus, these measures did not capture how complex those episodes were.

Table 1. A comparison of existing narrative discourse analysis tools.

Analysis Tool	Benefits	Drawbacks
Story Grammar (SG)	• Analyzes stories based on episodes • Analyzes story complexity through identification of episodic components (initiating events, attempts, and direct consequences)	• Lacks measures of accuracy, completeness, and sequencing of story content
Story Goodness Index (SGI)	• Measures story grammar organization • Measures story completeness	• Lacks measure of sequencing • Lacks consistent evidence to support its validity and clinical utility
Main Concept Analysis (MCA)	• Measures presence of story content • Measures accuracy of story content • Measures completeness of story content	• Lacks measures of story grammar organization • Lacks measures of sequencing of story content

Greenslade et al. (in submission) created a secondary analysis tool, Main Concept, Sequencing, and Story Grammar (MSSG), based on the work of Richardson and Dalton (2016) to provide preliminary quantitative data on the macrostructure of Cinderella narratives in NHCs. This tool extended Richardson and Dalton's MC analysis to include measures of logical

sequencing and story grammar, using the same 92 participant transcripts from AphasiaBank. Building on the work of Stark (2010), Stein and Glenn's (1975) story grammar framework was applied to explore the complexity of story episodes. MSSG calculates five total measures: 1) MC, 2) sequencing, 3) MC + sequencing, 4) episodic complexity, and 5) total episodic components.

MC total score is calculated according to Richardson and Dalton's (2016) MC analysis protocol by mapping transcript utterances to the list of 34 MCs. Based on the presence, accuracy, and completeness of that utterance, each MC is given a code of AC, AI, IC, or AB. The newly developed sequencing score is a measure of macrostructure that describes how well the main concepts of a participant's narrative follow a logical causal/effect arrangement throughout the story. This measure uses line numbers from the participant's transcript to determine if the MCs are organized in a logical sequence, roughly following the order outlined by Richardson and Dalton (2016; exceptions described in Greenslade et al., in submission). MC + sequencing total is a sum of a participant's MC and sequencing total scores and is used to quantify the overall quality of story content.

To calculate MSSG's episodic complexity and total episodic components scores, story grammar components were mapped onto MCs (Greenslade et al., in submission). All but five MCs map directly onto a single, story grammar component; the remainder require the coder to make a judgment on how the MC functions in the specific participant's narrative. For example, MC 20, "The prince and Cinderella danced around the room," can either function as an attempt or a direct consequence. If the prince first falls in love with Cinderella (MC 21), and they dance as a result, the latter would be coded as a direct consequence. If Cinderella and the prince first dance around the room, and then he falls in love with her, MC 20 would be coded as an attempt.

The measure of episodic complexity quantifies the total number of episodes, out of 5 identified episodes, that include two or more episodic components (initiating events, attempts, and direct consequences). The measure of total episodic components (TEC) quantifies whether the participant narrative includes at least one instance of the three required episodic components (initiating event, attempt, or direct consequence) per episode, for a total of 15 possible components. The higher the episodic complexity and total episodic components scores, the more intricate and cohesive one may expect a story to be.

Greenslade et al. (in submission) found that as participants aged, performance across all measures declined. These findings were consistent with prior research showing that the presence and organization of essential story content declines with age (e.g. Cannizzaro & Coelho, 2013; Marini, Boewe, et al., 2005). Further, the analyses were found to efficiently assess logical sequencing and story grammar use in healthy controls (Greenslade et al., in submission). These multi-level analyses provided preliminary evidence for the use of MSSG to detect declines in narrative discourse performance (Greenslade et al., in submission). It remains to be seen if this newly developed tool can be clinically useful in clinical populations.

The present study's primary aim is to evaluate the clinical utility of Greenslade et al.'s MSSG in identifying deficits that individuals with traumatic brain injury might exhibit when telling *Cinderella*. To do so, we sought to determine whether 1) adults with TBI show clinically significant differences in narrative discourse as compared to neurologically healthy controls, and 2) a larger proportion of adults with TBI show discrepant performance between their story content (accuracy, completeness, and sequencing) as compared to their story grammar component use or show consistently poor performance across measures. To address the first question, we examined between group differences in: a) the accuracy and completeness of story

content as indicated by true and correct information and the inclusion of all the essential elements of MCs; b) logical sequencing as indicated by main concepts that follow in a logical, causal/effect arrangement throughout the story; c) MC + sequencing total combined score d) episodic complexity as indicated by the number of complex episodes in the story; and e) total episodic components as indicated by the total number of episodic components included across the story's five episodes. We hypothesized that the narrative discourse performance of adults with TBI would be comprised of significantly less accurate, complete and logically sequenced story content, and fewer complex episodes with the inclusion of fewer episodic components. To address the second question, we examined whether performance of adults with TBI and NHCs was consistently good, consistently poor, or discrepant across content (MC + sequencing) and story grammar (total episodic components). We hypothesized that more adults with TBI would show a consistently poor performance across all measures. Answering these questions will allow us to document the clinical utility of this tool in identifying strengths and weaknesses in the narrative discourse of adults with TBI.

METHODS

Participants

Transcripts from 38 individuals with a TBI and 38 neurologically healthy controls were retrieved from online TalkBank databases in order to compare performance on the *Cinderella* retell task (see Table 2). Transcripts of neurologically healthy controls were contributed by the Wright, Richardson, Capilouto, and Boyle labs; transcripts of individuals with TBI were contributed by the Togher lab. Although the Togher lab contributed participant data at multiple

time points, only data from the time point 12-month post-injury were used for the current study. The 12-month checkpoint was chosen to assure stability of language performance in the chronic epoch of TBI recovery. Inclusionary criteria included: English as a primary language, at least 20 years of age, and no history of brain injury if neurologically healthy. Participants were matched pairwise based on age and sex, and as a group for years of formal education ($U = 495.000$, $z = -1.892$, $p = .059$).

Table 2. Mean age and education level and sex distributions for the 78 transcripts selected from the TalkBank database.

	N	Age (years)	Sex (F:M)	Education (years)
All	76	38.22	8:68	14.12
TBI	38	39.82	5:33	14.61
NHC	38	38.45	3:35	13.61

All transcripts were collected according to the AphasiaBank protocol (https://aphasia.talkbank.org/). Each participant signed an informed consent form prior to completing the protocol and specifying if their data may be shared and used for research purposes. Participants were first presented with a wordless *Cinderella* picture book to review. All participants confirmed that they had heard *Cinderella* before. When they finished, the examiner retrieved the book, and asked them to tell the best story they could. Participants continued until they concluded the story or made clear they were finished (Elbourn et al., 2019). Transcripts of *Cinderella* stories that had been uploaded to TalkBank were copied into word

processing documents and labeled with coding numbers to ensure that coders were naïve to each participant's diagnostic status.

MSSG Scoring

Transcripts were scored for MC, sequencing, MC + sequencing, episodic complexity, and total episodic components for the *Cinderella* story retell task using the multilevel macrostructural analysis tool, MSSG, created by Greenslade et al. (in submission).

To calculate the main concept total score, transcript utterances were matched to the list of 34 identified main concepts (Richardson & Dalton, 2016; see Appendices 2 and 4 for scoring details). Based on the accuracy and completeness of the matched utterance, each MC received a score up to 3. A score of 0 marked an MC as absent; a score of 1 marked an MC as inaccurate and incomplete (II); a score of 2 marked an MC as either accurate and incomplete (AI) or inaccurate and complete (IC); and finally a score of 3 marked an MC as both accurate and complete (AC). The total score was the sum of item-level scores for each MC and represented how much of the necessary content was present, accurate, and complete. A total of 104 points was possible if every MC was present, accurate, and complete.

The sequencing score was calculated using the participant's transcript line numbers to determine if the MCs followed a logical sequence (see Greenslade et al.'s Appendix A for sequencing scoring rules). Each MC received a score up to 3. A score of 0 is given if the MC is absent from the transcript. A score of 1 is given if the MC is present but marked out of order based on line number and rules. A score of 2 is given for an MC that is out of order but is marked as out of order by the participant themselves. For example, if a participant says "oh, I forgot to say that before she left for the ball, the fairy godmother told her she had to be home by

midnight," that MC would be given a score of 2. Rules are put into place for if participants noticed that they made a mistake and signals their error. Finally, an MC is given a score of 3 if it is present and in the correct order. The sequencing total score is the sum of the item-level sequencing scores assigned to each MC. A total of 102 points is possible if every MC is present and correctly sequenced. For the MC + sequencing measure, the MC total and sequencing total scores are summed for 204 total possible points.

To calculate MSSG's episodic complexity and total episodic components scores, story grammar components first were mapped onto MCs (see Greenslade et al.'s Appendix B for examples of story grammar coding/scoring). MSSG's episodic complexity was determined by calculating the number of complex episodes in the narrative. An episode that included two or more of the required episodic components (IE, A, DC) was assigned a score of 1. Episodes that have one or less episodic components received a score of 0. The maximum episodic complexity score was 5 (1 for each of the 5 episodes of *Cinderella*).

Total episodic components (TEC) was determined by calculating the number of episodic components (initiating event, attempt, or direct consequence) that occurred at least once in each episode. An episode that had at least one IE, one A, and one DC received a maximum of 3 points. If an episode included one IE, one A, but zero DCs, would receive a score of 2. When a single episodic component was included at least once in an episode, it received a score of 1. The maximum total episodic components score was 15 (i.e. one point per episodic component, with 3 possible components per episode across 5 episodes; Greenslade et al., in submission).

To ensure scoring reliability, two Communication Sciences and Disorders graduate students completed a training for assigning sequencing, total episodic components, and episodic complexity scores. Training was completed when the pair independently scored and reached at

least 80% reliability for point-to-point agreement and .7 for Cohen's Kappa for at least four of five consecutive practice samples across all scores. Practice samples were not included in the present data set. One scorer scored all 78 transcripts; 20% (18) were scored by a second scorer concurrently. Reliability meetings occurred throughout the data reduction to prevent scorer drift and to ensure that rules were being consistently and appropriately applied. Point-to point reliability for main concept, sequencing, episodic complexity, and total episodic components was 85.478%, 91.728%, 82.500%, and 92.121%, respectively. Corresponding Cohen's Kappa values were .853, .916, .808, and .919, respectively.

Data Analysis

IBM's Statistical Package for the Social Sciences (SPSS) version 26 was used to generate descriptive statistics and to perform statistical analyses to determine the clinical utility of Greenslade et al.'s MSSG in identifying deficits that adults with traumatic brain injury might exhibit when telling *Cinderella*. To determine whether MSSG detected different narrative performance in adults with TBI as compared to NHCs, we examined between-group differences on MC, sequencing, MC + sequencing, episodic complexity, and total episodic components scores. Each measure was assessed for the normality of distributions and outliers in the TBI and NHC groups. Shapiro-Wilk's test detected non-normal distributions for the NHC group across all variables (p's $\leq .013$) and for the TBI group for episodic complexity ($p = .002$). Inspection of box plots revealed outliers for episodic complexity and total episodic components. Thus, because the dependent variables did not meet all of the assumptions for an independent t-test, nonparametric Mann-Whitney U tests were run to determine if there were differences in MC, sequencing, MC + sequencing, episodic complexity, and total episodic components total scores

between adults with TBI and NHCs. Statistical significance for the Mann-Whitney *U* tests was set at $\alpha = .05$.

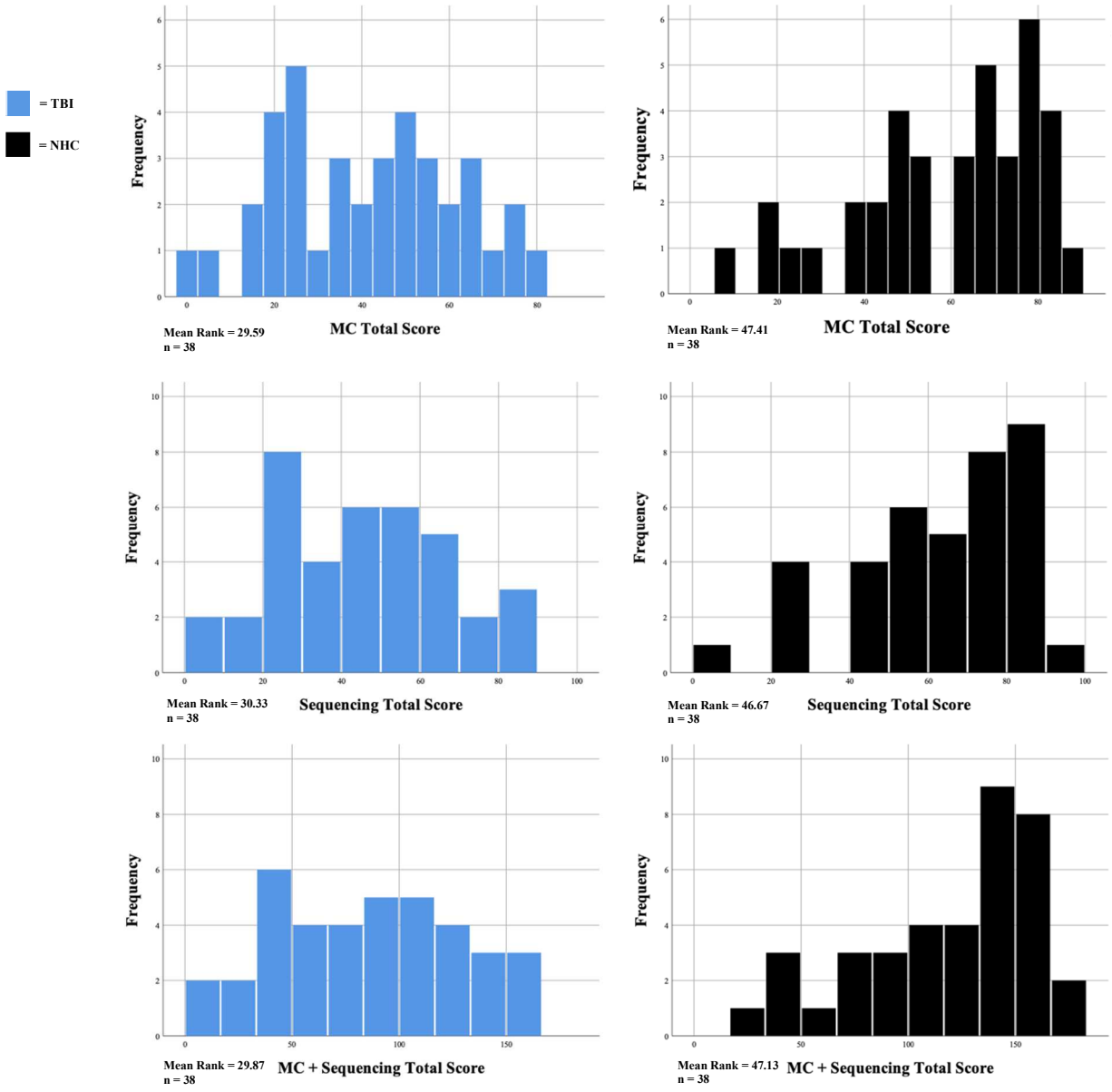
Additionally, we used visual analyses and chi-squared tests to determine whether a larger proportion of adults with TBI show either poor performance across measures or discrepant performance between their story content (accuracy, completeness, and sequencing) as compared to their story grammar component use. The *z*-score cut-offs for healthy controls from Greenslade et al. (in submission) at one and two standard deviations below the mean for MC + sequencing and total episodic components were applied to the current sample to determine whether a larger proportion of adults with TBI demonstrated consistently poor performance or discrepant performance across measures. By comparing current participants' scores to z-scores from the prior sample, each participant was identified as having consistently good accuracy, completeness, and logical sequencing of content and story grammar component use; poor story content in the presence of good story grammar component use; good story content in the presence of poor story grammar component use; or poor content and story grammar component use. Then, chi-squared tests were run to compare the proportion of adults with TBI versus NHCs who demonstrated either consistently poor or discrepant performance across content and story grammar use.

RESULTS

Accuracy, Completeness, and Sequencing of Content

Figure 1 shows paneled histograms of participants with TBI versus NHC on MC total scores, sequencing scores, and MC + sequencing scores. MC total scores for individuals with TBI (mean rank = 29.59) were significantly lower than for NHC (mean rank = 47.41), U = 383.500, Z = -3.517, $p < .001$. Similarly, sequencing total scores for adults with TBI (mean rank = 30.33) were significantly lower than for NHCs (mean rank = 46.67), U = 411.500, Z = -3.227, p = .001. Furthermore, MC + sequencing total scores for individuals with TBI (mean rank = 29.87) were statistically significantly lower than for NHCs (mean rank = 47.13), U = 394.000, Z = -3.408, p = .001.

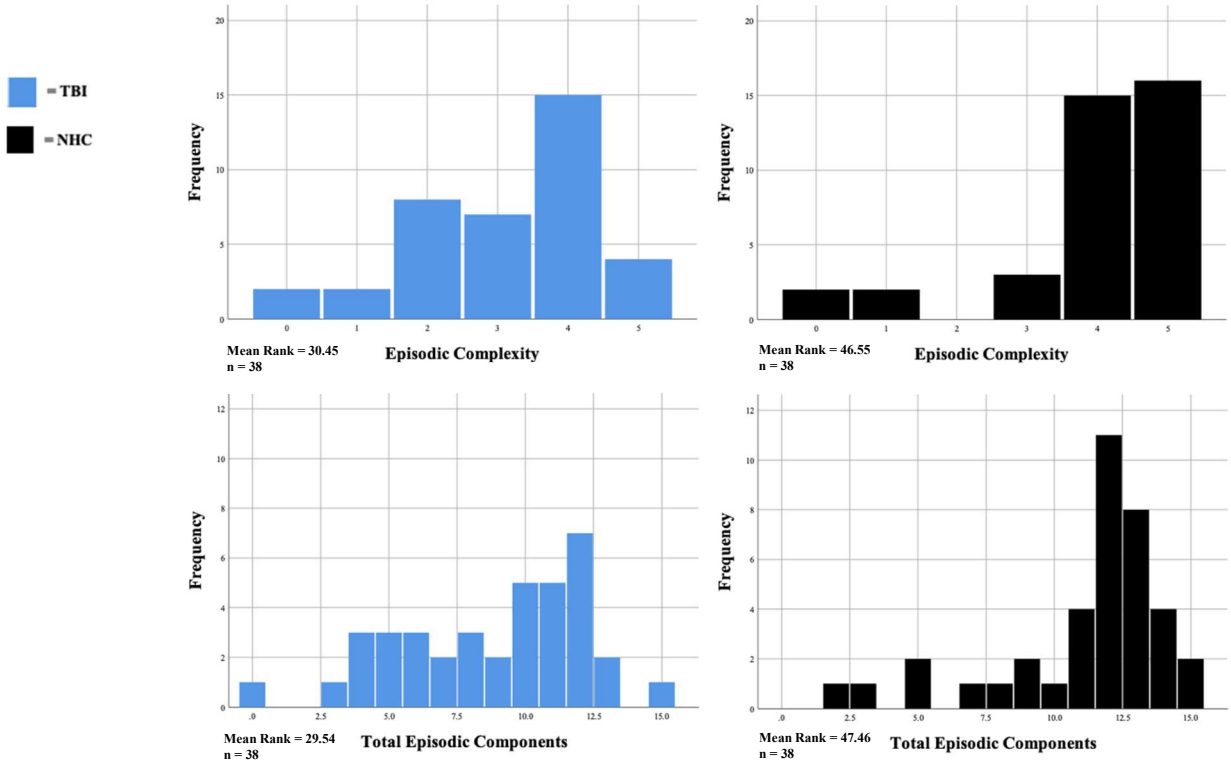
Figure 1. Paneled histograms of participants with TBI versus NHC on MC total scores, sequencing total scores, and MC + sequencing total scores. Histograms in blue represent the performance of participants with TBI, while the histograms in black represent the performance of NHCs. While the distributions for both adults with TBI and NHCs feature a single peak and are asymmetric, the distributions of adults with TBI are negatively skewed as compared to the positive skew of the NHCs. This makes sense as we hypothesized that adults with TBI would have a poorer performance.



Story Grammar Organization

Figure 2 shows paneled histograms of participants with TBI versus NHC on episodic complexity and total episodic components measures. Distributions of episodic complexity and total episodic components scores were not similar, as assessed by visual inspection. Episodic complexity scores for adults with TBI (mean rank = 30.45) were statistically significantly lower than for NHCs (mean rank = 46.55), $U = 416.000$, $Z = -3.320$, $p = .001$. Total episodic components scores for individuals with TBI (mean rank = 29.54) were statistically significantly lower than for NHCs (mean rank = 47.46), $U = 381.500$, $Z = -3.567$, $p < .001$.

Figure 2. Paneled histograms of participants with TBI versus NHC on the two story grammar measures: episodic complexity and total episodic components. Poorer scores of adults with TBI can be visually detected. Although both groups are positively skewed (as compared to Figure 1), distributions are not similar, and adults with TBI had more frequent low scores.



Discrepant or consistently poor performance in TBI versus NHC

Participant MC + sequencing total scores were plotted against corresponding total episodic components scores to determine whether participants who had difficulty telling logically sequenced, complete, and accurate story content also had difficulty with story grammar organization. MC + sequencing scores were plotted on the x-axis with total episodic components plotted on the y-axis (see Figure 3). Solid vertical and horizontal lines represent z-scores that are 1 SD below the means of neurologically healthy controls established by Greenslade et al. (in submission) (MC + Sequencing = 82, $z = -1.043$; $8 <$ total episodic components < 9, [$-.886 < z < -1.247$]. Dotted vertical and horizontal lines represent z-scores that are 2 SD below the means of Greenslade et al.'s controls (MC + Sequencing = 46, $z = -2.002$; $5 <$ total episodic components < 6, [$-1.969 < z < -2.330$]).

Visual inspection of Figure 1 revealed a linear relationship between MC + sequencing and total episodic components as expected. This indicated that individuals who produced more logically sequenced, complete, and accurate story content typically generated more story grammar episodic components. The bottom left quadrant of the plot identified 21 narratives, representing narratives with poor content (≥ 1 SD below on MC + Sequencing) and few episodic components (≥ 1 SD below on total episodic components). Of the 21 consistently poor narratives, the majority was produced by adults with TBI who were male and experienced a post-traumatic amnesia (PTA) of 60+ days. Please reference Table 3 for further demographic results. This lower quadrant illustrates an overall reduction in macrostructural narrative quality in some adults after TBI, more specifically males and those who experienced an extended period of post-traumatic amnesia. The lower right quadrant identified one adult with TBI who used fewer episodic components than expected, despite average accuracy, completeness, and logical

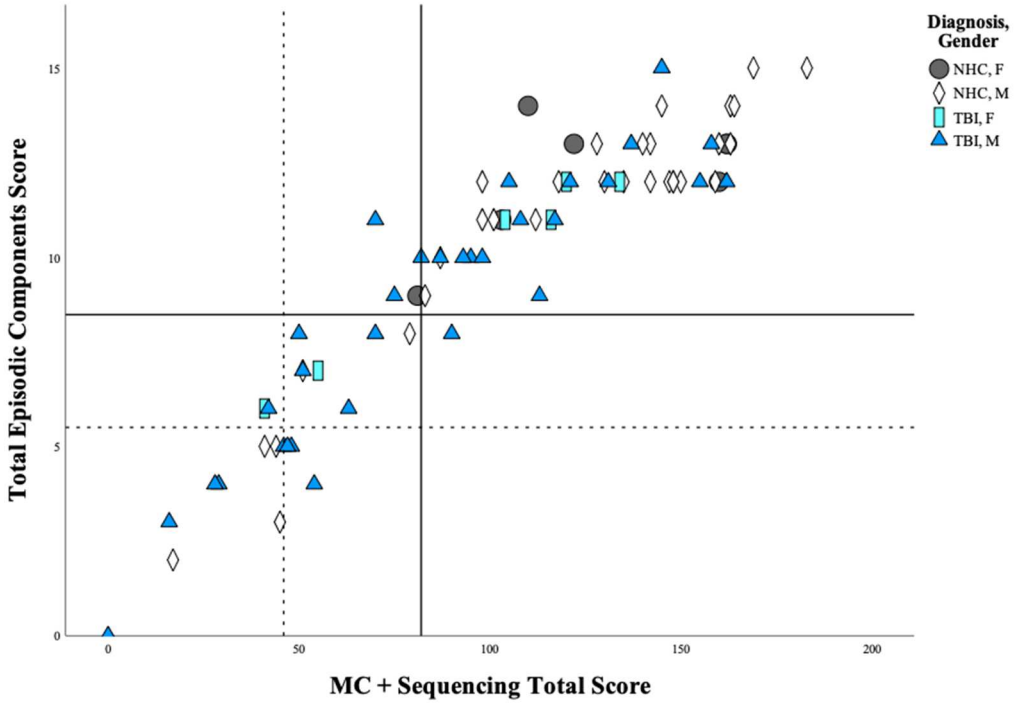
sequencing. Finally, in the upper left quadrant, two adults with TBI and one NHC were identified who had below average accuracy, completeness, and logical sequencing while maintaining an average number of episodic components.

A chi-square test for association was conducted between diagnostic status and performance that was at least 1 SD below the mean on MC + sequencing and/or total episodic components. All expected cell frequencies were greater than five. There was a statistically significant association between diagnostic status and narrative performance, with adults with TBI being more likely to demonstrate poor narrative performance, $X^2(1) = 8.418$, $p = .004$. This association was moderately strong, $\varphi = .333$, p=.004.

Table 3. Frequency of age and post-traumatic amnesia length and sex distributions for the 21 consistently poor narratives.

	N	18 to 39 Age Bracket (Frequency)	40 to 59 Age Bracket (Frequency)	60+ Age Bracket (Frequency)	Sex (F:M)	PTA 60+ Days (Frequency)
All	21	12	7	2	2:19	--
TBI	15	8	5	2	2:13	9
NHC	6	4	2	0	0:6	--

Figure 3. Scatterplot analysis comparing MC + sequencing total score to total episodic components score for the TBI and NHC groups.



Note: Solid lines indicate 1*SD* below the means of NHCs established by Greenslade et al. (in submission); dotted lines indicate 2*SD*s below the means. The solid lines divide the graph into four quadrants. Discrepant performances represented by top-left and bottom-right quadrants.

DISCUSSION

The aim of the present study was to evaluate the clinical utility of Greenslade et al.'s MSSG in identifying deficits that adults with traumatic brain injury might exhibit when telling *Cinderella* as compared to neurologically healthy controls. Across groups, five measures of narrative discourse were investigated: MC, sequencing, MC+ sequencing, episodic complexity, and total episodic components. Findings will be discussed below.

Statistically significant differences were found between groups for all measures, demonstrating that on average, adults with TBI told less logically sequenced, accurate, and complete stories while incorporating less complex episodes with fewer episodic components. As anticipated, these differences across measures aid in capturing the less cohesive storytelling of adults with TBI and provide construct validity evidence supporting the use of scores to detect diagnosis-related differences in narrative discourse performance. Further, when MC + sequencing total scores and total episodic components scores were plotted against each other, consistently good or poor performance was revealed for the majority of participants (n = 72). It is important to note that when these two measures were plotted against each other, discrepant performances were identified predominantly in adults with TBI, indicating that MSSG may not only have the ability to detect clinically significant differences in the narrative discourse of adults with TBI but to also allow clinicians to identify more nuanced deficits that may help specify future directions for treatment.

Diagnosis-Related Differences

Consistent with expectations, results of nonparametric analyses revealed that adults with TBI more poorly than controls across all five narrative discourse measures. Adults with TBI generally produced fewer MCs and fewer MCs that were logically sequenced, complete, and accurate, replicating prior findings of deficits in the production of story content following a TBI. Descriptively, these stories generally were significantly decreased in story length, used vague language, and lacked considerable informational content, as evidenced both by production of extraneous, unrelated content or absent content. Thus, the inaccurate information in narratives of those with TBI may be due to a variety of factors, including lack of specificity or word retrieval

difficulties. The large quantity of absent information in transcripts greatly contributed to an overall shorter story length produced by adults with TBI. These results are consistent with findings of prior studies that examined deficits in story content of adults with TBI (e.g. Carlomagno et al., 2011; Davis & Coelho, 2004; Stout et al., 2000).

In terms of story organization, the narratives of adults with TBI consisted of fewer episodes and fewer total episodic components (initiating event, attempt, and direct consequence), resulting in less complex story organization. For example, individuals with TBI commonly included only one episodic component for each episode or only included episodic components for certain episodes, resulting in story "gaps" or less exhaustive narratives. Of the 15 adults with TBI who had consistently poor performance across both MC + sequencing and total episodic components total scores, 14 were missing at least one full episode (zero episodic components). Nine out of 15 were missing at least two full episodes and had at least one other episode with only one out of three components. These results were consistent with the findings of previous research which indicating poor story organization in individuals with TBI (Coelho, Liles, & Duffy, 1995; Coelho, 2002; Power et al., 2020). Results are also consistent with the findings of Lê and colleagues (2011) and Lindsey and colleagues (2018) in which adults with TBI produced fewer utterances in an episodic format and used fewer episodic components (Lê et al., 2011; Lindsey et al., 2018).

It is of note that while our findings were consistent with Lindsey et al. (2018) for story grammar organization, Lindsey and colleagues did not detect between-group differences on their completeness measure. They found that more participants with TBI produced less organized stories while having adequate story content, attributing these results to the simplicity of the story *The Bear and the Fly (*a more straightforward story requiring no inferencing; Lindsey et al.,

2018). With this in mind, the present study's use of *Cinderella* for its complexity may have been a critical factor in our detection of differences in both completeness and organization between adults with TBI and NHCs. This indicates that the use of a complex, familiar story which provides more opportunities for high-level thinking (e.g. determining character motivations, establishing character reactions, making inferences) may be critical in detecting narrative discourse deficits in adults with TBI. Continued research is needed to determine if a story with multiple episodes and increased complexity may be more likely to detect differences and whether similar results would be obtained in populations of other acquired communication disorders such as aphasia or other cognitive communication disorders (Richardson & Dalton, 2016).

These diagnosis-related deficits in performance provide construct validity evidence for the use of scores from MSSG analyses to determine whether an individual is producing less accurate, complete, and logically sequenced content or fewer total episodic components secondary to traumatic brain injury.

Consistent vs. Discrepant Performance Across Macrostructural Analyses

Across the overall sample, MC + sequencing combined scores were highly consistent with the number of total episodic components. This is an expected finding, as the more MCs an individual produced and in a logical sequence, the more episodic components that individual would be expected to produce as the majority of MCs are coded as episodic components. In addition, as a result of TBI and potential damage to language, memory, or executive functioning centers in the brain, adults with TBI would be expected to show deficits in both story content and use of episodic components. This pattern of both poor story content and fewer episodic

components was seen in 21 participants, 15 of whom were from the TBI sample. These individuals were predominantly male (11 of 15) and experienced post traumatic amnesia of 60+ days (eight of 15). In fact, the two individuals who received the lowest scores, both for MC + sequencing and total episodic components, were both male and interestingly had post traumatic amnesia for over 150 days.

Along with the detection of differences between groups, discrepant performances were found within both groups. Out of the 76 participants, four participants demonstrated discrepant performances on the MC + sequencing and total episodic components measures. One participant exhibited average logical sequencing, accuracy, and completeness of story content in the presence of poorer than expected use of episodic components, while three participants demonstrated average episodic component use in the presence of below average sequencing, accuracy, and completeness of story content. For example, a 21-year-old male who was in the TBI sample produced 11 out of 15 required episodic components (total episodic complexity: z = - .165, based on Greenslade et al., in submission) indicating a fairly complex story that adheres to expected story grammar organization. In comparison, he obtained a total combined score of 70 out of 204 (z = -1.363, based on Greenslade et al., in submission) indicating the presence of poorly sequenced, inaccurate, and/or incomplete content. Discrepancies in scores such as these are significant because a narrative that has an adequate number of initiating events, attempts, and direct consequences may not be presented in the correct sequence or may feature information that misinforms or confuses the listener. For example, in the previously mentioned participant's narrative, he first stated that Cinderella goes to the ball, and then said that she got a beautiful dress and glass slippers. While receiving the dress and glass slippers is Cinderella's attempt to get to the ball, and Cinderella going to the ball is a direct consequence, these components do not

come in logical sequence. Importantly, three out of the four participants with discrepant scores were from the TBI sample. Further research is needed to determine whether adults from other clinical populations like aphasia or other cognitive communication disorders may demonstrate similar discrepancies and whether the MSSG will efficiently and effectively capture such narrative discourse deficits.

Limitations and Future Directions

While the results of the present study provide preliminary evidence for the use of MSSG analyses to detect differences in the narrative discourse of adults with TBI, readers are encouraged to consider the following limitations. First, although the total sample size was 78, the number of participants in each group was only 36, a relatively small sample size. In addition, while all participants in the sample spoke English as their primary language, the transcripts of NHCs came from labs across the United States, whereas the transcripts of adults with TBI came from the Togher lab in Australia. Although English is the common language, there are subtle linguistic nuances, phrases, and differences between the two cultures that may influence scoring and narrative interpretation. The sample also lacked diversity in terms of ethnicity, race, and sex. The control sample featured primarily Caucasians, and more males were included due to the higher rate of males in the TBI group. Finally, to increase the homogeneity of the sample, only those with closed-head brain injury were included in the TBI group. Thus, the present study's results cannot be generalized to those with open head injuries.

Future research is needed to address these limitations. Possible directions may include examining narrative discourse in individuals with differing types and degrees of head injuries, such as open or mixed head injuries, and in a larger and more diverse clinical sample to provide

a clearer understanding of the impact of TBI on story macrostructure. Additionally, further research into differences in the narrative discourse of adults with other acquired language disorders, such as aphasia and other cognitive communication disorders (e.g., right hemisphere disorder), should be investigated to provide further insights into the nature of discourse deficits in different clinical populations.

In addition, future research should continue to focus on the creation and development of clinically useful, relevant, and efficient tools sensitive to discourse impairments. Specifically, research should explore the suitability and reliability of current analyses in clinical settings, including acute care settings and adult outpatient settings in which early identification is of paramount importance. To investigate increasing analysis efficiency for such settings, research should examine whether real time scoring and/or scoring from videos without transcription would be possible while maintaining reliability. Additional evidence is needed to show the construct validity of MSSG scores for revealing client progress over time, including improvements throughout the acute recovery period and onward. Elbourn et al. (2019) showed an improvement in main concept scores during the first year post-TBI. Future research should explore whether measuring sequencing and story grammar organization close to onset of TBI and at checkpoints along recovery may prove useful in showing progress, and in re-evaluating strengths and weaknesses of an individual's communication over time. Finally, given that MSSG analyses did not examine every utterance in participants' transcripts, a comparison between these measures and traditional story grammar analyses would allow researchers to determine the new measures' effectiveness in capturing macrostructural challenges.

CONCLUSION

Story telling is an essential skill that makes life more fulfilling by allowing individuals to fully participate and establish meaningful connections and relationships. The current study provided preliminary data showing significant differences between the Cinderella narratives of adults with TBI and neurologically healthy controls across five measures: main concept, sequencing, main concept + sequencing, episodic complexity, and total episodic components. Overall, individuals with TBI told less accurate, complete, and logically sequenced stories. Their stories frequently featured fewer complex episodes with fewer episodic components, yielding less organized stories. MSSG analyses detected differences between populations with adequate reliability, supporting the construct validity of scores for identifying narrative macrostructure deficits. Future research is still needed to further confirm reliability and validity as well as to explore the clinical utility and feasibility of MSSG analyses for different clinical purposes and in different clinical settings. In the meantime, the present study's findings serve to aid clinicians in detecting and identifying areas of strengths and challenge in adults with TBI, which may improve the effectiveness and personalization of treatment planning.

REFERENCES

Barrett, J. (1998*). Old McDonald had an apartment house* (2nd ed.). New York, NY: Atheneum.

Biddle, K., McCabe, A., & Bliss, L. (1996). Narrative skills following traumatic brain injury in children and adults. *Journal of Communication Disorders*, *29*, 447-469.

Cannizzaro, M.S., & Coelho, C.A. (2013). Analysis of narrative discourse structure as an ecologically relevant measure of executive function in adults. *Journal of Pyscholinguistic Research, 42*(6), 527-549. http:// 10.1007/s10936-012-9231-5

Carlomagno, S., Giannotti, S., Vorano, L., & Marini, A. (2011). Discourse information content in non-aphasic adults with brain injury: A pilot study. *Brain Injury, 25*(10), 1010-1018. http://doi.org/10.3109/02699052.2011.605097

Coelho, C. A. (2002). Story Narratives of Adults With Closed Head Injury and Non- Brain-Injured Adults: Influence of Socioeconomic Status, Elicitation Task, and Executive Functioning. 18.

Coelho, C. A., Liles, B. Z., & Duffy, R. J. (1995). Impairments of discourse abilities and executive functions in traumatically brain-injured adults. *Brain Injury, 9*(5), 471–477. http://doi.org/10.3109/02699059509008206

Coelho, C. A., Ylvisaker, M., & Turkstra, L. S. (2005). Nonstandardized assessment approaches for individuals with traumatic brain injuries. *Seminars in Speech and Language, 26*(4) 223-241. http://doi.org/10.1055/s-2005-922102

Davis, G. A., & Coelho, C. A. (2004). Referential cohesion and logical coherence of narration after closed head injury. *Brain and Language*, *89*(3), 508–523. https://doi.org/10.1016/j.bandl.2004.01.003

Elbourn, E., Kenny, B., Power, E., Honan, C., McDonald, S., Tate, R., … Togher, L. (2019). Discourse recovery after severe traumatic brain injury: Exploring the first year. *Brain Injury*, *33*(2), 143–159. https://doi.org/10.1080/02699052.2018.1539246

Ghayoumi Anaraki, Z. (2014). Narrative discourse impairments in Persian-speaking persons with traumatic brain injury: a pilot study. *Folia Phoniatrica et Logopaedica.*, *66*(6), 273–279.

Greenslade, K., Stuart, J., Richardson, J.D., Dalton, S. G., & Ramage, A. (in submission).

Hartley, L. L., & Jensen, P.J. (1991). Narrative and procedural discourse after closed head injury. *Brain Injury, 5*(3), 267–285. http://doi.org/10.3109/02699059109008097

Hunt, K.W. (1965) Grammatical Structures Written at Three Grade Levels. National Council of Teachers of English Research Report No. 3, National Council of Teachers of English, Urbana.

Jorgensen, M. & Togher, L. (2009). Narrative after traumatic brain injury: A comparison of monologic and jointly-produced discourse. *Brain Injury*, *23*(9), 727–740. http://doi.org/10.1080/02699050903133954

Laerd Statistics. (2019). Retrieved from https://statistics.laerd.com/

Lê, K., Coelho, C., Mozeiko, J., Krueger, F., & Grafman, J. (2011). Measuring goodness of story narratives: Implications for traumatic brain injury. *Aphasiology*, *25*(6–7), 748–760. https://doi.org/10.1080/02687038.2010.539696

Lindsey, A., Hurley, E., Mozeiko, J., & Coelho, C. (2019). Follow-Up on the Story Goodness Index for Characterizing Discourse Deficits Following Traumatic Brain Injury. *American Journal of Speech-Language Pathology*, *28*(1S), 330–340. https://doi.org/10.1044/2018_AJSLP-17-0151

Marini, A., Boewe, A., Caltagirone, C., & Carlomagno, S. (2005). Age-related differences in the production of textual descriptions. *Journal of Psycholinguistic Research, 34*(5), 439–463. http://doi.org/10.1007/s10936-005-6203-z

Marini, A., Galetto, V., Zampieri, E., Vorano, L., Zettin, M., & Carlomagno, S. (2011). Narrative language in traumatic brain injury. *Neuropsychologia, 49*(10), 2904–2910. http://doi.org/10.1016/j.neuropsychologia.2011.06.017

Matsuoka, K., Kotani, I., & Yamasato, M. (2012). Correct information unit analysis for determining the characteristics of narrative discourse in individuals with chronic traumatic brain injury. *Brain Injury, 26*(13–14), 1723–1730. http://doi.org/10.3109/02699052.2012.698789

Mozeiko, J., Le, K., Coelho, C., Krueger, F., & Grafman, J. (2011). The relationship of story grammar and executive function following TBI. *Aphasiology*, *25*(6–7), 826–835. https://doi.org/10.1080/02687038.2010.543983

Nicholas, L. E., & Brookshire, R. H. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech, Language, and Hearing Research*, *38*(1), 145–156. https://doi.org/10.1044/jshr.3801.145

Peach, R. K. (2013). The cognitive basis for sentence planning difficulties in discourse after traumatic brain injury. *American Journal of Speech-Language Pathology, 22*(2), S285–S297. http://doi.org/10.1044/1058-0360(2013/12-0081)

Power, E., Weir, S., Richardson, S., Fromm, D., Forbes, M., MacWhinney, B., & Togher, L. (2020). Patterns of narrative discourse in early recovery following severe traumatic brain injury. *Brain Injury, 34*(1), 98-109. http://doi.org/10.1080/02699052.2019.1682192

Richardson, J. D., & Dalton, S. G. (2016). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology*, *30*(1), 45–73. https://doi.org/10.1080/02687038.2015.1057891

Richardson, J. D., & Dalton, S. G. (2019). A large-scale comparison of main concept production between persons with aphasia and persons without brain injury. American Journal of Speech-Language Pathology, 28(1 Suppl), 293-310. https://doi.org/10.1044/2018_AJSLP-17-0166

Roth, F. P., & Spekman, N. J. (1986). Narrative discourse: spontaneously generated stories of learning-disabled and normally achieving students. *Journal of Speech and Hearing Disorders*, *51*(1), 8–23. https://doi.org/10.1044/jshd.5101.08

Stark, J. A. (2010). Content analysis of the fairy tale *Cinderella* – A longitudinal single-case study of narrative production: "From rags to riches." *Aphasiology*, *24*(6–8), 709–724. https://doi.org/10.1080/02687030903524729

Stein, N. L., & Glenn, C. G. (1975). A developmental study of children's recall of story material. Paper presented at the Society for Research in Child Development. Denver, 1975.

Stein, N. L., & Glenn, G. (1979.). *An Analysis of Story Comprehension in Elementary School Children*. 65.

Stout, C. E., Yorkston, K. M., Pimentel, J. I. (2000). Discourse production following mild, moderate, and severe traumatic brain injury. *Journal of Medical Speech-Language Pathology, 8*(1), 15–25.