Research Article

A Large-Scale Comparison of Main Concept Production Between Persons With Aphasia and Persons Without Brain Injury

Sarah Grace Hudspeth Dalton[a] and Jessica D. Richardson[a]

Purpose: The purposes of this study are to provide clinicians and researchers with introductory psychometric data for the main concept analysis (MCA), a measure of discourse informativeness, and specifically, to provide descriptive and comparative statistical information about the performance of a large sample of persons not brain injured (PNBIs) and persons with aphasia (PWAs) on AphasiaBank discourse tasks.
Method: Transcripts of 5 semi-spontaneous discourse tasks were retrieved from the AphasiaBank database and scored according to detailed checklists and scoring procedures. Transcripts from 145 PNBIs and 238 PWAs were scored; descriptive statistics, median tests, and effect sizes are reported.
Results: PWAs demonstrated overall lower informativeness scores and more frequent production of statements that

were inaccurate and/or incomplete. Differences between PNBIs and PWAs were observed for all main concept measures and stories. Comparisons of PNBIs and aphasia subtypes revealed significant differences for all groups, although the pattern of differences and strength of effect sizes varied by group and discourse task.
Conclusions: These results may improve the investigative and clinical utility of the MCA by providing descriptive and comparative information for PNBIs and PWAs for standardized discourse tasks that can be reliably scored. The results indicate that the MCA is sensitive to differences in discourse as a result of aphasia.
Supplemental Material: https://doi.org/10.23641/asha. 7485647

T he goal of speech-language therapy for persons with aphasia (PWAs) is to restore communication abilities and reduce disruptions to activities and participation (Brady, Kelly, Godwin, Enderby, & Campbell, 2016). Primary treatment outcomes in both research and clinical practice (Kelly, Brady, & Enderby, 2010; Robey, 1998; Simmons-Mackie, Threats, & Kagan, 2005) have traditionally been formal measures of overall language severity (e.g., Aachen Aphasia Test, Huber, 1984; Western Aphasia Battery [WAB], Kertesz, 1982) or impairment-specific severity (e.g., Boston Naming Test, Kaplan, Goodglass, & Weintraub, 2001; Northwestern Assessment of Verbs and Sentences, Thompson, 2011). These standardized and widely used measures allow comparisons across patients, groups, and studies but may lack adequate

psychometrics, particularly when measuring treatment-induced change or change over time. Additionally, evidence suggests that these measures may not be sufficient to predict functional communication, quality of life, or participation (Larfeuil & Le Dorze, 1997; K. B. Ross & Wertz, 1999).

Reprioritization of outcomes now situates measures of functional communication as primary outcomes, whereas traditional impairment-based measures are secondary outcomes, useful as surrogates of the primary outcomes (Brady et al., 2016). Discourse is included as one of the new primary outcome measures (Brady et al., 2016) but remains underutilized by researchers and clinicians (Simmons-Mackie et al., 2005). Discourse includes a wide variety of speech acts, from telling stories, to giving directions, to conversation; through discourse, we not only express basic wants and needs but build relationships and community (van Dijk, 1997). Impairments in discourse, such as those seen in individuals with aphasia, can have a deeply negative impact on quality of life and life participation (Hilari, 2011; Hilari et al., 2010; K. Ross & Wertz, 2003).

Conversation, which generally encompasses different kinds of discourse (e.g., storytelling, debating, sharing humor), is the type of communication most frequently

[a]The University of New Mexico, Albuquerque

Correspondence to Sarah Grace Hudspeth Dalton: sarahgrace.hudspeth@gmail.com

utilized in daily life and is critical to the maintenance of healthy relationships (E. Armstrong & Ferguson, 2010; Wallace et al., 2017). Improvement in conversation as a direct result of therapy is the ultimate sign of an effective treatment but is difficult to reliably measure. Many micro-linguistic discourse measures may be applied to conversation (e.g., type–token ratio, mean length of utterance, percent nouns), but others above the level of words or utterances are either not fully compatible (e.g., main concept analysis [MCA], story grammar analysis) or would require modification (e.g., cohesion analysis) before being applied to conversation.

Semi-spontaneous discourse tasks, like retelling a story or describing a procedure, are widely used as a compromise between formal assessments and conversation. There are many different measures, elicitation procedures, and stimuli—yielding both advantages and disadvantages. Variety allows researchers and clinicians to select the most appropriate tool for their study or client but limits generalizability across studies. Also, most discourse measures lack normative data and are restricted to measuring treatment-induced changes in the same individual. Without norms, it is difficult to use discourse as a diagnostic tool, either for identification of aphasia broadly or for identifying specific aphasia subtypes. A recent review concluded that, although most discourse measures are theoretically well grounded, they lack sufficient psychometric definition to be utilized as diagnostic or outcome measures in isolation (Pritchard, Hilari, Cocks, & Dipper, 2017). Another review suggested that further studies using established discourse measures with large samples and varied severities of aphasia may shed light on some of the contradictions present in the literature currently (Linnik, Bastiaanse, & Höhle, 2016). What is needed are normative samples and large-scale analyses that can be used to describe and compare the productions of PWAs to themselves, other PWAs, and speakers without brain injury. Several barriers have limited such research, notably the sheer effort required to collect, transcribe, and analyze large corpora of discourse from both healthy controls and PWAs, the wide array of possible discourse tasks, and lack of consensus on the best measures to use. Development of databases such as AphasiaBank has paved the way for normative studies by establishing a standardized discourse protocol and facilitating data collected across multiple sites (see MacWhinney, Fromm, Forbes, & Holland, 2011); currently, there are five contributors of healthy control data, and 21 contributors of PWA data. While AphasiaBank has removed several barriers to normative studies, lack of consensus regarding measures is less easily overcome.

Discourse Informativeness Measures

There is a rich history of discourse measures that examine the content or informativeness of discourse produced by PWAs. Regardless of whether the unit of analysis under study is referred to as a correct information unit (CIU; Nicholas & Brookshire, 1993), content unit (Yorkston &

Beukelman, 1980), information unit (IU; McNeil, Doyle, Fossett, Park, & Goda, 2001), main concept (MC; Nicholas & Brookshire, 1995), main event (Capilouto, Wright, & Wagovich, 2006), proposition (Ulatowska, Freedman-Stern, Doyel, Macaluso-Haynes, & North, 1983), or theme (Gleason et al., 1980), the goal of the analysis is to quantify the amount of information relayed by PWAs. These measures can be divided based on whether they measure informativeness at a microstructural (below the level of a sentence) versus a macrostructural level (sentence level or higher). Microstructure informativeness measures include the content unit, CIU, and IU whereas macrostructure measures include main event, proposition, and theme analyses. The MCA has been described as a hybrid measure because it depends heavily on the lexical items produced (i.e., microstructure) but must also contain a verb and its constituent nouns (and potentially associated clauses) to receive full credit (i.e., macrostructure; E. Armstrong, 2000; Davis & Coelho, 2004).

CIUs are perhaps the most well-studied informativeness measure and are broadly defined as single words that are accurate, relevant, and informative regarding the topic or stimulus (see Appendix B in Nicholas & Brookshire, 1993, for details). McNeil et al. (2001) later developed a similar measure called *IUs*, which are words predetermined based on the stimuli in the story retell procedure (Doyle et al., 2000, 1998). Gleason et al. (1980) used production of target lexemes (lexical items identified by the authors) to investigate possible differences between individuals with Broca's and Wernicke's aphasia during narrative production. Studies utilizing these measures demonstrated that individuals with aphasia produced fewer informative lexical items compared with controls. Investigations of proposition-level analyses of informativeness reveal similar findings. PWAs produce only half as many themes (Gleason et al., 1980), fewer essential and peripheral propositions (Ulatowska et al., 1983), and a lower proportion of main events to total events (Capilouto et al., 2006) than healthy speakers. Importantly, the work by Ulatowska et al. demonstrated that although PWAs produced fewer propositions than healthy speakers, the propositions that they produced were essential to the successful telling of the story.

As mentioned above, MCA may inform researchers and clinicians about microstructure and macrostructure abilities and deficits (Nicholas & Brookshire, 1995). MCA measures how well an individual conveys the gist, or the essential elements, of a story. An *MC* is defined as an utterance (containing one main verb, its constituent nouns, and any associated clauses), which is scored based on accuracy (all essential information is correct) and completeness (all essential information is present). MCA is sensitive to differences between control speakers and individuals with aphasia (Kong, 2009; Kong, Whiteside, & Bargmann, 2016; Nicholas & Brookshire, 1995), as well as between individuals with fluent and nonfluent aphasia (Kong et al., 2016). Informativeness measures (CIUs, propositions, % accurate/complete [AC] MCs) may be useful for capturing treatment response (Albright & Purves, 2008; Avent & Austermann, 2003; Coelho,

McHugh, & Boyle, 2000; Cupit, Rochon, Leonard, & Laird, 2010; Stark, 2010) and are correlated with listener perceptions (Cupit et al., 2010; K. B. Ross & Wertz, 1999), suggesting they are consistent with person-centered frameworks. MCA has also been reported to correlate well with formal measures of overall severity (Kong, 2011; Kong et al., 2016), micro-linguistic measures (Dalton & Richardson, 2015), confrontation naming (Richardson et al., 2018), listener perceptions (Cupit et al., 2010; K. B. Ross & Wertz, 1999), and conversational informativeness (Doyle, Goda, & Spencer, 1995).

Importantly, previous research has demonstrated that MCA is reliable within and across judges, with reliability consistently above 80% (Boyle, 2014; Dalton & Richardson, 2015; Kong, 2011; Kong et al., 2016; Nicholas & Brookshire, 1995; Richardson & Dalton, 2016) and adequate test–retest reliability (Kong, 2011; Kong et al., 2016; Nicholas & Brookshire, 1995), although perhaps, only when multiple discourse tasks are combined into a single sample (Boyle, 2014, 2015; Brookshire & Nicholas, 1994a, 1994b). MCA can be done without phonetic transcription (L. Armstrong, Brady, Mackenzie, & Norrie, 2007), a critical factor for eventual widespread clinical adoption. Additionally, normative data for four discourse tasks are available for an unimpaired elderly control group and small groups of PWAs with fluent aphasia, PWAs with nonfluent aphasia, and individuals with Alzheimer's-type dementia (Kong et al., 2016).

Recently, MC lists were developed for the semi-spontaneous discourse tasks included in the AphasiaBank protocol (one picture scene description, two sequential picture descriptions, one story retell, and one procedure) using healthy control speakers from the database (Richardson & Dalton, 2016; Richardson & Dalton, 2019). The authors included scoring procedures and information regarding how healthy controls performed on each story. In this article, we present the results from a large MCA of PWAs with different subtypes and compare their performance to that of a large sample of persons without a reported history of brain difference or brain injury. We report descriptive statistical information about the performance of a large sample of healthy controls and PWAs on the semi-spontaneous discourse tasks included in the AphasiaBank database protocol. We also extend the work conducted by Kong et al. (2016) to provide clinicians with information about how PWAs perform on MCA compared with healthy control speakers for these AphasiaBank tasks, using previously developed MC checklists. This project represents important steps in the continued development of psychometric properties of MCA. It is our hope that making this information available will aid in the completion of future studies that will speak directly to the usefulness of MCA for diagnosis and outcomes measurement.

Method

Participants

Transcripts from 320 individuals with aphasia were retrieved from the AphasiaBank database in order to score

MCs for five tasks: sequential picture description (*Broken Window* and *Refused Umbrella*), picture scene description (*Cat Rescue*), story retell (*Cinderella*), and procedural discourse (how to make a peanut butter and jelly sandwich, hereafter referred to as *Sandwich*). These 320 transcripts represented all transcripts available on the database as of April 2017. All samples were elicited using the AphasiaBank protocol (http://aphasia.talkbank.org/protocol/), which instructs participants to tell a story with a beginning, middle, and end during picture description tasks. For *Cinderella,* participants review a wordless picture book prior to attempting the storytelling, and for the *Sandwich* procedure, a picture stimulus is only used when participants are unable to produce a verbal response. Using the computerized language analysis program available from AphasiaBank, each of these stories was isolated from the individual's full transcript for ease of scoring.

In order to compare the performance of PWAs to control performance, transcripts of 168 persons not brain injured (PNBIs) from the AphasiaBank database were retrieved. Within the AphasiaBank database are also transcripts of individuals who have had a stroke but who currently score above the WAB cutoff and are determined not to have aphasia (not aphasic by WAB; NABW). Consistent with previous research demonstrating differences in discourse performance between this group and PNBIs and other PWAs (Dalton & Richardson, 2015; Fromm et al., 2017; Fromm, Forbes, Holland, & MacWhinney, 2013), discourse performance of individuals NABW was investigated as a distinct subtype of aphasia as they often report limitations that prevent return to full preinjury functioning despite scoring above standardized test cutoffs.

All WAB aphasia subtypes were initially represented in the data; however, individuals with transcortical sensory (two), global (four), and transcortical motor (12) aphasia were excluded because of the small sample sizes, so that 302 participants remained. Individuals who did not complete all five discourse tasks were also excluded, resulting in 238 (104 female, 134 male) PWAs and 145 (77 female, 68 male) PNBIs in this study. There were 86 anomic, 61 Broca, 46 conduction, 26 NABW, and 19 Wernicke aphasics in the sample (based on WAB classification). The average Western Aphasia Battery–Aphasia Quotient (WAB-AQ) score was 72.2 (SD = 19.1). For PWAs, the average age was 61.7 years (SD = 12.6) with an average education of 15.5 years (SD = 2.8). The average age of PNBIs was 63.4 years (SD = 19.1) with an average education of 15.4 years (SD = 2.5). See Table 1 for complete demographics by group and aphasia subtype. All PNBIs in the database who completed all discourse tasks were included in this study in order to more closely match the average age of PWAs and to improve statistical power with more even groups.

MC Scoring

Transcripts were scored for MCs using standardized lists created from the stories of 92 PNBI speaker's transcripts

Table 1. Demographic information for all participants.

Variable	Not brain injured (*n* = 145)	All PWAs (*n* = 238)	NABW (*n* = 26)	Anomic (*n* = 86)	Broca's (*n* = 61)	Conduction (*n* = 46)	Wernicke's (*n* = 19)
Age (years)[a]	63.4 (±19.1) 20.0–89.5	61.7 (±12.6) 25.6–90.7	61.1 (±13.7) 26–80.7	62 (±12.1) 32.7–85.7	58 (±13.2) 25.6–85.4	64.5 (±11.9) 30.9–90.7	66.9 (±11.1) 42.6–81.3
Aphasia duration (months)[b]		64.1 (±58.3) 4–360	64.3 (±45.8) 12–188	55.7 (±51.1) 6–240	75.9 (±62.8) 5–309	61.3 (±57.2) 6–296	71.6 (±87) 4–360
WAB Aphasia Quotient[c]		72.2 (±19.1) 10.8–99.6	96.5 (±1.8) 93.8–99.6	84.7 (±6.8) 63.4–93.4	51.8 (±15.4) 10.8–77.6	70.5 (±9.1) 49.5–90	52.7 (±13.5) 28.2–72.6
Gender	77 female 68 male	104 female 135 male	18 female 8 male	38 female 48 male	20 female 41 male	22 female 24 male	6 female 13 male
Education (years)[d]	15.4 (±2.5) 11–23	15.5 (±2.8) 8–25	16 (±2.9) 12–21	15.9 (±2.8) 12–23	14.8 (±2.7) 8–23	15.5 (±3.1) 11–25	15.5 (±2.3) 12–19
Race/ethnicity	139 Caucasian 3 African American 3 Hispanic/Latino — —	207 Caucasian 19 African American 6 Hispanic/Latino 5 other 1 unknown	23 Caucasian 1 African American 2 Hispanic/Latino — —	81 Caucasian 3 African American 2 Hispanic/Latino — —	48 Caucasian 9 African American 2 Hispanic/Latino 2 other —	40 Caucasian 3 African American — 3 other —	15 Caucasian 3 African American — — 1 unknown

Note.   Em dashes indicate that there were no individuals of that race/ethnicity in the sample. PWAs = persons with aphasia; NABW = not aphasic by WAB; WAB = Western Aphasia Battery.

[a]Two individuals (one conduction, one Wernicke) are missing age data. [b]One individual (Wernicke) is missing aphasia duration data. [c]One individual (anomic) is missing WAB Aphasia Quotient data. [d]Seven individuals (three anomic, two Broca, two Wernicke) are missing education data.

retrieved from the AphasiaBank database (see Richardson & Dalton, 2016, for details of list development and scoring procedures and Nicholas & Brookshire, 1995, for scoring rules). The same normative sample was used for all five stories, except *Cat Rescue,* because five of the original normative participants did not complete it. The replacement normative participants for *Cat Rescue* were matched to the five participants they replaced on age, gender, education, *Broken Window* MC composite score, and the number of utterances produced for the *Broken Window* task.

Briefly, to create the MC lists used here, utterances relevant to each story were identified, and a master list was created with all relevant concepts. From this master list of relevant concepts, any concept that was produced by more than one third of the normative sample was considered an MC. Each MC consisted of two or more essential elements and any elements that were commonly, but not always, produced with the MC. The abbreviated MC lists (without additional scoring information) for each story are found in the Appendix C; however, these abbreviated lists should not be used to score MCs. Please refer to the full checklists (Richardson & Dalton, 2016; Richardson & Dalton, 2019) and Nicholas and Brookshire (1995) for scoring assistance.

Using these MC lists, each participant's stories were scored for the presence or absence of MCs and for the accuracy and completeness of MCs that were present. Each MC consists of two or more essential elements—minimally a verb and its constituent nouns—but could also include prepositional phrases or other clauses that operated on the main verb. Coding procedures from Nicholas and Brookshire (1995) were utilized, where missing MCs were coded as absent (AB) and MCs that were present could receive one of four codes. An AC code was assigned if all essential elements were present and correct. An accurate/incomplete (AI) code was assigned if one or more essential elements were missing but all essential information that was produced was correct. An inaccurate/complete (IC) code was assigned if all essential elements were present but some essential elements were inaccurate based on control speakers' productions. Finally, an inaccurate/incomplete (II) code was assigned if one or more essential elements were missing and one or more of the essential elements that were produced were inaccurate (see Table 2; Richardson & Dalton, 2016). MC codes were transformed to numeric scores using the formula adapted from Kong (2009): $AC(3) + AI(2) + IC(2) + II(1) + AB(0) = MC$ score. The adaptation from Kong's original formula is the separation of the IC and II categories, which he combined into a single "inaccurate" category. Nicholas and Brookshire (1995) combined the IC and II codes in their scoring procedure as well; however, we report them separately so that semantic paraphasias (which could result in incorrect information being produced) are not more heavily penalized than phonemic paraphasias (which can be scored as accurate if the target word is understood from context). The scores for each MC were summed within stories to yield a story MC composite score. The maximum score for *Cinderella* was 102 (34 MCs); for *Broken Window,* 24 (eight MCs); and for *Refused Umbrella, Cat Rescue,* and *Sandwich,* 30 (10 MCs). In addition, the number of MCs a participant attempted to produce for each story (MC attempts) was calculated by adding the number of statements receiving AC, AI, IC, and II codes.

Data Analysis

For each discourse task, the following descriptive statistics for MC composite, MC attempts, and each MC code are reported in Appendix A: mean, standard deviation, median, range, skew, and kurtosis. For all variables except MC composite, the mean should be interpreted as the average number of statements produced that received that code. Statistics are reported for PNBIs, PWAs collapsed across subtypes, and PWAs by individual subtype. The maximum value for the MC codes and MC attempts is equal to the number of MCs for that story, whereas the maximum value for MC composite scores is equal to the number of MCs multiplied by 3 (score for AC statements). Skew > ±2 and kurtosis > ±4 would indicate unacceptable nonnormality; please see Appendix A for skew and kurtosis values (Fabrigar, Wegener, MacCallum, & Strahan, 1999; West, Finch, & Curran, 1995).

Omnibus median tests were used to examine differences in MC composite scores, MC codes, and the number of MC attempts for each story between PNBIs and PWAs (with individuals NABW included in the PWAs group). Planned follow-up comparisons examined differences between PNBIs and each aphasia subtype. Median tests were selected because all groups (except NABW) were nonnormally distributed and had differently shaped distributions, and the variables under analysis are frequency counts (MC codes and attempts) or not truly interval (MC composite). Holm–Bonferroni correction for multiple comparisons was applied. Effect sizes (phi, φ) are reported for each comparison to focus on practically relevant differences using the traditional cutoffs to determine small (0.1–0.29), medium (0.3–0.49), and large (≥ 0.5) effects (Fritz, Morris, & Richler, 2012). We focus our results and discussion only on those comparisons with medium or large effect sizes, as we expect those to be most clinically relevant (but see Supplemental Materials S1–S6 for full results).

The MC composite score findings and the results for each MC code are reported in separate sections along with the percent overlap of the PNBI's distribution with the PWAs' distributions and vice versa, similar to the method reported by McNeil et al. (2001), to inform on the suitability of these measures for diagnostic use. Finally, we report preliminary cutoffs based on the limit of PNBI performance. Individuals that score below the cutoff for MC composite, MC attempts, and AC codes can be confidently classified as preventing with a language impairment. Similarly, any individual that scores above the cutoff for AI, IC, II, and AB codes can be classified as presenting with language impairment.

Table 2. Examples of each main concept code for the 5 discourse tasks.

| Discourse task | Target | Main concept code | | | |
		Accurate/ complete	Accurate/ incomplete	Inaccurate/ complete	Inaccurate/ incomplete
Broken Window	He was playing soccer.	"Boy was uh playing soccer with uh a ball."	"Um boy is ball."	"They kick this around." (with clear referent for this)	"And baseball or something ball and uh course he pushed."
Cinderella	Cinderella ran down the stairs.	"She was running down the steps." (with clear pronoun referent)	"And she um she had to run." (with clear pronoun referent)	"So he gets out."	"They run."
Sandwich	Get the peanut butter.	"Oh, well first you get the peanut butter out."	"Uh peanut butter."	"And you get out the butter."	"And then do the peanuts and the jellies."
Cat Rescue	The dog was barking.	"Here's a dog barking up the tree."	"Dog."	"And then the little boy is barking underneath her."	"Yelling up tree."
Refused Umbrella	It is raining.	"The rain starts falling."	"Raining."	"And now the water is falling."	"Draining."

Results

Comparisons Between PNBIs and PWAs

MC Composite

Examining the descriptive statistics for MC composite scores shows that PWAs had lower scores, with a more restricted range, compared with PNBIs. Results of the MC composite analysis revealed a significant difference between PNBIs and the entire group of PWAs for all stories (all $p < .001$), with large effect sizes (ϕ between 0.53 and 0.62), confirming that PWAs had overall lower MC composite scores (see Appendix B). However, when examining the results by aphasia subtype (see Figure 1A–E), it became clear that this relationship did not hold for individuals NABW. The boxplots for MC composite for each story show that the overlap between groups varied widely (see Table 3) with less than 50% overlap of PNBIs achieved for individuals with Broca's and Wernicke's aphasia for all stories. When examining the percent overlap of the aphasia subtypes with PNBIs, there was less than 50% overlap with PNBIs for *Cinderella* and *Refused Umbrella* (Broca's). PNBIs in this sample did not have an MC composite score lower than 5 for the *Broken Window* and *Sandwich* tasks. For the *Cat Rescue* and *Refused Umbrella* tasks, the lowest MC composite scores in the PNBI group were 8 and 9, respectively. *Cinderella* falls in the middle of the other tasks, with no PNBI scoring lower than 7 on MC composite.

MC Attempts

The descriptive statistics show that PWAs had fewer MC attempts than PNBIs with a restricted range. Results of the MC attempt analysis revealed a significant difference between PNBIs and the entire group of PWAs for all stories (all $p < .001$), with medium to large effect sizes (ϕ between 0.42 and 0.56), indicating that PWA's discourse samples exhibited reduced output overall compared with PNBIs (see
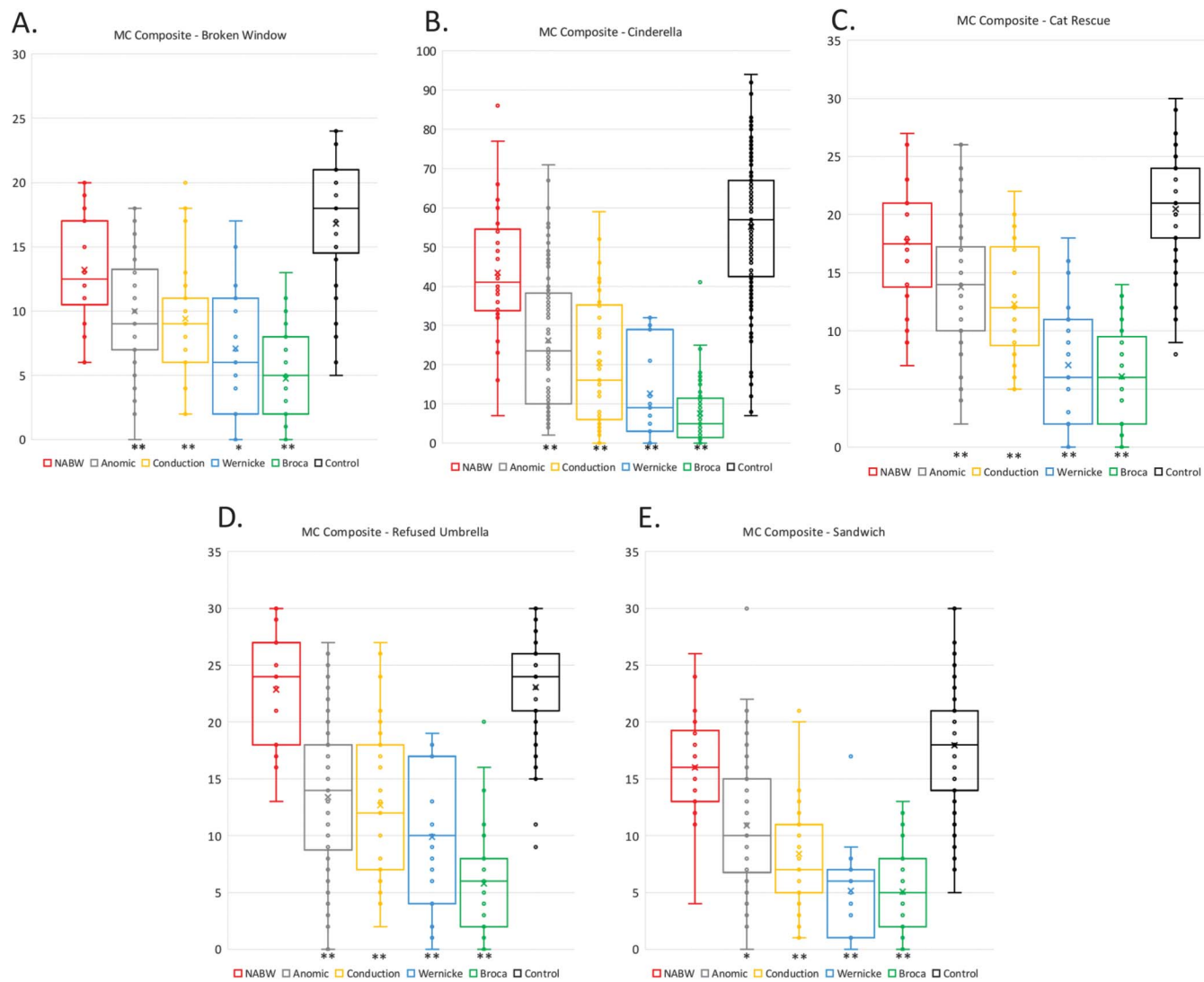
Appendix B), although this finding does not hold for individuals NABW. All other subtypes demonstrated differences with medium to large effect sizes for all stories except the *Broken Window* story for individuals with Wernicke's aphasia (see Figure 2A–E). There was less than 50% overlap of the PNBIs distribution with individuals with Broca's (*Broken Window, Cinderella*) and Wernicke's aphasia (*Cinderella*; see Table 3). When examining the percent overlap of the aphasia subtype distributions with PNBIs, no group had less than 50% overlap for any story. During the *Broken Window, Cinderella,* and *Sandwich* tasks, PNBIs attempted to produce no fewer than two MCs and, for *Refused Umbrella* and *Cat Rescue,* no fewer than three MCs.

MC Codes

AC. PWAs had fewer AC codes than PNBIs with a restricted range. Results of the AC code analysis revealed a significant difference between PNBIs and the entire group of PWAs for all stories (all $p < .001$), with large effect sizes (ϕ between 0.65 and 0.69), indicating that PWAs produced fewer AC statements across all subtypes and tasks (see Appendix B). This pattern holds for all subtypes (with medium to large effect sizes) except individuals NABW who showed significant differences from PNBIs for all stories but with small effect sizes for all comparisons (see Figure 3A–E). There was less than 50% overlap of the PNBIs distribution with individuals with Broca's and Wernicke's aphasia for all stories (see Table 3). When examining the percent overlap of the aphasia subtypes with PNBIs, there was less than 50% overlap with PNBIs for all stories (Broca's) and for *Cinderella, Cat Rescue,* and *Sandwich* (Wernicke's). PNBIs produced no fewer than one AC statement for *Broken Window, Refused Umbrella,* and *Sandwich,* no fewer than two for *Cinderella* and *Cat Rescue.*

AI. For these results and the following paragraphs about the IC, II, and AB codes, the direction of differences

Figure 1. Boxplots of the distribution of MC composite scores for all groups for the (A) *Broken Window,* (B) *Cinderella,* (C) *Cat Rescue,* (D) *Refused Umbrella,* and (E) *Sandwich* tasks. A single asterisk below the groups indicates a significant difference between that group and persons not brain injured with a medium effect size. A double asterisk indicates a significant difference with a large effect size. MC = main concept; NABW = not aphasic by WAB.

is opposite that seen for MC composite, MC attempts, and AC codes. Examining the descriptive statistics for the AI code shows that PWAs had more AI codes than PNBIs with a wider range. Results of the AI code analysis revealed a significant difference between PNBIs and the entire group of PWAs for all stories (all *p* < .001), with medium to large effect sizes (ϕ between 0.349 and 0.57), indicating that PWAs produced more AI statements than PNBIs (see Appendix B). The story is less clear-cut when examining the results by aphasia subtype. For individuals NABW, medium effect sizes indicating a clinically significant difference were found for the *Broken Window* and *Cinderella* tasks. For individuals with anomic aphasia, medium and large effects were

observed for the *Cat Rescue, Cinderella,* and *Sandwich* tasks. For individuals with conduction and Broca's and Wernicke's aphasia, medium to large effect sizes were seen for all stories except the *Cinderella* story for individuals with Wernicke's aphasia (see Figure 4A–E). No group showed less than 50% overlap of distributions for any story (see Table 3). PNBIs produced no more than four AI statements for *Cinderella;* no more than three for *Refused Umbrella, Cat Rescue,* and *Sandwich;* and no more than one for *Broken Window.*

IC. Examining the descriptive statistics for this code shows that PWAs had more IC codes than PNBIs with a wider range. Results of the IC code analysis revealed a

Dalton & Richardson: *MC Production in Aphasia and Controls* **299**

Table 3. Percent overlap of persons not brain injured (PNBIs) with each aphasia subtype (left) and percent overlap of each aphasia subtype with PNBIs (right) for main concept (MC) composite score.
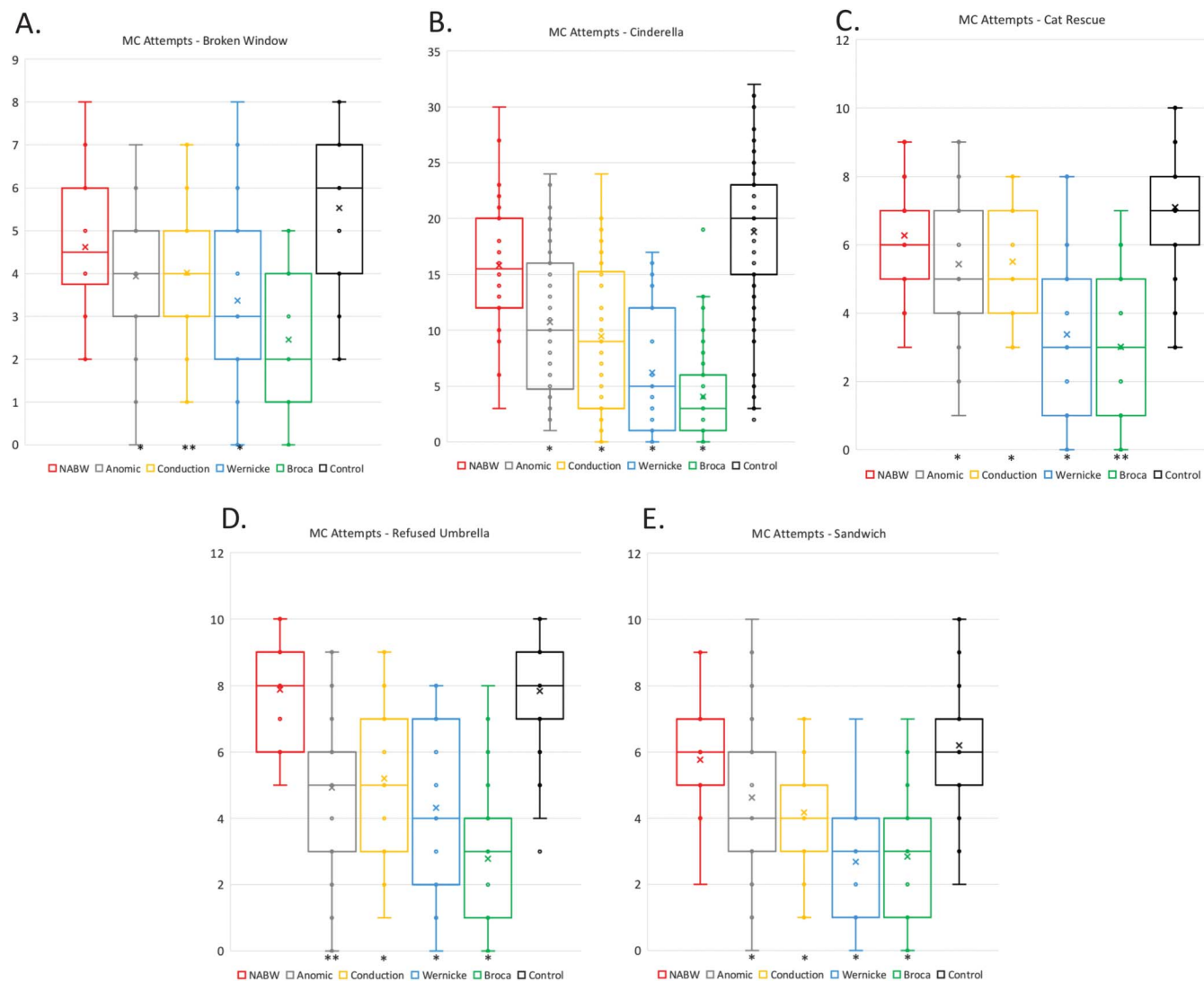
Variable	Broken Window	Cinderella	Sandwich	Cat Rescue	Refused Umbrella
MC composite					
PNBI/NABW ~ NABW/PNBI					
PNBI/anomic ~ anomic/PNBI	**69% ~ 91%**	**81% ~ 87%**		95% ~ 88%	
PNBI/conduction ~ conduction/PNBI	**70% ~ 91%**	**51% ~ 74%**	**79% ~ 80%**	**79% ~ 80%**	**82% ~ 65%**
PNBI/Wernicke's ~ Wernicke's/PNBI	45% ~ 63%	13% ~ 58%	40% ~ 58%	40% ~ 58%	19% ~ 58%
PNBI/Broca's ~ Broca's/PNBI	**23% ~ 52%**	**23% ~ 43%**	**21% ~ 54%**	**21% ~ 54%**	**21% ~ 23%**
Accurate/complete					
PNBI/NABW ~ NABW/PNBI					
PNBI/anomic ~ anomic/PNBI	74% ~ 80%	83% ~ 77%	100% ~ 71%	95% ~ 72%	97% ~ 86%
PNBI/conduction ~ conduction/PNBI	74% ~ 61%	26% ~ 50%	87% ~ 41%	52% ~ 41%	97% ~ 72%
PNBI/Wernicke's ~ Wernicke's/PNBI	3% ~ 53%	6% ~ 26%	12% ~ 16%	2% ~ 15%	5% ~ 68%
PNBI/Broca's ~ Broca's/PNBI	13% ~ 13%	5% ~ 8%	7% ~ 7%	< 1% ~ < 1%	5% ~ 20%
Accurate/incomplete					
PNBI/NABW ~ NABW/PNBI		100% ~ 88%			
PNBI/anomic ~ anomic/PNBI		100% ~ 74%			
PNBI/conduction ~ conduction/PNBI	100% ~ 54%	100% ~ 65%	100% ~ 87%	100% ~ 71%	
PNBI/Wernicke's ~ Wernicke's/PNBI	100% ~ 63%			100% ~ 78%	100% ~ 79%
PNBI/Broca's ~ Broca's/PNBI	100% ~ 51%	100% ~ 77%	100% ~ 80%	100% ~ 70%	100% ~ 74%
Inaccurate/complete					
PNBI/NABW ~ NABW/PNBI					
PNBI/anomic ~ anomic/PNBI					
PNBI/conduction ~ conduction/PNBI	**10% ~ 96%**				
PNBI/Wernicke's ~ Wernicke's/PNBI	100% ~ 89%				
PNBI/Broca's ~ Broca's/PNBI					
Inaccurate/incomplete					
PNBI/NABW ~ NABW/PNBI					
PNBI/anomic ~ anomic/PNBI		100% ~ 84%			
PNBI/conduction ~ conduction/PNBI		**100% ~ 52%**	**100% ~ 74%**		100% ~ 52%
PNBI/Wernicke's ~ Wernicke's/PNBI	100% ~ 74%	**100% ~ 79%**			100% ~ 84%
PNBI/Broca's ~ Broca's/PNBI	100% ~ 70%	99% ~ 85%	**100% ~ 82%**		100% ~ 85%
Absent					
PNBI/NABW ~ NABW/PNBI					
PNBI/anomic ~ anomic/PNBI		87% ~ 94%			**93% ~ 83%**
PNBI/conduction ~ conduction/PNBI		**82% ~ 85%**	77% ~ 93%	84% ~ 100%	**93% ~ 89%**
PNBI/Wernicke's ~ Wernicke's/PNBI	100% ~ 79%	36% ~ 68%	**77% ~ 74%**	84% ~ 57%	63% ~ 74%
PNBI/Broca's ~ Broca's/PNBI	**43% ~ 72%**	48% ~ 56%	92% ~ 75%	58% ~ 63%	63% ~ 60%
MC attempts					
PNBI/NABW ~ NABW/PNBI					
PNBI/anomic ~ anomic/PNBI		82% ~ 99%			**93% ~ 83%**
PNBI/conduction ~ conduction/PNBI		82% ~ 91%	77% ~ 93%	84% ~ 100%	93% ~ 89%
PNBI/Wernicke's ~ Wernicke's/PNBI	100% ~ 79%	36% ~ 74%	77% ~ 74%	84% ~ 57%	63% ~ 74%
PNBI/Broca's ~ Broca's/PNBI		48% ~ 66%	77% ~ 74%	**58% ~ 63%**	63% ~ 60%

Note. Empty cells reflect comparisons that were not significant, had small effect sizes, or had > 90% overlap between groups in both directions. Bold cells reflect comparisons with large effect sizes, and cells with regular font have medium effect sizes. See Supplemental Material S7 for complete table. NABW = not aphasic by WAB.

significant difference between PNBIs and the entire group of PWAs for the *Broken Window* and *Cat Rescue* tasks (all $p < .001$), with medium effect sizes ($\phi = 0.32$ and 0.35, respectively), indicating that PWAs produced significantly more statements judged IC than PNBIs (see Appendix B). Subtype comparisons were completed for all stories to ensure potentially significant differences in one group were not washed out by a lack of differences in the other groups. Results of the subtype analysis show a significant difference with a medium effect size ($\phi = 0.39$) for individuals NABW on the *Cat Rescue* task. For individuals with anomic aphasia there are significant differences with medium effect sizes for the *Broken Window, Cat Rescue,* and *Refused*

Umbrella tasks ($\phi = 0.38$, 0.45, and 0.36, respectively). For individuals with conduction and Wernicke's aphasia, significant differences were observed on the *Broken Window, Cat Rescue,* and *Refused Umbrella* tasks. Effect sizes were medium for all three comparisons for those with Wernicke's aphasia, and medium and large effect sizes were observed for individuals with conduction aphasia. For individuals with Broca's aphasia, no comparison yielded more than a small effect size, indicating that this code may not be clinically informative for this subtype (see Figure 5A–E). No group showed less than 50% overlap of distributions for any story (see Table 3). No PNBIs produced more than two IC statements for *Broken Window,* three for *Cat Rescue*

Figure 2. Boxplots of the distribution of MC attempts for all groups for the (A) *Broken Window,* (B) *Cinderella,* (C) *Cat Rescue,* (D) *Refused Umbrella,* and (E) *Sandwich* tasks. A single asterisk below the groups indicates a significant difference between that group and persons not brain injured with a medium effect size. A double asterisk indicates a significant difference with a large effect size. MC = main concept; NABW = not aphasic by WAB.
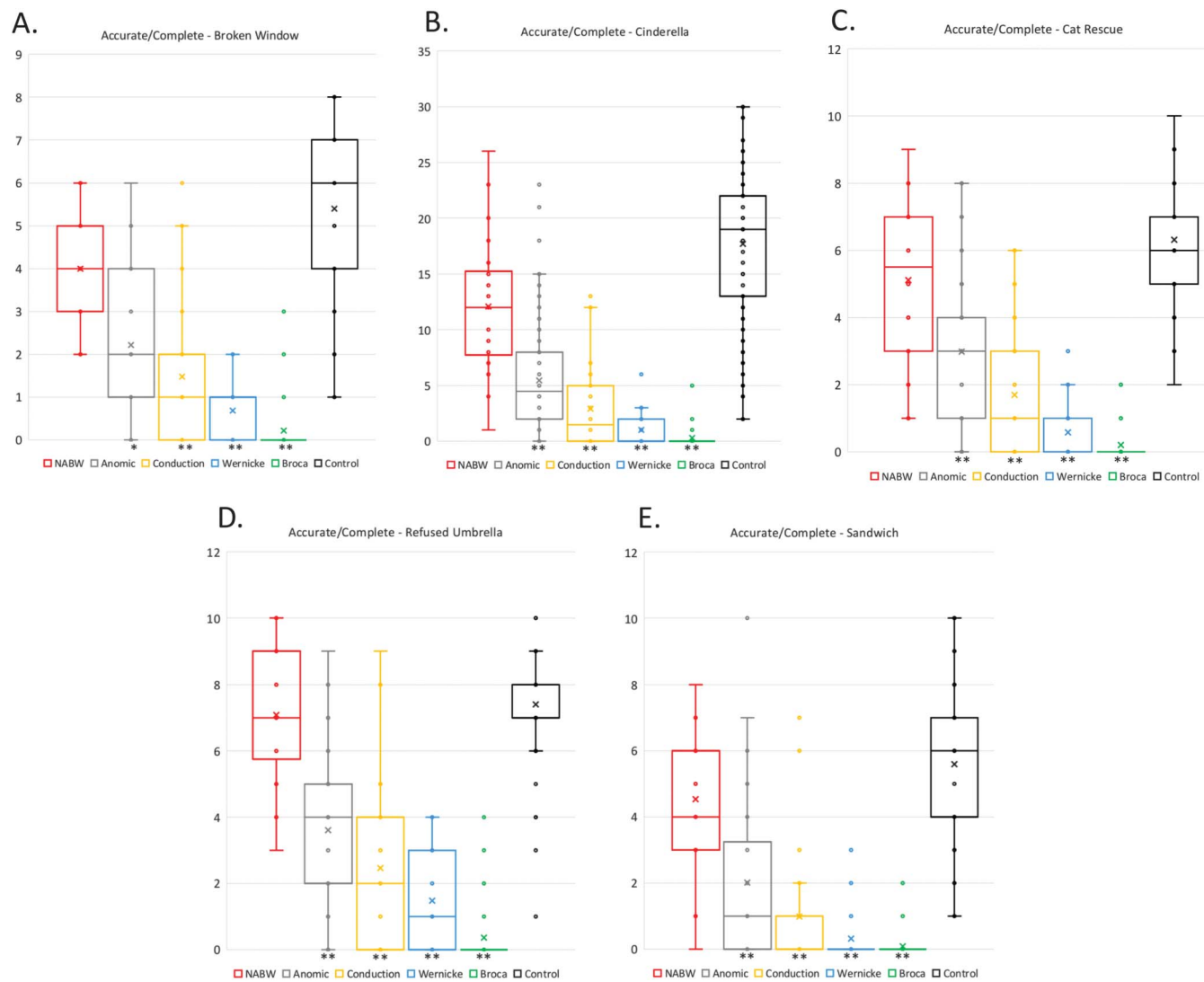
and *Sandwich,* and four for *Cinderella.* It is important to mention here that during the *Refused Umbrella* task, one of the PNBIs incorrectly referred to the small child in the picture as a girl rather than a boy, resulting in eight IC statements because the referent and pronouns used throughout referred to an essential element and were inaccurate (see Appendixes A and B in Nicholas & Brookshire, 1995). If this individual is excluded, no PNBIs produced more than two IC statements during *Refused Umbrella.*

II. Examining the descriptive statistics for this code shows that PWAs generally had more II codes than PNBIs, with a wider range. Relatively few statements were judged II in the PNBI and PWA samples, leading to a restricted range. Results of the II code analysis revealed a significant

difference between PNBIs and the entire group of PWAs for *Cinderella* and *Sandwich* tasks (all $p < .001$), with medium effect sizes ($\phi = 0.42$ and 0.4, respectively)— for these comparisons, PWAs produced more statements judged II than PNBIs (see Appendix B). Individuals with anomic aphasia demonstrated significant differences for the *Broken Window, Cinderella,* and *Sandwich* tasks, with medium effect sizes. For individuals with conduction aphasia, differences were observed for the *Cinderella, Sandwich, Cat Rescue,* and *Refused Umbrella* tasks, with medium and large effect sizes. Finally, individuals with Wernicke's and Broca's aphasia produced significantly more statements judged II on all tasks, with medium and large effect sizes observed (see Figure 6A–E). No

Dalton & Richardson: *MC Production in Aphasia and Controls* **301**

Figure 3. Boxplots of the distribution of accurate/complete codes for all groups for the (A) *Broken Window,* (B) *Cinderella,* (C) *Cat Rescue,* (D) *Refused Umbrella,* and (E) *Sandwich* tasks. A single asterisk below the groups indicates a significant difference between that group and persons not brain injured with a medium effect size. A double asterisk indicates a significant difference with a large effect size. NABW = not aphasic by WAB.

group showed less than 50% overlap of distributions for any story (see Table 3). For the *Broken Window* and *Refused Umbrella* tasks, no PNBIs produced a statement judged II. For the *Cinderella, Cat Rescue,* and *Sandwich* tasks, no PNBIs produced more than a single statement judged II.

AB. The descriptive statistics show that PWAs had more AB codes than PNBIs, with a wider range. Results of the AB code analysis revealed a significant difference between PNBIs and the entire group of PWAs for all stories (all $p < .001$), with medium to large effect sizes (ϕ between 0.35 and 0.53), indicating that PWAs had more statements judged AB than PNBIs (see Appendix B). When examining the results by aphasia subtype, this pattern holds for

individuals with anomic, Broca's, and Wernicke's aphasia (with medium to large effect sizes). Individuals with conduction aphasia had similar results, with significantly more AB statements for all stories except *Broken Window* (conduction). There were no significant comparisons with medium or large effect sizes for individuals NABW, indicating that this group produces roughly the same median number of MCs as PNBIs (see Figure 7A–E). There was less than 50% overlap of the PNBIs distribution with individuals with Broca's (*Broken Window, Cinderella*) and Wernicke's aphasia (*Cinderella;* see Table 3). When examining the percent overlap of the aphasia subtype distributions with PNBIs, no group had less than 50% overlap for any story. PNBIs omitted no more than six MCs for *Broken Window,* seven

Figure 4. Boxplots of the distribution of accurate/incomplete codes for all groups for the (A) *Broken Window,* (B) *Cinderella,* (C) *Cat Rescue,* (D) *Refused Umbrella,* and (E) *Sandwich* tasks. A single asterisk below the groups indicates a significant difference between that group and persons not brain injured with a medium effect size. A double asterisk indicates a significant difference with a large effect size. NABW = not aphasic by WAB.
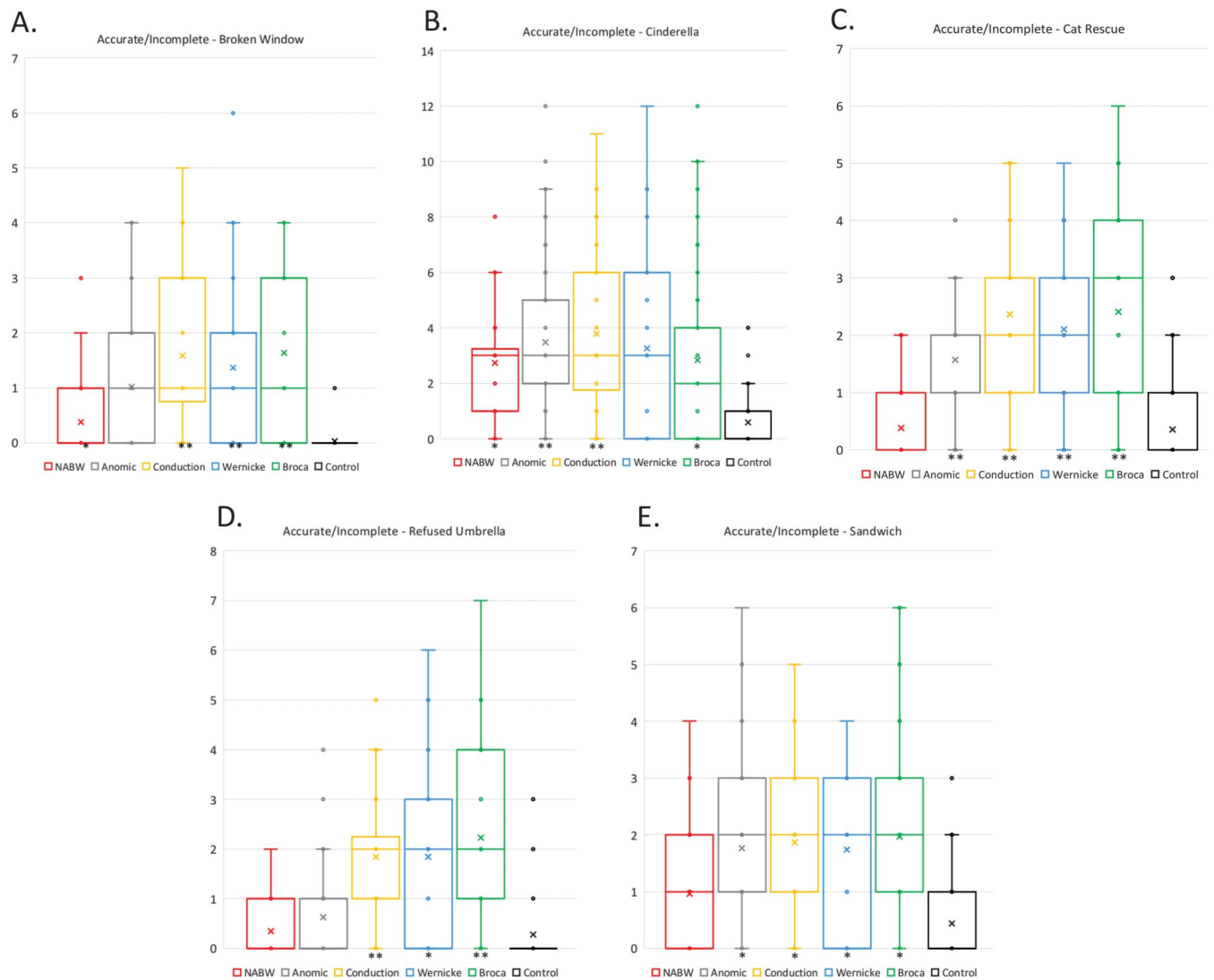

for *Cat Rescue* and *Refused Umbrella,* eight for *Sandwich,* and 31 for *Cinderella.*

Secondary Analyses

Because the aphasia subtypes can be approximately ordered based on severity, from least severe (individuals NABW) to most severe (individuals with Broca's aphasia), and the magnitude of differences tended to increase with severity, we completed a correlation analysis to ensure that our results were not simply due to aphasia severity (see Table 4). The results of the correlation revealed that for each story (except *Refused Umbrella*), aphasia severity as measured by WAB-AQ, was moderately to strongly

positively correlated with MC composite score ($r = .466$ to $r = .717$), the number of MC attempts ($r = .388$ to $r = .658$), and the AC code ($r = .547$ to $r = .684$). WAB-AQ was also strongly negatively correlated with the AB code for all stories except *Refused Umbrella* ($r = -.486$ to $r = -.658$). For codes AI, IC, and II, correlations were small (varying between positive and negative associations) or nonsignificant. This indicates that aphasia severity is indexed by some of the measures, but that in particular, the error codes AI, IC, and II are more likely indexing some specific aspect of language impairment rather than aphasia severity more broadly. We also completed correlations to examine the relationship between scores on each variable across the five tasks. MC composite, MC attempts, and

Dalton & Richardson: *MC Production in Aphasia and Controls* **303**

Figure 5. Boxplots of the distribution of inaccurate/complete codes for all groups for the (A) *Broken Window,* (B) *Cinderella,* (C) *Cat Rescue,* (D) *Refused Umbrella,* and (E) *Sandwich* tasks. A single asterisk below the groups indicates a significant difference between that group and persons not brain injured with a medium effect size. A double asterisk indicates a significant difference with a large effect size. NABW = not aphasic by WAB.
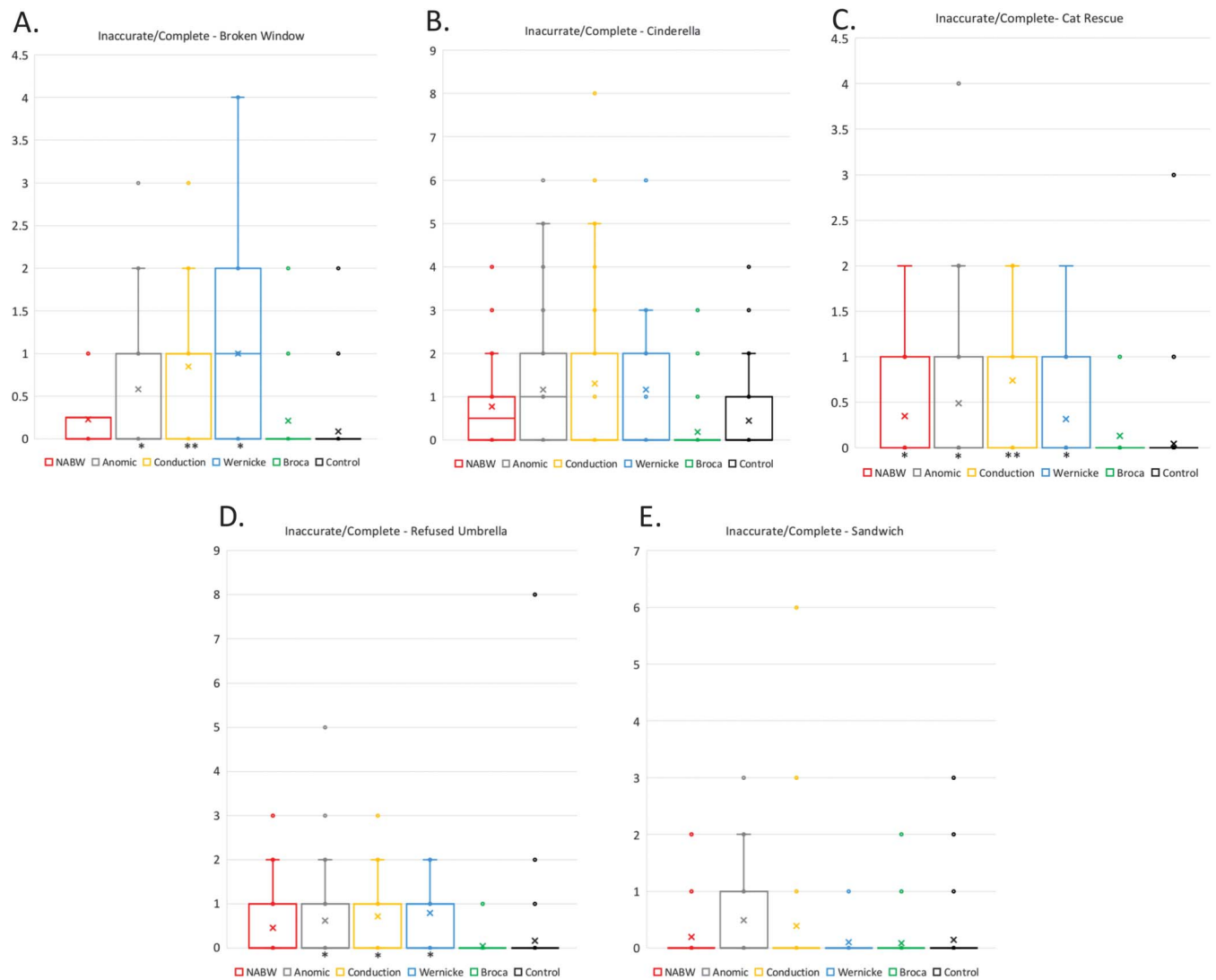

codes AC, AI, and AB showed medium to strong positive correlations for all variables. Correlations for IC were small and positive and, for II, were either small and positive or nonsignificant across all stories (for exact values, see Supplemental Materials S8–S14).

Discussion

Previous research established that persons with both fluent and nonfluent aphasia produce as little as half the information produced by PNBIs (e.g., Gleason et al., 1980; Nicholas & Brookshire, 1993). Our results confirmed this finding for individuals with Broca's and Wernicke's aphasia; however, for individuals NABW or with anomic and

conduction aphasia, informativeness was reduced but not as severely. Additionally, PNBIs in this study performed similarly to those reported by Richardson and Dalton (2016). Nicholas and Brookshire (1995) found that individuals with aphasia produced more inaccurate and AI MCs, had more AB MCs, and fewer AC MCs, which were replicated here. They suggested that the most revealing information at the individual level was inaccuracy or incompleteness (e.g., AI, IC, II) because few PNBIs made these errors; we observed the same tendency in this study. These findings are somewhat consistent with Kong et al. (2016), although they did not find significant differences between PNBIs and individuals with fluent aphasia for production of AI concepts or differences between PNBIs and

Figure 6. Boxplots of the distribution of inaccurate/incomplete codes for all groups for the (A) *Broken Window,* (B) *Cinderella,* (C) *Cat Rescue,* (D) *Refused Umbrella,* and (E) *Sandwich* tasks. A single asterisk below the groups indicates a significant difference between that group and persons not brain injured with a medium effect size. A double asterisk indicates a significant difference with a large effect size. NABW = not aphasic by WAB.
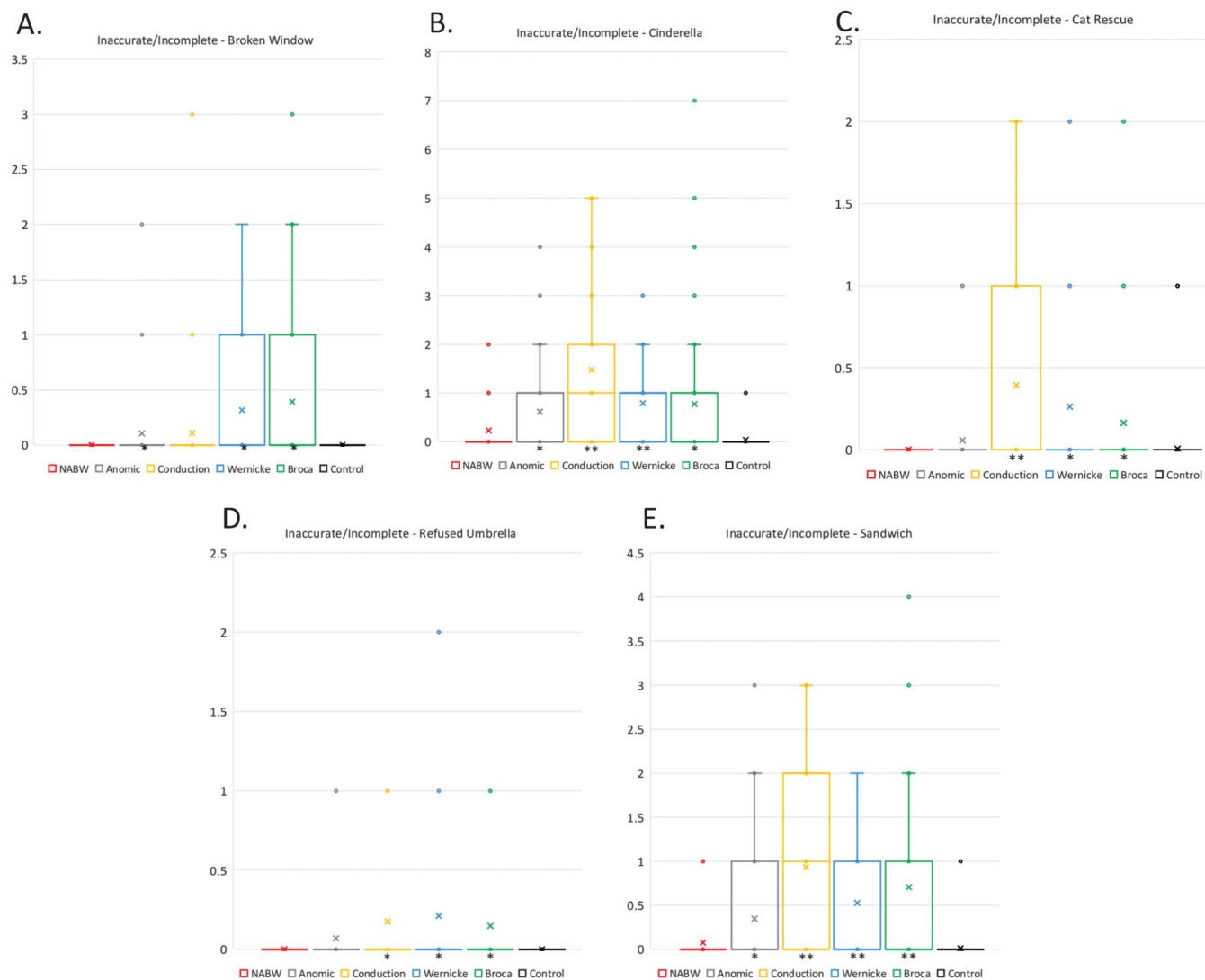
individuals with nonfluent aphasia for inaccurate (IC or II) concepts. Overall, these results demonstrate that MCA is sensitive to differences between PWAs and PNBIs. This is true even for some adults who do not have a clinical diagnosis of aphasia (i.e., NABW) but report residual language difficulties, which negatively affect life participation. However, the magnitude of these differences depended upon aphasia subtype (and perhaps severity) and discourse task.

Sensitivity and Specificity

The utility of diagnostic measures is determined in part by their sensitivity and specificity. A sensitive measure identifies an individual with an impairment as having that

impairment (i.e., few false negatives), whereas one that is specific correctly identifies healthy individuals as not having an impairment (i.e., few false positives; Lalkhen & McCluskey, 2008). To be used for diagnosis, MCA must demonstrate sufficient spread between the distribution of scores for PNBIs and PWAs. While MCA in this study was sensitive to differences between PNBIs and PWAs at the group level, the wide range of "normal" performance resulted in a great deal of overlap in the distributions of PNBIs and the aphasia subtypes. As a result, we observed poor sensitivity and specificity for individual NABW and those with anomic or conduction aphasia for all measures. However, good sensitivity was observed for individuals with Broca's and Wernicke's aphasia on all tasks for the AC code, with good specificity for all tasks (Broca's) or

Figure 7. Boxplots of the distribution of absent codes for all groups for the (A) *Broken Window,* (B) *Cinderella,* (C) *Cat Rescue,* (D) *Refused Umbrella,* and (E) *Sandwich* tasks. A single asterisk below the groups indicates a significant difference between that group and persons not brain injured with a medium effect size. A double asterisk indicates a significant difference with a large effect size. NABW = not aphasic by WAB.
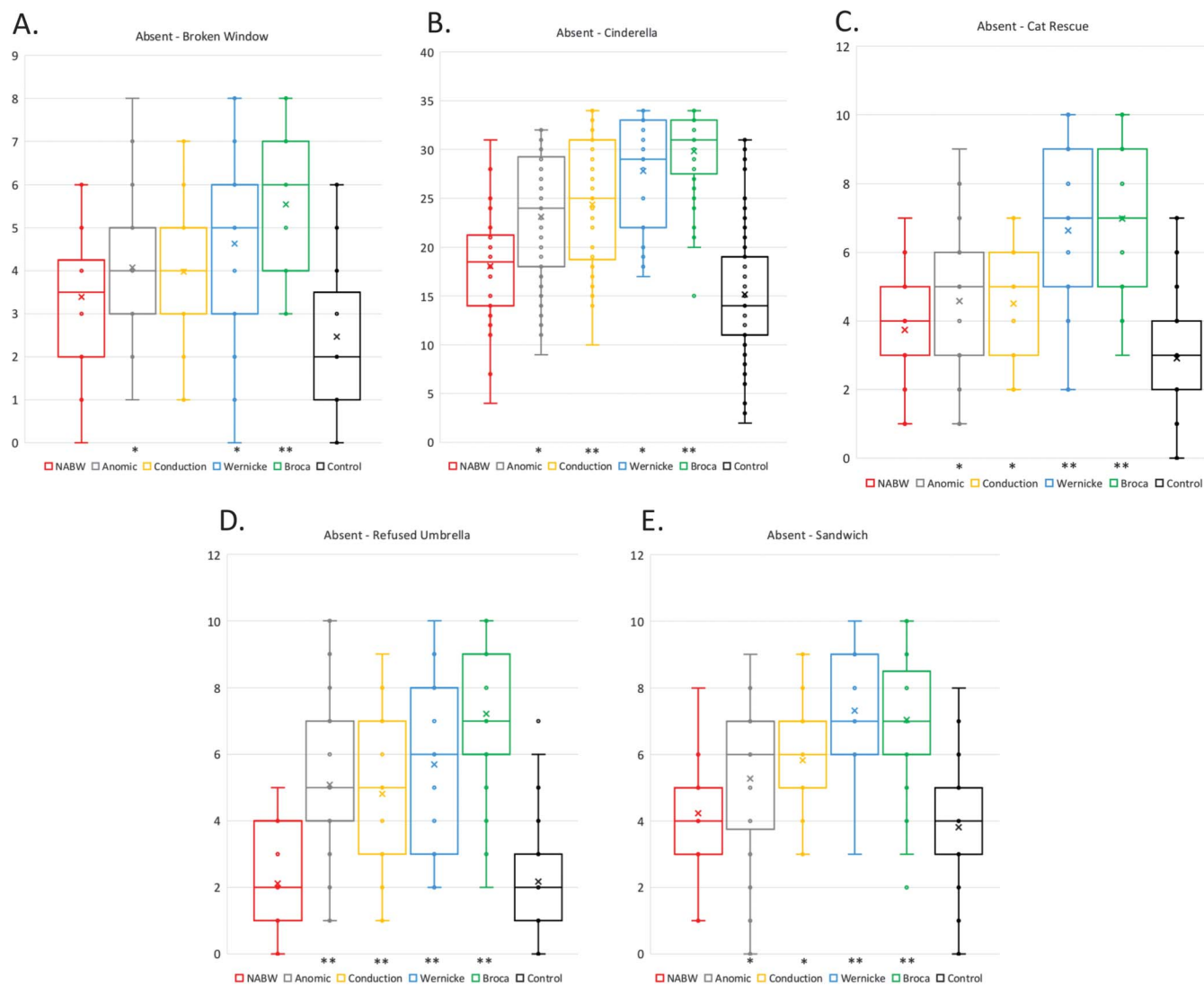

the *Cinderella, Sandwich,* and *Cat Rescue* tasks (Wernicke's). Therefore, MCA in its current state may not be the most appropriate tool to diagnose language impairment at the individual level for all subtypes. Still, researchers and clinicians can use the cutoffs identified in each section as diagnostic indicators of language impairment, given that no PNBIs in the sample scored beyond those cutoffs.

Future research investigating the utility of MCA should examine ways to improve sensitivity and specificity. The MC lists used in this study were created based on a cutoff where a concept was categorized as an MC if produced by 33% of the normative sample. However, the authors also reported MC lists with cutoffs of 50% and 66% of the normative sample, which might yield better

sensitivity. Additionally, previous research with CIUs has shown that the proportion of CIUs and %CIUs/min is generally more sensitive than the raw count of CIUs, and these measures should be investigated for MCA. Finally, it is important to note that on many standardized aphasia assessments, PNBIs perform at or near ceiling, whereas on the MCA, they showed a wide range of performance, indicating that MCA may be useful to gain a more nuanced understanding of language in those without a history of brain injury or disease.

MCA for Treatment Planning

Each of the measures reported here gives important information about an individuals' discourse that may be

Table 4. Correlation between Western Aphasia Battery–Aphasia Quotient and discourse measures for all five tasks.

Variable	Broken Window	Cinderella	Sandwich	Cat Rescue	Refused Umbrella
MC composite	$r = .543$	$r = .631$	$r = .588$	$r = .714$	$r = .466$
	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
Accurate/complete	$r = .642$	$r = .613$	$r = .585$	$r = .664$	$r = .547$
	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
Accurate/incomplete	$r = -.105$	$r = .198$	$r = .013$	$r = -.079$	$r = -.271$
	ns	$p = .002$	ns	ns	$p < .001$
Inaccurate/complete	$r = .040$	$r = .243$	$r = .149$	$r = .162$	$r = .141$
	ns	$p < .001$	$p = .013$	$p = .012$	$p = .02$
Inaccurate/incomplete	$r = -.204$	$r = -.016$	$r = -.211$	$r = -.195$	$r = -.267$
	$p = .001$	ns	$p < .001$	$p = .003$	$p < .001$
Absent	$r = -.549$	$r = -.604$	$r = -.568$	$r = -.661$	$r = -.486$
	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
MC attempts	$r = .651$	$r = .603$	$r = .647$	$r = .661$	$r = .388$
	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$

Note. MC = main concept; ns = not significant.

useful for treatment planning purposes. For example, the MC composite score gives an estimate of the overall informativeness of a discourse sample but does not provide specific information regarding the quality of the discourse. On the other hand, the scoring codes provide information related to the quantity and quality of the sample. Although the composite score is derived from the error codes, there is not a 1:1 correspondence between a composite score and the underlying codes. For example, an MC composite of 9 could be achieved by 3 AC codes, 1 AC and 3 AI codes, 1 AC and 3 IC codes, 4 AI and 1 II codes, 4 IC and 1 II codes, or 9 II codes. The discourse samples underlying each of these possibilities would be markedly different.

For example, a transcript including a large number of statements scored as AC with relatively few scored AB would indicate a mostly complete story, likely falling within the range of performance seen by PNBIs. The reverse scenario, where most MCs were scored AB with very few scored AC, would indicate a paucity of information and might fall below the performance of PNBIs. For both of these scenarios, the MC composite score would serve as a good proxy measure for both the quality and quantity of discourse. However, when one or more of the codes indicating an error is assigned to a discourse, the MC composite becomes more difficult to interpret. In these instances, it may be better to examine the breakdown of how many statements receive each code, rather than the composite score. For example, a discourse sample that receives a large number of AI statements might be generally informative but would likely exhibit more errors of syntax and more abandoned utterances (see Table 2). A sample of this kind might still retain overall flow and could still be recognizable as the intended story. In contrast, a discourse sample with a high proportion of statements scored IC would likely have fewer syntactic errors but would be more difficult to follow because of inaccuracies in semantic content. Finally, a sample with a large proportion of statements scored II would exhibit syntactic errors and would be

confusing and difficult to follow, in the extreme case barely recognizable as the intended story. Given this, MCA also has the potential to inform about the coherence and cohesion of discourse because inaccurate or incomplete statements may lack linking elements (e.g., pronoun referents or temporal sequencing) that can be targeted in therapy.

Practically, an individual with more AI than IC statements would likely benefit from treatment that focused on increasing output of specific types of lexical items. For example, if the AI codes were a result of a patient omitting or producing only vague verbs, then a treatment targeting verb production might be appropriate. If an individual has a higher proportion of inaccurate statements, then therapy with barrier tasks or a focus on providing accurate information might be best. Finally, if an individual has a large proportion of AB codes, then a clinician might want to focus on increasing overall output. This of course does not cover all possible appropriate treatment approaches, and the question of which treatments might result in improvement in discourse measures such as MCA is an important one for our field to continue to pursue.

Discourse Tasks

Given previous reports that the type of discourse task and instructions used to elicit discourse can result in productions of different length and quality (e.g., Olness, 2006; Wright & Capilouto, 2009), it is important to consider the impact of these elicitation procedures and stimuli on PWA performance. Looking at the number of significant differences, *Cat Rescue* and *Broken Window* appear to be the most sensitive to differences across all subtypes, with 36 and 37 statistically significant comparisons, respectively, followed by *Refused Umbrella* and *Cinderella* with 33 statistically significant comparisons, and *Sandwich* with 31. However, when considering effect sizes, *Cinderella* had the greatest number of large effects (20) followed by *Cat Rescue* and *Refused Umbrella* (19), then *Broken Window* (16) and, finally,

Dalton & Richardson: *MC Production in Aphasia and Controls* **307**

Sandwich (15). For clinicians or researchers who are operating under time and/or personnel constraints, choosing either the *Cinderella, Cat Rescue,* or *Refused Umbrella* tasks may allow for a small but sensitive discourse sample.

Although we report results for each story separately, measures calculated from longer speech samples are likely to be more stable and reliable across time (Boyle, 2014, 2015; Brookshire & Nicholas, 1994a, 1994b). Future research should investigate which combination of these tasks yields the greatest sensitivity and stability while minimizing the time needed to administer and score. However, even at the individual story level, statistically and practically significant differences were apparent for all PWAs (and most subtypes) compared with PNBIs.

Different storylines were produced during *Refused Umbrella* by some participants. The main initiating actions reported by the majority of participants involved the mother offering her son an umbrella and the boy refusing it. However, a minority of participants (a few PNBIs and a larger number of PWAs) produced narratives where the boy requested the umbrella and the mother refused to let him have it. It is not clear whether this picture stimulus is more ambiguous than the other tasks, more reliant on intact syntactic abilities, or if perhaps the *Refused Umbrella* task is associated with a higher cognitive load, but this warrants further examination.

Multidimensional Discourse Assessment

Although our results demonstrate some sensitivity of MCA to identifying difficulties relating to verbally conveying the gist or essential elements of a story, MCA does not take into account the full richness of the samples produced. Only spoken statements that pertain to the identified MCs are examined, potentially leaving a great deal of the sample unanalyzed (e.g., other relevant concepts, MCs communicated via gesture). We echo previous researchers (e.g., E. Armstrong, 2000; Linnik et al., 2016) who have urged the use of multidimensional or multilevel discourse analyses that can leverage a larger proportion of the sample, such as reported by Marini, Andreetta, Del Tin, and Carlomagno (2011). Another potential avenue to link MCA with higher level discourse features is the Story Goodness Index (SGI; Lê, Coelho, Mozeiko, & Grafman, 2011), which estimates the organization and completeness of a discourse. The organization is quantified by story grammar episodes. Story completeness is scored, essentially, as statements that receive an AC code during MC analysis. Using the SGI may allow clinicians and researchers to quickly quantify performance on macrostructural and superstructural levels, especially if a story's episodes consist primarily of information captured by the established MCs. To date, the SGI has been used primarily with individuals with traumatic brain injury, so research would be needed to confirm its utility in individuals with stroke-induced deficits. Further development of SGI, using the more nuanced MCA codes, may improve both the diagnostic and outcome tracking usefulness of this measure. Listener ratings may also provide valuable information regarding the acceptability of the produced discourse to conversational partners. It is our hope that the results reported here will contribute to the use of MCA as part of a multilevel discourse analysis.

Limitations

Although the current study leveraged the AphasiaBank to conduct large-scale comparisons, there are limitations associated with using such a database. First, results are only as reliable as the data being contributed. Previous research has investigated the assessment and transcription fidelity of AphasiaBank and found that both are excellent (Richardson & Dalton, 2016), indicating that we can be confident the results reported here reflect true performance by PNBIs and PWAs. Second, when utilizing a database, one is limited to the measures that are included as part of that database. Although the tasks included in the AphasiaBank protocol are widely used, they are not the only frequently used tasks and stimuli; thus, results may only be useful for researchers and clinicians who utilize the protocol. Third, the AphasiaBank database utilizes WAB-AQ (Kertesz, 1982) scores to determine aphasia diagnosis and type. However, the WAB and WAB-R use a subjective fluency rating with questionable reliability to determine AQ (e.g., Hillis-Trupe, 1984; Hula, Donovan, Kendall, & Gonzalez-Rothi, 2010), so interpretation and application of subtype findings should be completed with classification accuracy in mind. Although the WAB-R-AQ introduces limitations, it is widely used, is suggested as a core outcome measure for aphasia (Wallace, Worrall, Rose, & Le Dorze, 2016), and remains a useful tool to examine differences among aphasia subtypes and communicate those results to a wide audience.

Although we were able to obtain large samples for PNBIs and PWAs as a whole group, the sample sizes for the aphasia subtypes varied. Sample sizes for normative data should be greater than 50 (Mitrushina, Boone, Razani, & D'Elia, 2005) or even 100 (American Psychological Association, 1999). The samples for individuals with Broca's and anomic aphasia (and, perhaps, conduction aphasia) are sufficiently large to serve as adequate normative samples, but the samples of individuals NABW and with Wernicke's aphasia are small, and individuals with transcortical and global aphasia were excluded altogether for the subtype analysis. As the AphasiaBank database continues to grow and representation of these groups increases, these results should be updated to ensure appropriate sample sizes are used and that norms are available for all aphasia subtypes.

Finally, the discourse tasks used were semi-spontaneous tasks rather than conversation. Although this improves replicability and makes norming possible, it potentially limits insights into functional, everyday communication. Given previous research, which has shown strong correlations between informativeness measures and conversation (Doyle et al., 1995) and listener perceptions of aphasic speech (Cupit et al., 2010, K. B. Ross & Wertz, 1999), we

feel that this is an appropriate compromise. Because these participants completed the discourse tasks once, it is not possible to determine test–retest reliability for this sample. However, MCA reportedly has good test–retest reliability at 2 weeks (Kong et al., 2016), 1 month (Nicholas & Brookshire, 1995), and 1 year (Kong, 2011), suggesting that test–retest reliability may be sufficient to use as a treatment outcome measure, if adequate sensitivity and specificity are also achieved.

Conclusions

Aphasiology has seen recent shifts in primary outcome measures (Brady et al., 2016) and calls for improved discourse measures that are stable and reliable for research and clinical use (E. Armstrong, 2000; Boyle, 2014, 2015; Brady et al., 2016; Linnik et al., 2016; Pritchard et al., 2017). We sought to contribute information regarding a measure that is informative, reliable, and has potential for clinical utilization. We also sought to increase investigative and clinical utility by providing descriptive and comparative information for PNBIs and PWAs for several standardized discourse tasks with detailed checklists that enable reliable scoring. The checklists, combined with the scoring procedures and rules developed by Nicholas and Brookshire (1995), and the information reported here, may support further development of MCA as a valuable source of information for diagnosis and treatment outcomes for PWAs following additional research. This is a measure that may be reliable enough for non-transcription–based scoring, which would further promote clinical adoption. We demonstrated that MCA is sensitive to group differences between PNBIs and PWAs, and the extension of these results to other clinical populations (like traumatic brain injury and dementia) are avenues of research that should be explored. Future research should also investigate ways to improve sensitivity, specificity, and reliability.

Acknowledgments

References

Albright, E., & Purves, B. (2008). Exploring SentenceShaperTM: Treatment and augmentative possibilities. *Aphasiology, 22*(7–8), 741–752.

American Psychological Association. (1999). *Standards for education and psychological tests*. Washington, DC: American Educational Research Association.

Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology, 14*(9), 875–892.

Armstrong, E., & Ferguson, A. (2010). Language, meaning, context, and functional communication. *Aphasiology, 24*(4), 480–496.

Armstrong, L., Brady, M., Mackenzie, C., & Norrie, J. (2007). Transcription-less analysis of aphasic discourse: A clinician's dream or a possibility? *Aphasiology, 21*(3–4), 355–374.

Avent, J., & Austermann, S. (2003). Reciprocal scaffolding: A context for communication treatment in aphasia. *Aphasiology, 17*(4), 397–404.

Boyle, M. (2014). Test–retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and Hearing Research, 57*(3), 966–978.

Boyle, M. (2015). Stability of word-retrieval errors with the AphasiaBank stimuli. *American Journal of Speech-Language Pathology, 24*(4), S953–S960.

Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews, 6,* CD000425. https://doi.org/10.1002/14651858.CD000425.pub4

Brookshire, R. H., & Nicholas, L. E. (1994a). Speech sample size and test–retest stability of connected speech measures for adults with aphasia. *Journal of Speech and Hearing Research, 37,* 399–407.

Brookshire, R. H., & Nicholas, L. E. (1994b). Test-retest stability of measures of connected speech in aphasia. *Clinical Aphasiology, 22,* 119–133.

Capilouto, G. J., Wright, H. H., & Wagovich, S. A. (2006). Reliability of main event measurement in the discourse of individuals with aphasia. *Aphasiology, 20*(2–4), 205–216.

Coelho, C. A., McHugh, R. E., & Boyle, M. (2000). Semantic feature analysis as a treatment for aphasic dysnomia: A replication. *Aphasiology, 14*(2), 133–142.

Cupit, J., Rochon, E., Leonard, C., & Laird, L. (2010). Social validation as a measure of improvement after aphasia treatment: Its usefulness and influencing factors. *Aphasiology, 24*(11), 1486–1500.

Dalton, S. G., & Richardson, J. D. (2015). Core-lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. *American Journal of Speech-Language Pathology, 24*(4), S923–S938.

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.

Doyle, P. J., Goda, A. J., & Spencer, K. A. (1995). The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. *American Journal of Speech-Language Pathology, 4*(4), 130–134.

Doyle, P. J., McNeil, M. R., Park, G., Goda, A., Rubenstein, E., Spencer, K., . . . Szwarc, L. (2000). Linguistic validation of four parallel forms of a story retelling procedure. *Aphasiology, 14*(5–6), 537–549.

Doyle, P. J., McNeil, M. R., Spencer, K. A., Goda, A. J., Cottrell, K., & Lustig, A. P. (1998). The effects of concurrent picture presentations on retelling of orally presented stories by adults with aphasia. *Aphasiology, 12*(7–8), 561–574.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2–18.

Fromm, D., Forbes, M., Holland, A., Dalton, S. G., Richardson, J., & MacWhinney, B. (2017). Discourse characteristics in aphasia beyond the Western aphasia battery cutoff. *American Journal of Speech-Language Pathology, 26*(3), 762–768.

Fromm, D., Forbes, M., Holland, A., & MacWhinney, B. (2013). *PWAs and PBJs: Language for describing a simple procedure.* Paper presented at the Clinical Aphasiology Conference, Tucson, AZ.

Gleason, J. B., Goodglass, H., Obler, L., Green, E., Hyde, M. R., & Weintraub, S. (1980). Narrative strategies of aphasic and normal-speaking subjects. *Journal of Speech and Hearing Research, 23*(2), 370–382.

Hilari, K. (2011). The impact of stroke: Are people with aphasia different to those without? *Disability and Rehabilitation, 33*(3), 211–218.

Hilari, K., Northcott, S., Roy, P., Marshall, J., Wiggins, R. D., Chataway, J., & Ames, D. (2010). Psychological distress after stroke and aphasia: The first six months. *Clinical Rehabilitation, 24*(2), 181–190.

Hillis-Trupe, A. E. (1984, May). *Reliability of rating spontaneous speech in the Western Aphasia Battery: Implications for classification.* Clinical Aphasiology: Proceedings of the Conference 1984, Seabrook Island, SC, BRK Publishers.

Huber, W. (1984). The Aachen Aphasia Test. *Advanced Neurology, 42,* 291–303.

Hula, W., Donovan, N. J., Kendall, D. L., & Gonzalez-Rothi, L. J. (2010). Item response theory analysis of the Western aphasia battery. *Aphasiology, 24*(11), 1326–1341.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test.* Austin, TX: Pro-Ed.

Kelly, H., Brady, M. C., & Enderby, P. (2010). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews, 5,* CD000425. https://doi.org/10.1002/14651858.CD000425.pub2

Kertesz, A. (1982). *The Western Aphasia Battery.* New York, NY: Grune & Stratton.

Kong, A. P. H. (2009). The use of main concept analysis to measure discourse production in Cantonese-speaking persons with aphasia: A preliminary report. *Journal of Communication Disorders, 42*(6), 442–464.

Kong, A. P. H. (2011). The main concept analysis in Cantonese aphasic oral discourse: External validation and monitoring chronic aphasia. *Journal of Speech, Language, and Hearing Research, 54*(1), 148–159.

Kong, A. P. H., Whiteside, J., & Bargmann, P. (2016). The main concept analysis: Validation and sensitivity in differentiating discourse produced by unimpaired English speakers from individuals with aphasia and dementia of Alzheimer type. *Logopedics Phoniatrics Vocology, 41*(3), 129–141.

Lalkhen, A. G., & McCluskey, A. (2008). Clinical tests: Sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care & Pain, 8*(6), 221–223.

Larfeuil, C., & Le Dorze, G. (1997). An analysis of the word-finding difficulties and of the content of the discourse of recent and chronic aphasic speakers. *Aphasiology, 11,* 783–811.

Lê, K., Coelho, C., Mozeiko, J., & Grafman, J. (2011). Measuring goodness of story narratives. *Journal of Speech, Language, and Hearing Research, 54*(1), 118–126.

Linnik, A., Bastiaanse, R., & Höhle, B. (2016). Discourse production in aphasia: A current review of theoretical and methodological challenges. *Aphasiology, 30*(7), 765–800.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*(11), 1286–1307.

Marini, A., Andreetta, S., Del Tin, S., & Carlomagno, S. (2011). A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology, 25*(11), 1372–1392.

McNeil, M. R., Doyle, P. J., Fossett, T. R., Park, G. H., & Goda, A. J. (2001). Reliability and concurrent validity of the information unit scoring metric for the story retelling procedure. *Aphasiology, 15*(10–11), 991–1006.

Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (Eds). (2005). *Handbook of normative data for neuropsychological assessment.* Oxford, United Kingdom: Oxford University Press.

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research, 36*(2), 338–350.

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 and Hearing Research, 38*(1), 145–156.

Olness, G. S. (2006). Genre, verb, and coherence in picture-elicited discourse of adults with aphasia. *Aphasiology, 20,* 175–187.

Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2017). Reviewing the quality of discourse information measures in aphasia. *International Journal of Language & Communication Disorders, 52*(6), 689–732.

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.

Richardson, J. D., & Dalton, S. G. (2019). Main concepts for two picture description tasks: An addition to Richardson & Dalton, 2016. *Aphasiology,* Advance online publication. https://doi.org/10.1080/02687038.2018.1561417

Richardson, J. D., Dalton, S. G. H., Fromm, D., Forbes, M., Holland, A., & MacWhinney, B. (2018). The relationship between confrontation naming and story gist production in aphasia. *American Journal of Speech-Language Pathology, 27,* 406–422.

Robey, R. R. (1998). A meta-analysis of clinical outcomes in the treatment of aphasia. *Journal of Speech, Language, and Hearing Research, 41*(1), 172–187.

Ross, K., & Wertz, R. (2003). Quality of life with and without aphasia. *Aphasiology, 17*(4), 355–364.

Ross, K. B., & Wertz, R. T. (1999). Comparison of impairment and disability measures for assessing severity of, and improvement in, aphasia. *Aphasiology, 13,* 113–124.

Simmons-Mackie, N., Threats, T. T., & Kagan, A. (2005). Outcome assessment in aphasia: A survey. *Journal of Communication Disorders, 38*(1), 1–27.

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.

Thompson, C. K. (2011). *Northwestern Assessment of Verbs and Sentences.* Evanston, IL: Northwestern University.

Ulatowska, H. K., Freedman-Stern, R., Doyel, A. W., Macaluso-Haynes, S., & North, A. J. (1983). Production of narrative discourse in aphasia. *Brain and Language, 19*(2), 317–334.

van Dijk, T. A. (Ed.). (1997). *Discourse as social interaction* (Vol. 2). London, United Kingdom: Sage.

Wallace, S. J., Worrall, L., Rose, T., & Le Dorze, G. (2016). Core outcomes in aphasia treatment research: An e-Delphi consensus study of international aphasia researchers. *American Journal of Speech-Language Pathology, 25*(4S), S729–S742.

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Cruice, M., Isaksen, J., . . . Gauvreau, C. A. (2017). Which outcomes are most important to people with aphasia and their families? An international nominal group technique study framed within the ICF. *Disability and Rehabilitation, 39*(14), 1364–1379.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Thousand Oaks, CA: Sage Publications, Inc.

Wright, H. H., & Capilouto, G. J. (2009). Manipulating task instructions to change narrative discourse performance. *Aphasiology, 23*(10), 1295–1308.

Yorkston, K. M., & Beukelman, D. R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders, 45*(1), 27–36.

Appendix A (p. 1 of 5)

Descriptive Statistics for Each Story

Table A1. *Broken Window.*

Descriptive statistic	PNBI	PWA	NABW	ANO	CON	WER	BRO
Accurate/complete							
M	5.4	1.6	4.0	2.2	1.5	0.7	0.2
SD	1.6	1.8	1.3	1.7	1.7	0.7	0.6
Mdn	6	1	4	2	1	1	0
Range	1–8	0–6	2–6	0–6	0–6	0–2	0–3
Skew	−0.396	0.807	−0.128	0.255	1.224	0.616	3.419
Kurtosis	−0.302	−0.557	−0.933	−1.054	0.862	−0.856	11.852
Accurate/incomplete							
M	0.03	1.2	0.3	1.0	1.6	1.4	1.6
SD	0.2	1.3	0.7	1.1	1.3	1.8	1.4
Mdn	0	1	0	1	1	1	1
Range	0–1	0–6	0–3	0–4	0–5	0–6	0–4
Skew	5.156	0.872	2.257	0.888	0.516	1.317	0.365
Kurtosis	24.928	0.024	5.289	0.012	−0.489	1.129	−1.057
Inaccurate/complete							
M	0.1	0.5	0.2	0.6	0.8	1.0	0.2
SD	0.3	0.8	0.4	0.8	0.8	1.3	0.5
Mdn	0	0	0	0	1	1	0
Range	0–2	0–4	0–1	0–3	0–3	0–4	0–2
Skew	3.6	1.710	1.358	1.425	0.767	1.385	2.285
Kurtosis	13.3	2.995	−0.177	1.403	0.044	1.288	4.686
Inaccurate/incomplete							
M	0.0	0.2	0.0	0.1	0.1	0.3	0.4
SD	0.0	0.5	0.0	0.3	0.5	0.6	0.7
Mdn	0	0	0	0	0	0	0
Range	0–0	0–3	0–0	0–2	0–3	0–2	0–3
Skew	0.000	3.324	0.000	3.475	5.295	1.766	2.096
Kurtosis	0.000	12.458	0.000	12.592	30.252	2.540	4.554
Absent							
M	2.5	4.4	3.4	4.1	4.0	4.6	5.5
SD	1.5	1.7	1.5	1.4	1.3	2.4	1.6
Mdn	2	4	3.5	4	4	5	6
Range	0–6	0–8	0–6	1–8	1–7	0–8	3–8
Skew	0.353	0.353	−0.276	0.165	−0.074	−0.474	0.172
Kurtosis	−0.469	−0.469	−0.270	−0.411	0.240	−0.786	−1.147
MC attempts							
M	5.5	3.6	4.6	3.9	4.0	3.7	2.5
SD	1.5	1.7	1.5	1.4	1.3	2.4	1.6
Mdn	6	4	4.5	4	4	3	2
Range	2–8	0–8	3–8	0–7	1–7	0–8	0–5
Skew	−0.348	−0.138	0.276	−0.165	0.074	0.474	−0.172
Kurtosis	−0.521	−0.171	−0.270	−0.411	0.240	−0.786	−1.147
MC composite							
M	16.8	8.7	13.2	10.0	9.4	7.1	4.7
SD	4.5	4.8	4.1	4.1	4.0	5.2	3.3
Mdn	18	8	12.5	9	9	6	5
Range	5–24	0–20	6–20	0–18	2–20	0–17	0–13
Skew	−0.389	0.209	−0.036	0.015	0.732	0.432	0.099
Kurtosis	−0.521	−0.476	−0.804	−0.706	0.790	−0.861	−0.770

Note. PNBI = persons not brain injured; PWA = persons with aphasia; NABW = not aphasic by Western Aphasia Battery; ANO = anomic; CON = conduction; WER = Wernicke's; BRO = Broca's; MC = main concept.

Dalton & Richardson: *MC Production in Aphasia and Controls* **311**

Descriptive Statistics for Each Story

Table A2. *Refused Umbrella*.

Descriptive statistic	PNBI	PWA	NABW	ANO	CON	WER	BRO
Accurate/complete							
M	7.4	2.8	7.1	3.6	2.5	1.5	0.4
SD	1.6	2.7	1.9	2.3	2.3	1.4	0.83
Mdn	8	2	7	4	2	1	0
Range	1–10	0–10	3–10	0–9	0–9	0–4	0–4
Skew	−1.172	0.709	−0.425	0.005	0.950	0.570	2.572
Kurtosis	1.933	−0.493	−0.516	−0.641	0.799	−1.117	6.652
Accurate/incomplete							
M	0.3	1.3	0.3	0.6	1.8	1.8	2.2
SD	0.5	1.5	0.6	0.9	1.2	1.8	1.8
Mdn	0	1	0	0	2	2	2
Range	0–3	0–7	0–2	0–4	0–5	0–6	0–7
Skew	2.134	1.088	1.403	1.526	0.307	0.863	0.376
Kurtosis	4.990	0.594	1.216	1.815	0.081	−0.045	−0.782
Inaccurate/complete							
M	0.2	0.5	0.5	0.6	0.7	0.8	0.0
SD	0.7	0.8	0.8	0.9	0.8	0.8	0.2
Mdn	0	0	0	0	1	1	0
Range	0–8	0–5	0–3	0–5	0–3	0–2	0–1
Skew	8.732	2.019	1.837	2.093	1.100	0.410	4.275
Kurtosis	90.283	5.351	2.935	6.033	1.023	−1.208	16.830
Inaccurate/incomplete							
M	0.0	0.1	0.0	0.1	0.2	0.2	0.1
SD	0.0	0.3	0.0	0.3	0.4	0.5	0.4
Mdn	0	0	0	0	0	0	0
Range	0–0	0–2	0–0	0–1	0–1	0–2	0–1
Skew	0.000	2.800	0.000	3.438	1.779	2.658	2.038
Kurtosis	0.000	7.132	0.000	10.052	1.216	6.883	2.226
Absent							
M	2.2	5.3	2.1	5.1	4.8	5.7	7.2
SD	1.4	2.6	1.5	2.2	2.2	2.4	2.0
Mdn	2	5	2	5	5	6	7
Range	0–7	0–10	0–5	1–10	1–9	2–10	2–10
Skew	0.880	0.010	0.246	0.283	−0.068	−0.012	−0.388
Kurtosis	1.043	−0.844	−1.064	−0.634	−1.105	−1.030	−0.329
MC attempts							
M	7.8	4.7	7.9	4.9	5.2	4.3	2.8
SD	1.4	2.6	1.5	2.2	2.2	2.4	2.0
Mdn	8	5	8	5	5	4	3
Range	3–10	0–10	5–10	0–9	1–9	0–8	0–8
Skew	−0.880	−0.010	−0.246	−0.283	0.068	0.012	0.388
Kurtosis	1.043	−0.844	−1.064	−0.634	−1.105	−1.030	−0.329
MC composite							
M	23.0	12.1	22.8	13.4	12.7	9.9	5.8
SD	4.2	7.5	4.7	6.5	6.4	6.1	4.4
Mdn	24	11	24	14	12	10	6
Range	9–30	0–30	13–30	0–27	2–27	0–19	0–20
Skew	−0.868	0.292	−0.355	−0.190	0.358	−0.018	0.624
Kurtosis	0.893	−0.773	−0.877	−0.665	−0.764	−1.178	0.670

Note. PNBI = persons not brain injured; PWA = persons with aphasia; NABW = not aphasic by Western Aphasia Battery; ANO = anomic; CON = conduction; WER = Wernicke's; BRO = Broca's; MC = main concept.

Descriptive Statistics for Each Story

Table A3. *Cat Rescue.*

Descriptive statistic	PNBI	PWA	NABW	ANO	CON	WER	BRO
Accurate/complete							
M	6.3	2.1	5.1	3	1.7	0.6	0.2
SD	1.5	2.3	2.2	2.1	1.8	0.9	0.5
Mdn	6	1	5.5	3	1	0	0
Range	2–10	0–9	1–9	0–8	0–6	0–3	0–2
Skew	−0.231	0.940	−0.285	0.396	0.977	1.517	2.629
Kurtosis	−0.322	−0.122	−0.721	−0.373	−0.183	1.593	6.143
Accurate/incomplete							
M	0.7	2.1	0.8	1.9	2.7	2.2	2.5
SD	0.8	1.4	0.9	1.1	1.3	1.7	1.7
Mdn	1	2	1	2	2	2	3
Range	0–3	0–6	0–3	0–4	0–5	0–6	0–6
Skew	0.953	0.423	1.132	0.154	0.265	0.590	0.085
Kurtosis	0.578	−0.413	0.953	−0.460	−0.720	−0.270	−0.915
Inaccurate/complete							
M	0.04	0.4	0.3	0.5	0.7	0.3	0.1
SD	0.3	0.7	0.6	0.7	0.7	0.6	0.3
Mdn	0	0	0	0	1	0	0
Range	0–3	0–4	0–2	0–4	0–2	0–2	0–1
Skew	8.617	1.678	1.403	1.858	0.468	1.766	2.241
Kurtosis	82.985	3.364	1.216	4.711	−1.027	2.540	3.123
Inaccurate/incomplete							
M	0.007	0.2	0.0	0.06	0.4	0.3	0.2
SD	0.08	0.4	0.0	0.2	0.6	0.6	0.4
Mdn	0	0	0	0	0	0	0
Range	0–1	0–2	0–0	0–1	0–2	0–2	0–2
Skew	12.042	2.578	0.000	3.844	1.174	2.158	2.554
Kurtosis	145.000	6.256	0.000	13.079	0.472	4.253	6.302
Absent							
M	2.9	5.2	3.7	4.6	4.5	6.6	7
SD	1.4	2.2	1.6	1.7	1.5	2.5	1.9
Mdn	3	5	4	5	5	7	7
Range	0–7	1–10	1–7	1–9	2–7	2–10	3–10
Skew	0.554	0.374	0.042	0.231	0.097	−0.341	0.096
Kurtosis	0.185	−0.352	−0.465	−0.206	−1.003	−0.675	−1.047
MC attempts							
M	7.1	4.8	6.3	5.4	5.5	3.3	3
SD	1.4	2.2	1.6	1.7	1.5	2.5	1.9
Mdn	7	5	6	5	5	3	3
Range	3–10	0–9	3–9	1–9	3–8	0–8	0–7
Skew	−0.554	−0.374	−0.042	−0.231	−0.097	0.341	0.096
Kurtosis	−0.185	−0.352	−0.465	−0.206	−1.003	−0.675	−1.047
MC composite							
M	20.5	11.4	17.7	13.8	12.3	7.1	6.1
SD	4.2	6.3	5.3	5.3	4.7	5.6	4
Mdn	21	11	17.5	14	12	6	6
Range	8–30	0–27	7–27	2–26	5–22	0–18	0–14
Skew	−0.461	0.161	−0.137	0.018	0.244	0.466	−.009
Kurtosis	0.076	−0.551	−0.501	−0.338	−0.991	−0.738	−1.083

Note. PNBI = persons not brain injured; PWA = persons with aphasia; NABW = not aphasic by Western Aphasia Battery; ANO = anomic; CON = conduction; WER = Wernicke's; BRO = Broca's; MC = main concept.

Descriptive Statistics for Each Story

Table A4. *Cinderella.*

Descriptive statistic	PNBI	PWA	NABW	ANO	CON	WER	BRO
Accurate/complete							
M	17.7	4	12.1	5.4	2.9	1	0.3
SD	6.1	5.3	5.9	5	3.5	1.6	0.8
Mdn	19	2	12	4.5	1.5	0	0
Range	2–30	0–26	1–26	0–23	0–13	0–6	0–5
Skew	−0.521	1.632	0.464	1.318	1.295	2.006	3.950
Kurtosis	−0.160	2.502	0.188	1.808	1.152	4.389	18.663
Accurate/incomplete							
M	0.6	3.3	2.7	3.5	3.8	3.3	2.8
SD	0.9	2.8	1.9	2.6	2.8	3.7	3.2
Mdn	0	3	3	3	3	3	2
Range	0–4	0–12	0–8	0–12	0–11	0–12	0–12
Skew	1.339	0.998	1.025	0.981	0.589	0.903	1.349
Kurtosis	1.228	0.539	1.483	0.580	−0.332	0.016	1.169
Inaccurate/complete							
M	0.4	0.9	0.8	1.2	1.3	1.2	0.2
SD	0.7	1.4	1	1.4	2	1.6	.6
Mdn	0	0	.5	1	0	0	0
Range	0–4	0–8	0–4	0–6	0–8	0–6	0–3
Skew	1.967	1.984	1.686	1.242	1.682	1.687	3.511
Kurtosis	4.975	4.346	3.016	1.362	2.310	3.312	12.798
Inaccurate/incomplete							
M	0.04	0.8	0.2	0.6	1.5	0.8	0.8
SD	.2	1.2	0.6	0.9	1.4	0.9	1.4
Mdn	0	0	0	0	1	1	0
Range	0–1	0–7	0–2	0–4	0–5	0–3	0–7
Skew	4.654	1.914	2.510	1.471	0.713	0.944	2.658
Kurtosis	19.932	4.671	5.324	1.945	0.017	0.129	8.136
Absent							
M	15.1	24.9	18.1	23.1	24.4	27.8	29.8
SD	6.1	6.8	6	6.2	6.5	5.9	4.2
Mdn	14	26	18.5	24	25	29	31
Range	2–31	4–34	4–31	9–32	10–34	17–34	15–34
Skew	0.561	−0.479	−0.231	−0.210	−0.246	−0.710	−1.284
Kurtosis	0.076	−0.681	0.636	−1.026	−1.056	−0.896	1.721
MC attempts							
M	18.6	9	15.8	10.7	9.5	6.2	4.1
SD	6.2	6.7	6	6.2	6.6	5.9	4.1
Mdn	20	8	15.5	10	9	5	3
Range	0–32	0–30	3–30	1–24	0–24	0–17	0–19
Skew	−0.700	0.495	0.277	0.195	0.280	0.710	1.317
Kurtosis	0.309	−0.66	0.609	−1.093	−1.019	−0.896	1.881
MC composite							
M	55	21.1	43.5	26.2	20.4	12.6	7.6
SD	18.4	18	17.7	16.6	15.7	12.2	8.2
Mdn	57	16	41	23.5	16	9	5
Range	7–94	0–86	7–86	2–71	0–59	0–32	0–41
Skew	−0.547	0.857	0.409	0.514	0.548	0.648	1.632
Kurtosis	0.049	0.178	0.590	−0.482	−0.717	−1.245	3.514

Note. PNBI = persons not brain injured; PWA = persons with aphasia; NABW = not aphasic by Western Aphasia Battery; ANO = anomic; CON = conduction; WER = Wernicke's; BRO = Broca's; MC = main concept.

Descriptive Statistics for Each Story

Table A5. *Sandwich.*

Descriptive statistic	PNBI	PWA	NABW	ANO	CON	WER	BRO
Accurate/complete							
M	5.6	1.5	4.5	1.6	1.0	0.3	0.1
SD	1.9	2.1	1.9	2.1	1.8	0.8	0.3
Mdn	6	0	4	1	0	0	0
Range	1–10	0–10	0–8	0–10	0–7	0–3	0–2
Skew	−0.258	1.472	−0.358	1.190	2.423	2.695	4.424
Kurtosis	−0.202	1.431	0.204	1.500	5.691	6.781	20.766
Accurate/incomplete							
M	0.4	1.6	1.0	1.8	1.9	1.7	2.0
SD	0.7	1.4	1.1	1.4	1.3	1.4	1.7
Mdn	0	2	1	2	2	2	2
Range	0–3	0–6	0–4	0–6	0–5	0–4	0–6
Skew	1.528	0.628	1.111	0.714	0.369	−0.144	0.616
Kurtosis	1.805	−0.171	0.544	0.303	−0.693	−1.613	−0.424
Inaccurate/complete							
M	0.1	0.3	0.2	0.5	0.4	0.1	0.1
SD	0.4	0.7	0.5	0.8	1.1	0.3	0.3
Mdn	0	0	0	0	0	0	0
Range	0–3	0–6	0–2	0–3	0–6	0–1	0–2
Skew	3.614	3.523	2.676	1.504	3.533	2.798	4.424
Kurtosis	16.300	17.180	7.053	1.581	13.504	6.509	20.766
Inaccurate/incomplete							
M	0.0	0.5	0.1	0.3	0.9	0.5	0.7
SD	0.1	0.8	0.3	0.7	1.0	0.6	0.9
Mdn	0	0	0	0	1	0	0
Range	0–1	0–4	0–1	0–3	0–3	0–2	0–4
Skew	8.425	1.613	3.373	2.163	0.792	0.703	1.386
Kurtosis	69.944	2.140	10.156	4.616	−0.506	−0.312	1.721
Absent							
M	3.8	5.9	4.2	5.3	5.8	7.3	7.0
SD	1.7	2.1	1.5	2.0	1.7	1.6	1.9
Mdn	4	6	4	6	6	7	7
Range	0–8	0–10	1–8	0–9	3–9	3–10	2–10
Skew	0.170	−0.240	−0.059	−0.517	−0.014	−0.762	−0.298
Kurtosis	−0.625	−0.216	0.913	−0.333	−0.491	1.676	−0.130
MC attempts							
M	6.2	4.0	5.8	4.6	4.2	2.7	2.8
SD	1.7	2.0	1.5	2.0	1.7	1.6	1.8
Mdn	6	4	6	4	4	3	3
Range	2–10	0–10	2–9	0–10	1–7	0–7	0–7
Skew	−0.170	0.157	0.059	0.318	0.014	0.762	0.112
Kurtosis	−0.625	−0.314	0.913	−0.391	−0.491	1.676	−0.549
MC composite							
M	18.0	9.0	16.0	10.9	8.4	5.2	5.0
SD	5.2	5.8	4.7	5.6	4.7	3.9	3.5
Mdn	18	8	16	10	7	6	5
Range	5–30	0–30	4–26	0–30	1–21	0–17	0–13
Skew	−0.201	0.732	−0.115	0.706	0.980	1.359	0.225
Kurtosis	−0.455	0.180	0.814	0.275	1.164	3.630	−0.736

Note. PNBI = persons not brain injured; PWA = persons with aphasia; NABW = not aphasic by Western Aphasia Battery; ANO = anomic; CON = conduction; WER = Wernicke's; BRO = Broca's; MC = main concept.

Median Test Results for Main Concept (MC) Composite and MC Codes by Story and Group

Table B1. Median tests comparing persons not brain injured and aphasia subtypes for MC composite scores.

Test statistic	Variable	Broken Window	Cinderella	Sandwich	Cat Rescue	Refused Umbrella
PWA	(χ^2)	**119.966**	**128.655**	**110.683**	**136.99**	**148.484**
	(ρ)	**.56**	**.58**	**.538**	**.598**	**.623**
NABW	(χ^2)	—	—	—	—	—
	(ρ)					
Anomic	(χ^2)	**52.595**	**66.92**	49.474	**73.393**	**88.041**
	(ρ)	**.477**	**.538**	.463	**.564**	**.617**
Conduction	(χ^2)	59.347	58.486	63.924	56.208	65.563
	(ρ)	.557	.553	.579	.542	.586
Wernicke	(χ^2)	29.266	68.877	45.894	55.738	53.902
	(ρ)	.422	.648	.529	.583	.573
Broca	(χ^2)	97.124	129.274	107.445	125.866	118.995
	(ρ)	**.687**	**.792**	**.722**	**.782**	**.76**

Note. Bold cells reflect comparisons with large effect sizes, and cells with regular font have medium effect sizes. Em dashes indicate comparisons that were not significant or had small effect sizes. PWAs = persons with aphasia; NABW = not aphasic by Western Aphasia Battery.

Table B2. Median tests comparing persons not brain injured and aphasia subtypes for MC attempts.

Test statistic	Variable	Broken Window	Cinderella	Sandwich	Cat Rescue	Refused Umbrella
PWA	(χ^2)	72.164	**98.97**	68.562	83.114	**119.996**
	(ρ)	.434	**.508**	.423	.466	**.56**
NABW	(χ^2)	—	—	—	—	—
	(ρ)					
Anomic	(χ^2)	**31.413**	44.106	26.278	37.155	**84.485**
	(ρ)	**.369**	.454	.337	.401	**.605**
Conduction	(χ^2)	26.708	58.486	63.924	20.953	65.563
	(ρ)	.374	.553	.579	.331	.586
Wernicke	(χ^2)	29.266	39.282	33.552	25.909	42.43
	(ρ)	.422	.454	.419	.397	.471
Broca	(χ^2)	—	32.09	25.821	**81.072**	32.148
	(ρ)		.442	.397	**.627**	.443

Note. Bold cells reflect comparisons with large effect sizes, and cells with regular font have medium effect sizes. Em dashes indicate comparisons that were not significant or had small effect sizes. PWA = persons with aphasia; NABW = not aphasic by Western Aphasia Battery.

Table B3. Median tests comparing persons not brain injured and aphasia subtypes for the accurate/complete code.

Test statistic	Variable	Broken Window	Cinderella	Sandwich	Cat Rescue	Refused Umbrella
PWA	(χ^2)	**164.041**	**180.073**	**176.41**	**183.564**	**186.91**
	(ρ)	**.654**	**.686**	**.679**	**.692**	**.699**
NABW	(χ^2)	—	—	—	—	—
	(ρ)					
Anomic	(χ^2)	82.5	**104.109**	**93.568**	**102.634**	**108.418**
	(ρ)	.598	**.671**	**.636**	**.667**	**.685**
Conduction	(χ^2)	89.222	109.877	105.572	91.768	108.332
	(ρ)	.683	.758	.743	.693	.753
Wernicke	(χ^2)	**71.255**	**104.378**	**73.762**	**73.762**	**76.408**
	(ρ)	**.659**	**.798**	**.671**	**.671**	**.683**
Broca	(χ^2)	**136.493**	**168.372**	**139.318**	**139.318**	**142.215**
	(ρ)	**.814**	**.904**	**.822**	**.822**	**.831**

Note. Bold cells reflect comparisons with large effect sizes, and cells with regular font have medium effect sizes. Em dashes indicate comparisons that were not significant or had small effect sizes. PWA = persons with aphasia; NABW = not aphasic by Western Aphasia Battery.

Median Test Results for Main Concept (MC) Composite and MC Codes by Story and Group

Table B4. Median tests comparing persons not brain injured and aphasia subtypes for the accurate/incomplete code.

Test statistic	Variable	Broken Window	Cinderella	Sandwich	Cat Rescue	Refused Umbrella
PWA	(χ^2)	**126.468**	**96.226**	69.462	87.075	46.526
	(ρ)	**.575**	**.501**	.426	.477	.349
NABW	(χ^2)	18.619	35.967	—	—	—
	(ρ)	.33	.459			
Anomic	(χ^2)	—	**82.901**	53.974	**61.946**	—
	(ρ)		**.599**	.483	**.518**	
Conduction	(χ^2)	**111.291**	**56.126**	45.129	**78.208**	**55.637**
	(ρ)	**.763**	**.542**	.486	**.64**	**.54**
Wernicke	(χ^2)	**48.902**	—	32.193	**21.314**	16.618
	(ρ)	**.546**		.443	**.361**	.318
Broca	(χ^2)	**115.52**	28.835	50.428	**58.786**	**51.703**
	(ρ)	**.749**	.374	.495	**.534**	**.501**

Note. Bold cells reflect comparisons with large effect sizes, and cells with regular font have medium effect sizes. Em dashes indicate comparisons that were not significant or had small effect sizes. PWA = persons with aphasia; NABW = not aphasic by Western Aphasia Battery.

Table B5. Median tests comparing persons not brain injured and aphasia subtypes for the inaccurate/complete code.

Test statistic	Variable	Broken Window	Cinderella	Sandwich	Cat Rescue	Refused Umbrella
PWA	(χ^2)	40.238	—	—	49.168	—
	(ρ)	.324			.358	
NABW	(χ^2)	—	—	—	26.510	—
	(ρ)				.394	
Anomic	(χ^2)	35.014	—	—	48.701	31.245
	(ρ)	.389	—	—	.459	.368
Conduction	(χ^2)	**58.343**	—	—	**76.24**	38.862
	(ρ)	**.553**			**.632**	.451
Wernicke	(χ^2)	28.455	—	—	17.974	26.822
	(ρ)	.417			.331	.404
Broca	(χ^2)	—	—	—	—	—
	(ρ)					

Note. Bold cells reflect comparisons with large effect sizes, and cells with regular font have medium effect sizes. Em dashes indicate comparisons that were not significant or had small effect sizes. PWA = persons with aphasia; NABW = not aphasic by Western Aphasia Battery.

Table B6. Median tests comparing persons not brain injured and aphasia subtypes for the inaccurate/incomplete code.

Test statistic	Variable	Broken Window	Cinderella	Sandwich	Cat Rescue	Refused Umbrella
PWA	(χ^2)	—	69.973	62.5	—	—
	(ρ)		.427	.404		
NABW	(χ^2)	—	—	—	—	—
	(ρ)					
Anomic	(χ^2)	13.972	49.425	33.963	—	—
	(ρ)	.246	.463	.383		
Conduction	(χ^2)	—	**89.455**	84.876	**50.06**	26.32
	(ρ)		**.684**	.667	**.512**	.371
Wernicke	(χ^2)	39.358	**44.87**	56.779	23.566	23.321
	(ρ)	.49	**.523**	.588	.379	.377
Broca	(χ^2)	46.884	45.597	**68.405**	18.389	22.371
	(ρ)	.477	.47	**.576**	.379	.33

Note. Bold cells reflect comparisons with large effect sizes, and cells with regular font have medium effect sizes. Em dashes indicate comparisons that were not significant or had small effect sizes. PWA = persons with aphasia; NABW = not aphasic by Western Aphasia Battery.

Median Test Results for Main Concept (MC) Composite and MC Codes by Story and Group

Table B7. Median tests comparing persons not brain injured and aphasia subtypes for the absent code.

Test statistic	Variable	*Broken Window*	*Cinderella*	*Sandwich*	*Cat Rescue*	*Refused Umbrella*
PWA	(χ^2)	47.563	**103.691**	59.687	93.733	**107.987**
	(ρ)	.352	**.52**	.395	.495	**.531**
NABW	(χ^2)	—	—	—	—	—
	(ρ)					
Anomic	(χ^2)	25.854	52.806	24.253	45.47	**69.354**
	(ρ)	.335	.478	.324	.444	**.548**
Conduction	(χ^2)	—	**49.339**	22.717	32.182	**58.26**
	(ρ)		**.508**	.345	.41	
Wernicke	(χ^2)	26.822	40.747	**45.894**	**52.693**	59.534
	(ρ)	.404	.498	**.529**	**.567**	.603
Broca	(χ^2)	**64.522**	117.977	76.608	117.567	144.913
	(ρ)	**.56**	.757	.61	.755	.839

Note. Bold cells reflect comparisons with large effect sizes, and cells with regular font have medium effect sizes. Em dashes indicate comparisons that were not significant or had small effect sizes. PWA = persons with aphasia; NABW = not aphasic by Western Aphasia Battery.

Main Concept Lists for the 5 Discourse Tasks

Broken Window
1. The boy was outside.
2. The boy was playing soccer.
3. The ball breaks the window.
4. The man is sitting.
5. The man was startled.
6. The ball broke a lamp.
7. The man picked up the ball.
8. The man looked out of the window.

Refused Umbrella
1. It is going to rain.
2. You need to take the umbrella.
3. The boy (*does something to refuse*) the umbrella.
4. The boy walks to school.
5. It is raining.
6. The boy gets soaking wet.
7. The boy runs back.
8. The mother is (*negative emotional state*).
9. The boy gets the umbrella.
10. The boy goes back to school.

Cat Rescue

1. The little girl was riding her bicycle.
2. The cat was in the tree.
3. The dog was barking.
4. The man climbed up the tree.
5. The man tries to rescue the cat.
6. The ladder fell down.
7. The father is stuck in the tree.
8. Someone called the fire department.
9. The fire department comes with a ladder.
10. The fire department rescues them.

Cinderella

1. Dad remarried a woman.
2. Cinderella lives with stepmothers/stepsisters.
3. The stepmother/stepsisters were mean to Cinderella.
4. Cinderella was a servant.
5. Cinderella has to do the housework.
6. The prince needs to get married.
7. There is going to be a ball.
8. The got an invitation.
9. They are excited.
10. Cinderella cannot go.
11. The stepsisters tore Cinderella's dress.
12. Stepmother/stepsisters went.
13. Cinderella was upset.
14. A fairy godmother appeared.
15. The fairy godmother makes (*item[s]*) turn into (*item[s]*).
16. The fairy godmother makes Cinderella into a beautiful princess.
17. Cinderella went to the ball.
18. She had to be home by midnight.
19. The prince and Cinderella danced.
20. The prince falls in love with Cinderella.
21. It is midnight.
22. She ran down the stairs.
23. She lost one of her glass slippers.
24. The prince finds Cinderella's slipper.
25. Everything turns back to its original form.
26. She returned home.
27. The prince searched for Cinderella.
28. The prince comes to Cinderella's house.

29.   The stepsisters try on the glass slipper.

30.   The slipper didn't fit the stepsisters.

31.   He put the slipper on.

32.   The slipper fits.

33.   Cinderella and the prince are married.

34.   Cinderella and the prince lived happily ever after.

Sandwich

1.   Get the bread out.

2.   Get two slices of bread.

3.   Get the peanut butter.

4.   Get the jelly.

5.   Get a knife.

6.   Put the bread on a plate.

7.   Put peanut butter on the bread.

8.   Put jelly on the bread.

9.   Put the two pieces together.

10.   Cut the sandwich in pieces.

The main concepts for the *Broken Window, Cinderella,* and *Sandwich* tasks are reprinted with permission of the publisher, Taylor & Francis Ltd. (http://www.tandfonline.com), from the following: Richardson, J. D., & Dalton, S. G. (2015). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology*, 1–29. https://doi.org/10.1080/02687038.2015.1057891