AJSLP

Research Article

The Relationship Between Confrontation Naming and Story Gist Production in Aphasia

Jessica D. Richardson,[a] Sarah Grace Hudspeth Dalton,[a] Davida Fromm,[b] Margaret Forbes,[b] Audrey Holland,[c] and Brian MacWhinney[b]

Purpose: The purpose of this study was to examine the relationship between picture naming performance and the ability to communicate the gist, or essential elements, of a story. We also sought to determine if this relationship varied according to Western Aphasia Battery–Revised (WAB-R; Kertesz, 2007) aphasia subtype.

Method: Demographic information, test scores, and transcripts of 258 individuals with aphasia completing 3 narrative tasks were retrieved from the AphasiaBank database. Narratives were subjected to a main concept analysis to determine gist production. A correlation analysis was used to investigate the relationship between naming scores and main concept production for the whole group of persons with aphasia and for WAB-R subtypes separately.

Results: We found strong correlations between naming test scores and narrative gist production for the large sample of persons with aphasia. However, the strength of the correlations varied by WAB-R subtype.

Conclusions: Picture naming may accurately predict gist production for individuals with Broca's and Wernicke's aphasia, but not for other WAB-R subtypes. Given the current reprioritization of outcome measurement, picture naming may not be an appropriate surrogate measure for functional communication for all persons with aphasia.

Supplemental Materials: https://doi.org/10.23641/asha.5851848

Anomia, the inability to retrieve object and action names, is the primary diagnostic feature of all aphasia subtypes (Benson, 1988; Goodglass & Geschwind, 1976; Goodglass & Wingfield, 1997). Picture naming assessment can reveal differences in language abilities between persons with aphasia (PWAs) and neurotypical controls as well as between different aphasia subtypes (Bastiaanse & Jonkers, 1998; Kohn & Goodglass, 1985). Such assessments are easy to administer, well validated, and reliable across users and settings. These qualities, in addition to the common practice of using impairment-based outcome measures to determine response to impairment-based intervention, make standardized picture naming assessments one of the most widely used tools for PWAs in clinical and research settings (Brady, Kelly, Godwin, & Enderby, 2012; Grinnon et al., 2012; Simmons-Mackie, Threats, & Kagan, 2005; Wallace, Worrall, Rose, & Le Dorze, 2016b). Treatment studies frequently take advantage of the ease and reliability of naming administration to create study-specific naming assessments to document improvement on treated and untreated items.

The usefulness of picture naming assessment in the evolving landscape of outcome reprioritization is in question. In response to emphasis placed on activity and participation by the World Health Organization (2007) *International Classification of Functioning, Disability, and Health* framework and the complementary Life Participation Approach to Aphasia (Chapey et al., 2001; Elman, 2016; Simmons-Mackie & Kagan, 2007), developers and providers of aphasia treatment are encouraged to adopt a life and social participation approach when planning treatment and measuring outcomes. This is further fueled by a recent focus on stakeholder input into outcomes, in which PWAs, caregivers, and providers consistently identify activity and participation goals (e.g., participation in life roles, participation in leisure, improved psychosocial well-being, increased autonomy) in addition to language and/or communication goals as priorities (Brown, Worrall, Davidson, & Howe, 2012; Wallace et al., 2016b; Worrall et al., 2011).

A welcome change to the recent Cochrane Library systematic review (Brady, Kelly, Godwin, Enderby, & Campbell, 2016) was the redefining of primary outcomes: "The primary

[a]University of New Mexico, Albuquerque
[b]Carnegie Mellon University, Pittsburgh, PA
[c]University of Arizona, Tucson

Correspondence to Jessica D. Richardson: jdrichardson@unm.edu

Editor: Margaret Blake
Associate Editor: Gerasimos Fergadiotis

Received November 1, 2016
Revision received April 18, 2017
Accepted October 6, 2017
https://doi.org/10.1044/2017_AJSLP-16-0211

Publisher Note: This article is part of the Special Issue: Select Papers From the 46th Clinical Aphasiology Conference.

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

outcome chosen to indicate the effectiveness of an intervention that aims to improve communicative ability must reflect communication activity in real world settings, that is, functional communication" (p. 6). Since Martha Taylor Sarno first introduced this concept, researchers and clinicians have struggled to define and quantify functional communication (Sarno, 1969; Taylor, 1965). In the review, Brady and colleagues (2016) broadly defined functional communication as the ability to transmit messages through any combination of communication modalities and identified measurement approaches that included discourse analysis (DA), Communication Activities of Daily Living (Holland, 1980; Holland, Frattali, & Fromm, 1998), Communicative Effectiveness Index (Lomas et al., 1989), Functional Communication Profile (Sarno, 1969), therapy outcome measures (Enderby, 2006), spontaneous speech subtests of the Western Aphasia Battery–Revised (WAB-R; Kertesz, 2007) and the Aachen Aphasia Test (Huber, Poeck, & Wilmes, 1983), and others. Impairment-based measures were relegated to secondary outcomes that could serve as a surrogate or delegate for functional communication in the absence of psychometrically sound and widely adopted measures of functional communication.

As research into development and refinement of functional communication measures progresses, it continues to be useful to explore the relationships between the reprioritized primary and secondary outcome measures (Brady et al., 2016). The relationship between picture naming assessments and DA may be useful for estimating the potential impact of the aphasic language deficit in connected speech. Research to date has focused on describing relationships between lexical retrieval during picture naming and lexical retrieval during narrative and conversational tasks; so far, the nature of this relationship is unclear. Differences in participant inclusion criteria and assessment measures likely account for much of the variability in the findings. For example, investigations often limited the types of aphasia to one (Kambanaros, 2010; Pashek & Tompkins, 2002) or two (Bastiaanse & Jonkers, 1998; Herbert, Hickin, Howard, Osborne, & Best, 2008; Mayer & Murray, 2003) subtypes. Of the research that examined two subtypes, only one reported correlations separately for the groups (agrammatic and anomic; Bastiaanse & Jonkers, 1998); other studies combined different types and severities of aphasia when examining relationships between naming ability and lexical retrieval in discourse (Fergadiotis & Wright, 2016; Herbert et al., 2008; Mayer & Murray, 2003). Additional contributors to the unclear results are the wide range of discourse measures that have been used and whether the study examined noun naming or verb naming ability. What follows is a brief summary of findings to date, organized by similarities in population studied when possible (see also Table 1).

In persons with anomic and agrammatic/Broca's aphasia, a significant correlation between picture naming (objects) and percentage of nouns and of content units per turn has been observed (Herbert et al., 2008). In a similar population, no significant relationships between picture naming (actions) and verb retrieval or diversity during connected speech were present (Bastiaanse & Jonkers, 1998). In persons with mild–moderate fluent aphasia, significant correlations between object naming and number of nouns produced in spontaneous speech were observed (Kambanaros, 2010; Pashek & Tompkins, 2002), but none were observed between object naming and type-token ratio or between action naming and number or diversity of verbs (Kambanaros, 2010). In a combined group with mild–moderate aphasia (subtypes not specified), significant correlations between object naming and percentage of words retrieved (in picture description and conversation) and percentage of substantive verbs (in picture description) were present (Mayer & Murray, 2003). There was also a significant relationship between action naming and both connected speech tasks for percentage of words retrieved and percentage of substantive verbs (Mayer & Murray, 2003). In a large group of all subtypes except transcortical sensory aphasia, an object and action naming composite score was strongly related to informativeness (correct information units [CIUs]) but not to proportion of paraphasias in connected speech (Fergadiotis & Wright, 2016).

Because of the variability in study populations and measures, a clear picture of the relationship between picture naming and discourse variables is not available, but it does seem that relationships between picture naming and some discourse variables are present for some types and severities of aphasia. A repeated and strong finding is a relationship between confrontation naming of nouns and retrieval of nouns in picture description and conversation, where noun naming explains between 41% and 71% of the variance in noun retrieval during discourse in individuals across a wide range of aphasia subtypes and severities (Herbert et al., 2008; Mayer & Murray, 2003; see Table 1). Noun and/or verb naming may also correspond to informativeness during connected speech tasks, as measured by content words or CIUs in a mixed sample of PWAs (Fergadiotis & Wright, 2016; Herbert et al., 2008). Finally, although not a repeated finding as yet, the lack of relationship between naming and production of paraphasias during discourse in a large and varied sample of PWAs is remarkable (Fergadiotis & Wright, 2016). Replication of these studies as well as shared discourse outcomes across future studies would help to facilitate the identification of relationships between naming and word retrieval during discourse.

Although having simple tools to predict lexical retrieval during connected speech would be valuable, this would not necessarily predict a PWA's ability to convey the gist of a story or procedure, a discourse variable more closely related to the idea of functional communication and the ability to communicate a desired message (Brady et al., 2016). To illustrate, a person with Broca's aphasia may have a perfect correlation between picture naming scores and the number of nouns or content units produced during a discourse task; however, this person's discourse sample may be mostly an inventory of items in the picture, with the gist of the story untold. Conversely, a person with anomic aphasia may also have a perfect correlation—high score on naming tests and high number of content units produced in connected speech—and may be more likely to express the

Table 1. Previous research examining confrontation naming and discourse performance.

Study	Sample	Tasks	Measures used	Correlation outcomes
Bastiaanse & Jonkers, 1998	16 PWAs -Eight agrammatic -Eight anomic	-Object naming -Action naming -Stroke story	-# correct objects -# correct actions -# spontaneous verbs -TTR of verbs -Inflection index -# of copulas and modals -# of arguments produced per verb	# correct actions named in relationship with: -# spontaneous verbs produced: -Agrammatic: $\rho = .41$, *ns* -Anomic: $\rho = -.14$, *ns* -TTR of verbs: -Agrammatic: $\rho = .64$, *ns* -Anomic: $\rho = -.14$, *ns*
Fergadiotis et al., 2016	98 PWAs -33 anomic -22 Broca's -22 conduction -14 Wernicke's -Four global -Three TCM	-WAB-R naming subtest -BNT short form -VNT -Free speech -Picture description -Story retell	Confrontation naming factor -Proportion of correct items named on WAB-R -Proportion of correct items named on BNT -Proportion of correct items named on VNT Discourse factor -Proportion of paraphasias during free speech -Proportion of paraphasias during picture description -Proportion of paraphasias during story retell CIU factor -Proportion of CIUs per number of words	Confrontation naming factor in relationship with: -Discourse factor: paraphasias, $r^2 = .27$ -CIU factor: $r^2 = .46$ Proportion of correct WAB in relationship with: -Paraphasias in free speech: $r^2 = .14$, us -Paraphasias in picture description: $r^2 = .18$, us -Paraphasias during story retell: $r^2 = .08$, us -Proportion of CIUs: $r^2 = .31$, us Proportion of correct BNT in relationship with: -Paraphasias in free speech: $r^2 = .15$, us -Paraphasias in picture description: $r^2 = .2$, us -Paraphasias during story retell: $r^2 = .07$, us -Proportion of CIUs: $r^2 = .36$, us Proportion of correct VNT in relationship with: -Paraphasias in free speech: $r^2 = .21$, us -Paraphasias in picture description: $r^2 = .2$, us -Paraphasias during story retell: $r^2 = .1$, us -Proportion of CIUs: $r^2 = .38$, us
Herbert et al., 2008	10 PWAs -Four anomic -Five Broca's -One unclassified	-Picture (noun) naming test -Conversation with familiar partner (tasks completed at two time points)	-% correctly named items -% nouns/content words per speech unit -% nouns/content words per turn taken -% nouns/content words per substantive turn -% substantive turns per turn	% Correctly named items in relationship with: -% nouns per turn taken -Time 1: $r^2 = .45$; Time 2: $r^2 = .45$ -% nouns per substantive turn -Time 1: $r^2 = .62$; Time 2: $r^2 = .41$ -% content words per speech unit -Time 1: $r^2 = .36$ -% content words per turn taken -Time 1: $r^2 = .36$; Time 2: $r^2 = .48$ -% content words per substantive turn -Time 1: $r^2 = .45$; Time 2: $r^2 = .64$

(*table continues*)

Table 1. *(Continued).*

Study	Sample	Tasks	Measures used	Correlation outcomes
Kambanaros, 2010	12 PWAs with anomic aphasia, Greek–English bilingual	-GOAT (administered in Greek and English) -Conversation in Greek and English	-# correct objects/actions named -TTR of nouns, TTR of verbs -# types of nouns -# tokens of nouns -# types of verbs -# tokens of verbs	# actions named in relationship with: -# verb tokens -Greek: ρ = .45, *ns*; English: ρ = .28, *ns* -Verb TTR -Greek: ρ = .65; English: ρ = .31, *ns* # objects named in relationship with: -# noun tokens -Greek: ρ = .82; English: ρ = .51, *ns* -Noun TTR -Greek: ρ = .54, *ns*; English: ρ = .62, *ns*
Mayer et al., 2003	14 PWAs: -Five anomic -Four Broca's -Two conduction -One TCM -Two Wernicke's	-TAWF (nouns and verbs) -Description of three-picture sequences -Conversation with the first author	-% correct production of TAWF nouns -% correct production of TAWF verbs -%WR-D -%WR-C -%SV-D -%SV-C	% TAWF nouns named in relationship with: -%WR-D: nouns, r^2 = .61; verbs, r^2 = .64 -%WR-C: nouns, r^2 = .71; verbs, r^2 = .64 -%SV: discourse, r^2 = .71; conversation, r^2 = .46, *ns* % TAWF verbs named in relationship with: -%WR-D: nouns, r^2 = .76; verbs, r^2 = .69 -%WR-C: nouns, r^2 = .79; verbs, r^2 = .77 -%SV: discourse, r^2 = .71; conversation, r^2 = .48
Pashek et al., 2002	20 PWAs with anomic aphasia	-BNT -Narration of short video clips	-BNT standard score -Proportion of word-finding difficulty during narration	BNT standard score in relationship with proportion of word-finding difficulty during narration: r^2 = .58

Note. r^2 is the Pearson product–moment correlation coefficient squared or the coefficient of determination. It may be interpreted as the amount of variance between the two variables explained; ρ is the Spearman correlation coefficient based on ranked data and should be interpreted as an effect size informing on the strength of the relationship between the ranks of the two variables. We did not choose to square ρ, but ρ^2 can be interpreted as the amount of variance in the ranks of the two variables that is explained, but not the variance between the two variables themselves. *ns* = not significant; us = unknown *p* value, unable to determine significance; PWA = person with aphasia; TCM = transcortical motor aphasia; BNT = Boston Naming Test; CIU = correct information unit; GOAT = Greek Object and Action Test; %WR-D = percentage of word retrieval in discourse; %WR-C = percentage of word retrieval in conversation; %SV-D = proportion of substantive vs. light verbs in discourse; %SV-C = proportion of substantive vs. light verbs in conversation; TAWF = Test of Adolescent and Adult Word Finding; TTR = type/token ratio; VNT = Verb Naming Test; WAB-R = Western Aphasia Battery–Revised.

story gist. To add to this complexity, research demonstrates that even those with mild aphasia, who often have the highest naming scores, produce discourse samples that are characterized by reduced coherence, complexity, content, and lexical diversity (e.g., Andreetta, Cantagallo, & Marini, 2012; Capilouto, Wright, & Wagovich, 2006; Fergadiotis & Wright, 2011; Nicholas & Brookshire, 1995; Ulatowska, North, & Macaluso-Haynes, 1981). This could translate to a person with mild anomic aphasia or perhaps a person performing at test ceiling and therefore not classifiable, who may perform well on naming assessment but cannot express gist.

Investigations of the relationship between naming and production of main concepts (MCs) would inform whether naming ability is an adequate surrogate for the vital ability to convey gist in connected speech. Importantly, word-level informativeness measures such as CIUs (Nicholas & Brookshire, 1993) and content units (Yorkston & Beukelman, 1980) do not make a distinction between information that is correct versus information that is essential. MC analysis (MCA) is a proposition level measure of the informativeness of an individual's discourse that differs from CIUs and content units because it assesses whether speakers are able to communicate concepts considered to be essential for expressing the overall gist of a picture or story about which those involved in the communication exchange share knowledge (Nicholas & Brookshire, 1993, 1995). MCA requires the existence of a closed checklist of MCs, defined as an utterance containing a subject, a main verb, and, as appropriate, an object and subordinate clauses (Nicholas & Brookshire, 1995). Discourse samples are examined for the presence or absence of each MC—if it is present, the concept's accuracy and completeness are coded. MCA is accepted as a measure of communication adequacy that correlates well with listener perceptions of improved functional communication and conversational abilities (Doyle, Goda, & Spencer, 1995; Kong, 2009; Nicholas & Brookshire, 1995; Ross & Wertz, 1999). MCA is also reliable and sensitive to increments or decrements in information content (Hopper, Holland, & Rewega, 2002; Nicholas & Brookshire, 1993, 1995). Developers Nicholas and Brookshire (1995) reported point-to-point interrater reliability greater than 80% for all except one participant (for whom reliability was 74%) and intrarater reliability greater than 80% for all participants. Test–retest stability across three sessions has demonstrated high correlations (between .7 and .96; Boyle, 2014; Nicholas & Brookshire, 1995). Kong and colleagues (Kong, 2009; Kong, Whiteside, & Bargmann, 2016) similarly report high interrater, intrarater, and test–retest reliabilities ($r > .90$) for most MC codes and for overall MC scores but lower reliability (between .49 and .52) for concepts receiving inaccurate or incomplete scores. Using enhanced checklists, Richardson and Dalton (2016) reported point-to-point intrarater reliability from 91% to 100% and interrater reliability from 90% to 100% across three discourse (Broken Window, Cinderella, and How to Make a Sandwich) tasks in healthy controls.

There is little information about the relationship between lexical retrieval (during confrontation naming or connected speech) and the ability to produce the MCs necessary to tell the gist of a story or procedure. Lexical retrieval during discourse (as measured by percentage of lexical information units and incomplete sentences) has been correlated with narrative coherence in individuals with anomic aphasia (Marini, Andreetta, del Tin, & Carlomagno, 2011). In addition, core lexicon production of task-specific content and function words during narrative correlated with MC production in controls and PWAs (Dalton & Richardson, 2015). However, revisiting data for African American persons with moderate anomic, Broca's, or conduction aphasia, we found no correlation between object naming (on the WAB naming subtest) and coherence, reference, or emplotment (defined as the ability to provide information about an event in a narrative structure while including all elements; Ulatowska et al., 2003). Whether information about lexical retrieval (in picture naming or connected speech) can provide predictive information about one's ability to convey gist in narrative is not yet well studied.

The relationship between performance on noun and verb naming assessments and discourse performance has not been systematically examined in a large and diverse cohort of PWAs. With the exception of Fergadiotis and Wright (2016), all of the studies reviewed included relatively small sample sizes (≤ 16 PWAs) and highly variable and restricted subtypes for comparison. Furthermore, most of the discourse measures investigated to date are microlinguistic in nature. Because discourse involves a set of highly complex speech acts, microlinguistic measures may not be sufficient as a functional communication outcome when investigating the relationship between confrontation naming and discourse performance. With recent advances in DA, it is possible to investigate the relationship between naming performance and gist in a large sample of PWAs. The first advancement relates to the development of AphasiaBank (http://talkbank.org/AphasiaBank), an online database of standardized testing results and transcriptions (with accompanying audio and/or video files) of PWAs and controls performing various standardized discourse tasks (Forbes, Fromm, & MacWhinney, 2012; MacWhinney, Fromm, Forbes, & Holland, 2011). As contributors to AphasiaBank continue to grow in number and as it continues to be used to generate new knowledge, AphasiaBank is experiencing success as a source for Big Data (MacWhinney & Fromm, 2016), and we used it in this manner for this investigation. The second advancement relates to the development of MC checklists for select AphasiaBank discourse tasks, specifically a picture sequence narrative (Broken Window), a storytelling narrative (Cinderella), and a procedural description (How to Make a Peanut Butter and Jelly Sandwich; see Richardson & Dalton, 2016, for lists, available via AphasiaBank).

The first aim of this investigation was to determine the relationship between (a) performance on the Boston Naming Test–Short Form (BNT-15; Kaplan, Goodglass, & Weintraub, 2001) and discourse performance as measured by MCA (Nicholas & Brookshire, 1995; Richardson & Dalton, 2016) and (b) performance on the Verb Naming Test (VNT; Cho-Reyes & Thompson, 2012) and MCA in a large sample of PWAs. The second aim of this investigation was to determine if these relationships vary by WAB-R (Kertesz, 2007) aphasia subtype.

Method

Participants

Demographic information, test scores, and transcripts from 258 PWAs (148 men, 110 women) were retrieved from the AphasiaBank database for use in this study. Participants primarily reported identifying as White (221), with 23 identifying as African American, eight identifying as Hispanic/Latino, three identifying as Asian, and one identifying as American Indian. One person reported being of mixed races, and race/ethnicity was unknown for one other person. Using the group classification criteria from the WAB-R Aphasia Quotient (WAB-R-AQ; Kertesz, 2007) for diagnosis, the sample included 86 individuals with anomic aphasia (M = 84.77, SD = 6.86, range = 63.4–93.4), 66 with Broca's aphasia (M = 51.5, SD = 14.95, range = 10.8–77.6), 47 with conduction aphasia (M = 70, SD = 9.35, range = 49–90), two with global aphasia (AQ = 20.3 and 20.5), nine with transcortical motor aphasia (M = 70.22, SD = 5.91, range = 59.8–79.3), two with transcortical sensory aphasia (AQ = 54.1 and 79.3), 21 with Wernicke's aphasia (M = 53.68, SD = 13.69, range = 28.2–74.4), and 25 who were not aphasic by WAB-R-AQ score (not aphasic by WAB-R [NABW]; M = 96.33, SD = 1.67, range = 93.8–99.6). The average WAB-R-AQ score for the entire sample was 70.9 (SD = 19.33, range = 10.8–99.6), spanning the entire range of severity, from very severe to within normal limits (see Table 2 and Supplemental Material S1). Individuals with global aphasia, transcortical sensory aphasia, transcortical motor aphasia, and unknown aphasia were excluded from the analyses for the second aim as there were insufficient numbers in each group for accurate comparison. In total, 244 participants were included in the analysis for the second aim.

Naming Assessment

Scores for the BNT-15 and VNT for each participant were retrieved from the AphasiaBank database. The BNT-15 assesses noun confrontation naming with 15 pictures, selected empirically from the original 60-item BNT by Mack, Freed, Williams, and Henderson (1992) with a sample of individuals with Alzheimer's disease and older healthy controls. The BNT-15 (Kaplan et al., 2001; Mack et al., 1992) is supported by a recent study that sought to create a BNT short form specifically for PWAs (del Toro et al., 2011). Although the two short versions in question share only three original BNT items (comb, bench, and stethoscope), they both demonstrated acceptable psychometrics (e.g., unidimensionality, person/group separation), matched the item difficulty distribution of the original 60-item BNT, and were identified as appropriate to administer to PWAs. The BNT-15 (Kaplan et al., 2001; Mack et al., 1992) correlated well with longer-form performance in PWAs (r = .99; del Toro et al., 2011).

The VNT, a subtest of the Northwestern Assessment of Verbs and Sentences (Cho-Reyes & Thompson, 2012), contains 22 black and white drawings of actions. The Northwestern Assessment of Verbs and Sentences was developed with a sample of individuals with anomic aphasia and Broca's aphasia with agrammatism and age-matched healthy controls. The VNT is sensitive to differences between

Table 2. Demographic information and test results for all participants and by aphasia subtype.

Descriptive statistics		Broca's (n = 66)	Wernicke's (n = 21)	Conduction (n = 47)	Anomic (n = 86)	NABW (n = 25)	All (N = 258)
Age	Mean (*SD*)	57.9 (13.1)	66.9 (10.9)	64.5 (12.5)	61.6 (12.3)	61.1 (14.0)	61.8 (12.7)
	Median	56.8	69.1	64.5	62.9	59.6	73.2
	Range	25.6–85.4	42.6–81.3	30.9–90.7	32.7–85.7	26.0–80.7	25.6–90.7
WAB-AQ	Mean (*SD*)	51.5 (14.9)	53.6 (13.7)	70.0 (9.3)	84.8 (6.9)	96.3 (1.7)	70.9 (19.3)
	Median	54.7	53.0	70.9	86.5	96.0	73.2
	Range	10.8–77.6	28.2–74.4	49.0–90.0	63.4–93.4	93.8–99.6	10.8–99.6
	Severity	6 Very severe	—	—	—	N/A	8 Very severe
		16 Severe	9 Severe	2 Severe	—		27 Severe
		43 Moderate	12 Moderate	29 Moderate	10 Moderate		104 Moderate
		1 Mild	—	16 Mild	75 Mild		94 Mild
		—	—	—	1 Unknown		—
BNT	Mean (*SD*)	4.0 (3.1)	4.1 (4.4)	6.0 (3.6)	9.8 (3.7)	13.2 (2.2)	7.3 (4.6)
	Median	4	2	6	11	14	7
	Range	0–12	0–14	0–13	0–15	7–15	0–15
VNT	Mean (*SD*)	9.1 (6.0)	9.4 (5.9)	14.1 (4.9)	18.2 (4.0)	21.4 (0.8)	14.5 (6.6)
	Median	8.5	9	15	20	22	16
	Range	0–22	0–20	5–21	4–22	20–22	0–22
MC composite	Mean (*SD*)	16.5 (12.5)	25.5 (17.5)	35.7 (20.2)	47.3 (22.7)	69.2 (21.4)	36.3 (25.1)
	Median	14.5	19	29	48	71	29
	Range	0–62	6–60	10–100	10–113	13–121	0–121

Note. Age is reported in years. MC = main concept; VNT = Verb Naming Test; BNT = Boston Naming Test; WAB-AQ = Western Aphasia Battery–Aphasia Quotient; NABW = not aphasic by Western Aphasia Battery–Revised.

groups, and authors demonstrated good interrater reliability (99.8%; Cho-Reyes & Thompson, 2012). The VNT shows good external validity with high correlations with the WAB-AQ and WAB naming subtest, Boston Diagnostic Aphasia Examination action naming (Goodglass & Kaplan, 1983), and Verb and Sentence Test naming (Bastiaanse, Edwards, & Rispens, 2002).

Transcripts

All narratives were elicited using the standardized discourse protocol developed by AphasiaBank. Narratives are transcribed using the Codes for the Human Analysis of Transcripts format (MacWhinney, 2000). We limited our analyses to three narrative tasks (which are referred to as “gems”): (a) a picture sequence narrative, Broken Window, during which speakers look at a four-panel picture sequence while telling a story with a beginning, middle, and end; (b) a storytelling, Cinderella, during which participants tell the story relying on any details they can remember about the story and/or those details prompted by the brief pre-elicitation review of a wordless picture book; and (c) a procedural narrative, How to Make a Peanut Butter and Jelly Sandwich. These narrative tasks are the AphasiaBank database tasks for which standardized and norm-referenced MC lists, developed based on a large sample of healthy control speakers, are currently available (Richardson & Dalton, 2016). Therefore, the study included three different semi-spontaneous tasks that, when combined, might better predict spontaneous speech while still relying on relatively shared knowledge across many speakers, thus maintaining assessment validity and reliability (Boyle, 2015; Brookshire & Nicholas, 1994; Richardson & Dalton, 2016).

Using the Computerized Language Analysis tool, the story gems of interest for this project were isolated from the rest of the transcript using the command: +g +s(story name) +d1 +fs(story name) +t*PAR *.cha. This command output a file with the specified gem for each participant into a specified directory. The file included the participant’s utterances and header information with age, gender, and other pertinent demographic information. Using this procedure, three transcript files (one for each of the above-named tasks) were created for each participant.

MCA

Each story transcript for each participant was scored for MC production using guidelines for determining accuracy and completeness (Nicholas & Brookshire, 1995) and MC checklists with an accompanying scoring system developed from a large database of neurologically healthy control speakers (Richardson & Dalton, 2016). Each MC consists of several essential elements that correspond to the subject, verb, and any objects or subordinate clauses (as appropriate). The checklists developed by Richardson and Dalton (2016) include a canonical version of the MC representing the most common form and vocabulary used to express the concept as well as examples of alternate words commonly used by healthy controls for that concept. The checklists also contain information about which MCs were produced in forms that differed substantially from their canonical form, pairs of MCs that were sometimes produced in a single statement, and examples of related statements that should not be applied to a given concept. With this system, MCs receive a numeric score based on the criteria of accuracy (essential elements are correct) and completeness (essential elements are present).

A score of 0 was assigned for MCs that were missing (absent) or not attempted in the participant’s transcript. A score of 3 was assigned when the concept produced contained all of the essential elements and all elements were accurate (accurate and complete [AC]). The following codes were arranged between those two extreme codes: a score of 1 assigned to concepts that had one or more essential elements produced inaccurately and one or more essential elements missing (inaccurate/incomplete [II]) and a score of 2 assigned to concepts that either had an inaccurate essential element with all other elements produced accurately (inaccurate/complete [IC]) or had a missing essential element with all other essential elements produced accurately (accurate/incomplete [AI]). Table 3 contains the definitions for each MC code, the corresponding numeric score, and examples of statements that would receive a given code.

Both AI and IC concepts receive a score of 2 because they contain an error in only a single dimension. This scoring differs from others (e.g., Kong, 2011; Kong et al., 2016) in an attempt to ensure that equivalent errors are not weighted differently in the scoring. The formula used to yield an overall MC score for each story was $(3 \times AC) + (2 \times AI) + (2 \times IC) + (1 \times II)$. The overall story scores for the three stories were then summed to yield an MC composite score for each PWA. There are eight MCs for the Broken Window gem (for a maximum story score of 8×3 points if all concepts are accurate and complete = 24), 34 MCs for the Cinderella gem (for a story maximum story score of 102), and 10 Sandwich MCs (for a maximum story score of 30). The maximum possible MC composite score is therefore 156, which is achieved if an individual produced all MCs accurately and completely. Lower scores would indicate that fewer MCs were produced and/or that the MCs produced were inaccurate and/or incomplete. It is not possible to determine from the MC composite score how many MCs are absent compared with inaccurate and/or incomplete. Scoring reliability was examined in this sample. Point-to-point interrater reliability was 97.5%, and intrarater reliability was 90%.

Data Analysis

For Aim 1, data were screened to identify outliers and ensure that the assumptions of the correlation analysis were met. The variables were monotonic, and the MC composite score is not a true continuous variable, so Spearman’s rho was used. For Aim 2, only WAB-R subtypes with samples greater than 10 (i.e., Broca’s, Wernicke’s, conduction, anomic, NABW) were included for visualization and

Table 3. Main concept from Richardson and Dalton (2016) and scoring procedures for the main concept analysis with examples of statements that would receive each code.

Example of a main concept and associated information from the Richardson and Dalton (2016) lists	[1]A/The boy [2]was playing [3]soccer. 1. "He" since referent is unambiguous; some give the boy a name 2. Played, is kicking, kicks, is practicing, etc. 3. With the soccer ball, with the ball, with the football* (*if not from U.S.) Note: "He has a ball" or "He has a soccer ball" does not count toward this concept because it does not imply any kind of action with the soccer ball, and boy-action-ball was the concept that met criterion.	
Main concept code	**Definition**	**Score**
Accurate/complete	All elements of the main concept are produced, and all elements are accurate. • "The boy was playing outside with a ball." • "The boy kicked the soccer ball." • "Johnny was practicing playing soccer."	3
Accurate/incomplete	One or more of the elements of the main concept are missing, but the elements that are produced are accurate. • "Boy ball." • "He's kicking around."	2
Inaccurate/complete	All elements of the main concept are produced, but one or more elements are inaccurate. • "She's playing soccer." • "He's kicking the football." • "He's throwing the ball."	2
Inaccurate/incomplete	One or more of the elements of the main concept are missing, and one or more elements that are produced are inaccurate. • "And baseball or something." • "Throws the ball." • "She's football."	1
Absent	No attempt was made to produce any elements of the main concept.	0

Note. For the full list of main concepts, please see "Main concepts for three different discourse tasks in a large non-clinical sample," by J. D. Richardson and S. G. Dalton, 2016, *Aphasiology, 30*(1), 45–73. Copyright 2016 Taylor & Francis. Adapted with permission.

description. Spearman's rho was again used for all subtypes except for NABW, which did not meet the assumption of monotonicity. Interpretation of the strength of the correlation analyses is based on criteria suggested by Hinkle, Wiersma, and Jurs (2003).

Results

The average BNT score was 7.3 (SD = 4.6, range = 0–15), with WAB-R aphasia subtypes (with $N > 10$) ordered from the lowest score to the highest score as follows: Broca's and Wernicke's, conduction, anomic, and then NABW. The average VNT score was 14.5 (SD = 6.6, range = 0–22), with an identical order of subtypes as BNT scores. The average MC composite was 36.3 (SD = 25.1, range = 0–121), and subtypes were ordered from the lowest score to the highest score as follows: Broca's, Wernicke's, conduction, anomic, and then NABW (see Table 2).

A moderate–high correlation between performance on the BNT and MC composite scores ($\rho = .676$, $p < .001$) and a high correlation between VNT and MC composite scores ($\rho = .704$, $p < .001$) were present in the full sample of PWAs (Figures 1–2). Correlations between BNT and MC composite score by WAB-R aphasia subtype ranged from low to high. The correlation between BNT and MC composite scores was low for individuals with anomic ($\rho = .367$, $p < .001$) and conduction ($\rho = .351$, $p = .011$) aphasia and high for individuals with Broca's ($\rho = .705$, $p < .001$) and Wernicke's ($\rho = .834$, $p < .001$) aphasia. Similarly, results for the correlation between VNT and MC composite score ranged from low to high. Low correlations were observed for individuals with anomic ($\rho = .390$, $p < .001$) and conduction ($\rho = .442$, $p = .002$) aphasia. High correlations were observed for individuals with Broca's ($\rho = .738$, $p < .001$) and Wernicke's ($\rho = .835$, $p < .001$) aphasia (Figures 3, 4, 5, 6, 7, 8, 9, and 10). Scatter plots for individuals NABW show apparent ceiling effects for VNT and a less predictable relationship between BNT and MC composite scores (Figures 11–12).

Discussion

To our knowledge, this study included the largest and most diverse sample of PWAs to examine the relationship between naming and the ability to convey a narrative gist. Object and action naming abilities were significantly, positively correlated with gist production, although the significance and strength of the correlations varied by WAB-R aphasia subtype—correlations were highest for individuals with Broca's and Wernicke's aphasia and lower for those with conduction and anomic aphasia. This pattern of high-to-low correlations could also be influenced by severity, as the order of subtypes is also generally the order of average aphasia severity scores (from most to least severe) in our sample (Table 2) as well as in WAB-R standardization samples (Kertesz, 2007). A monotonic relationship between naming and discourse variables was not observed for

Figure 1. Correlation between performance on the Boston Naming Test (BNT) and main concept (MC) composite scores for all participants. NABW = not aphasic by Western Aphasia Battery.

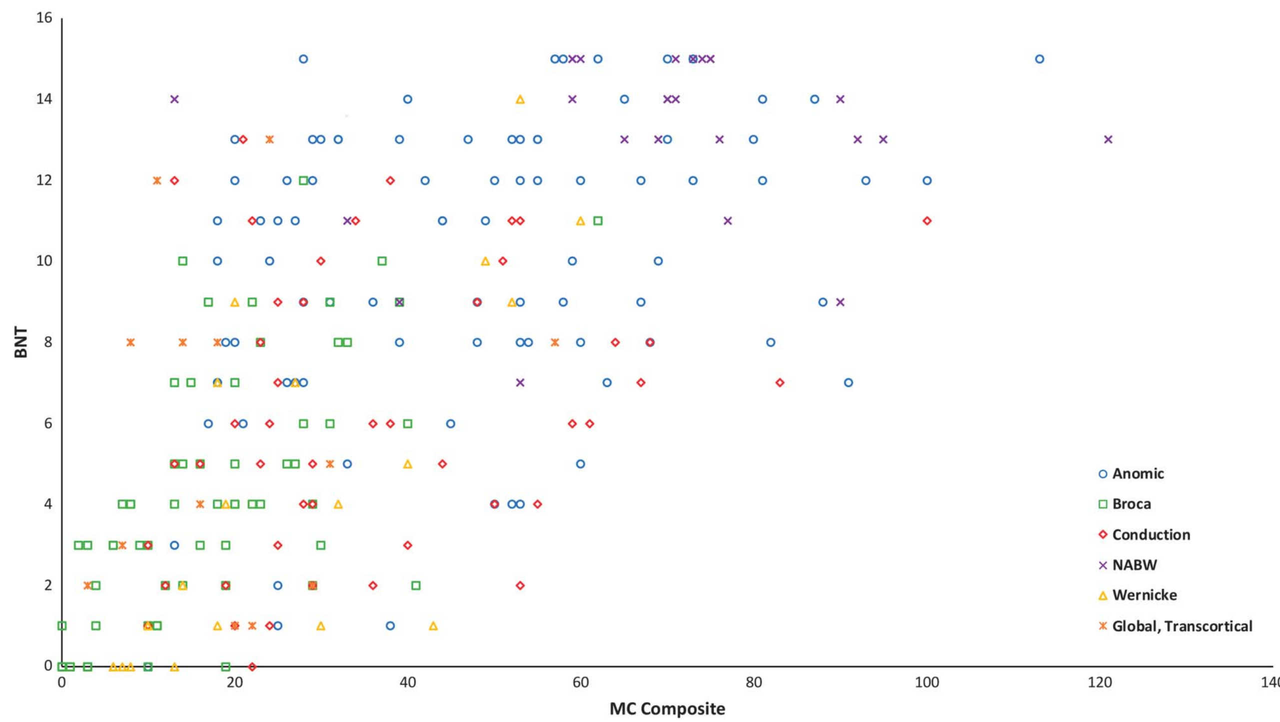


Figure 2. Correlation between performance on the Verb Naming Test (VNT) and main concept (MC) composite scores for all participants. NABW = not aphasic by Western Aphasia Battery.

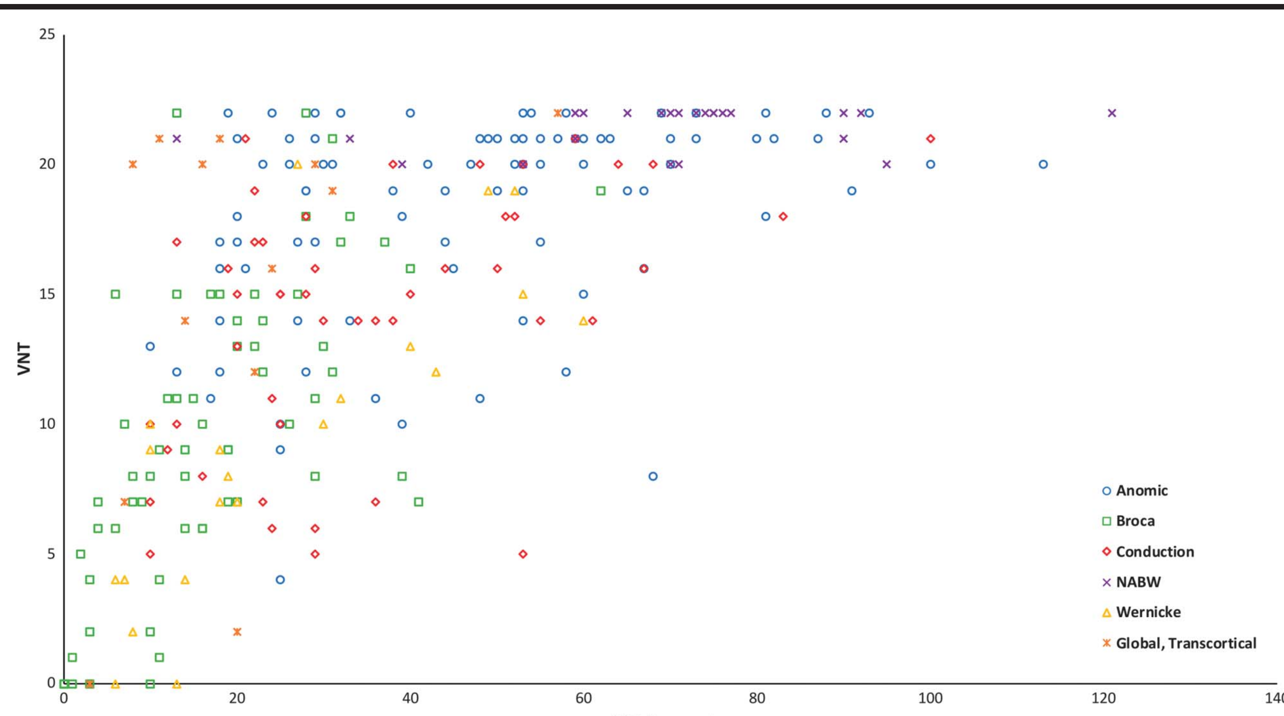


Figure 3. Correlation between performance on the Boston Naming Test (BNT) and main concept (MC) composite scores for participants with Broca's aphasia.

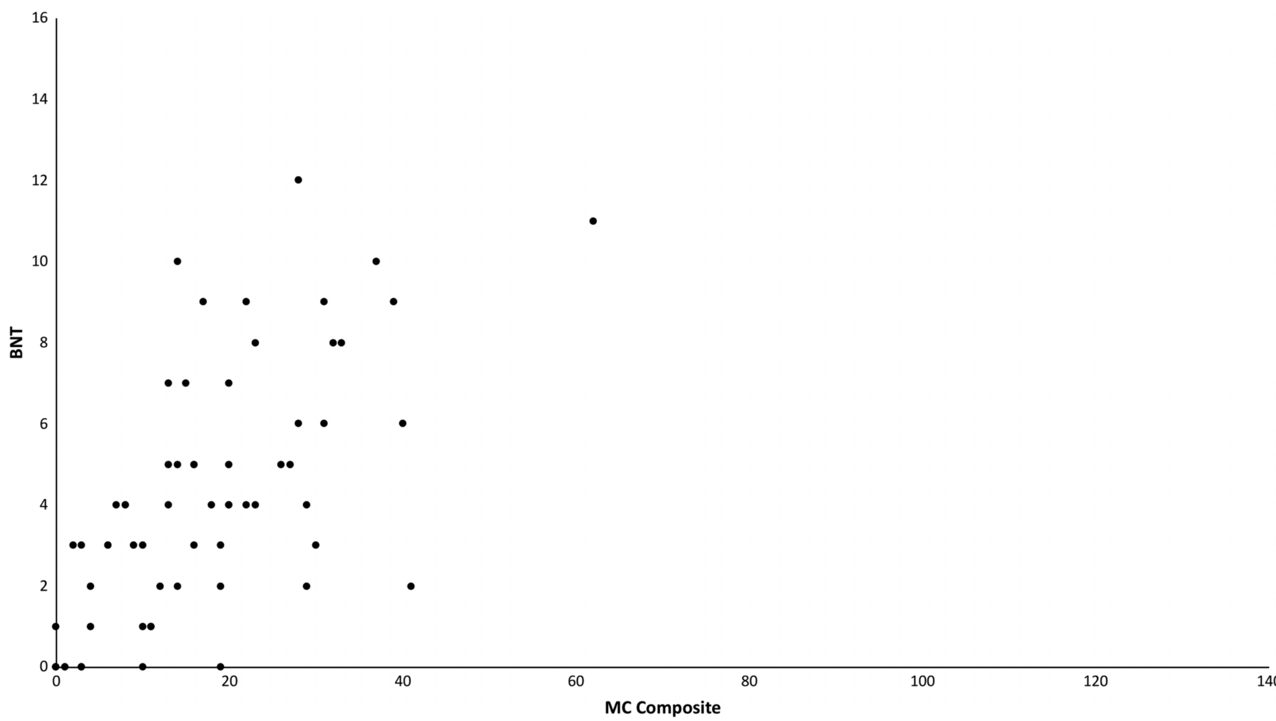


Figure 4. Correlation between performance on the Verb Naming Test (VNT) and main concept (MC) composite scores for participants with Broca's aphasia.

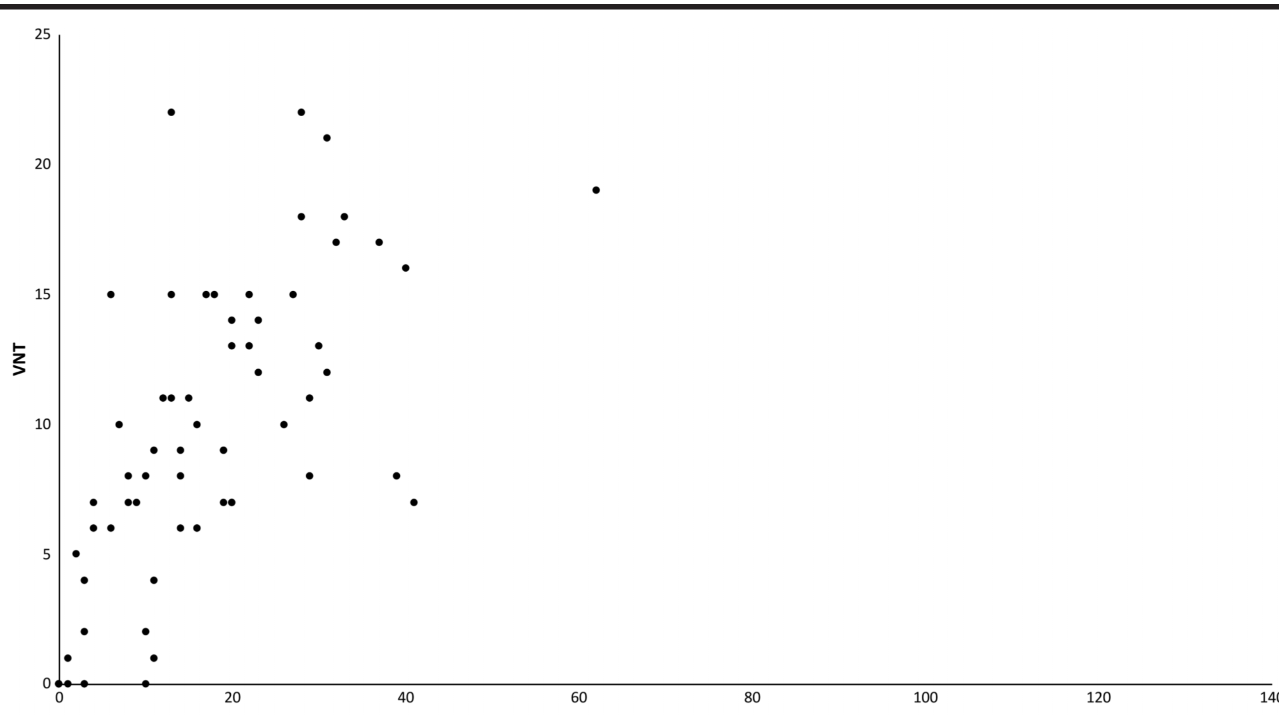


Figure 5. Correlation between performance on the Boston Naming Test (BNT) and main concept (MC) composite scores for participants with Wernicke's aphasia.

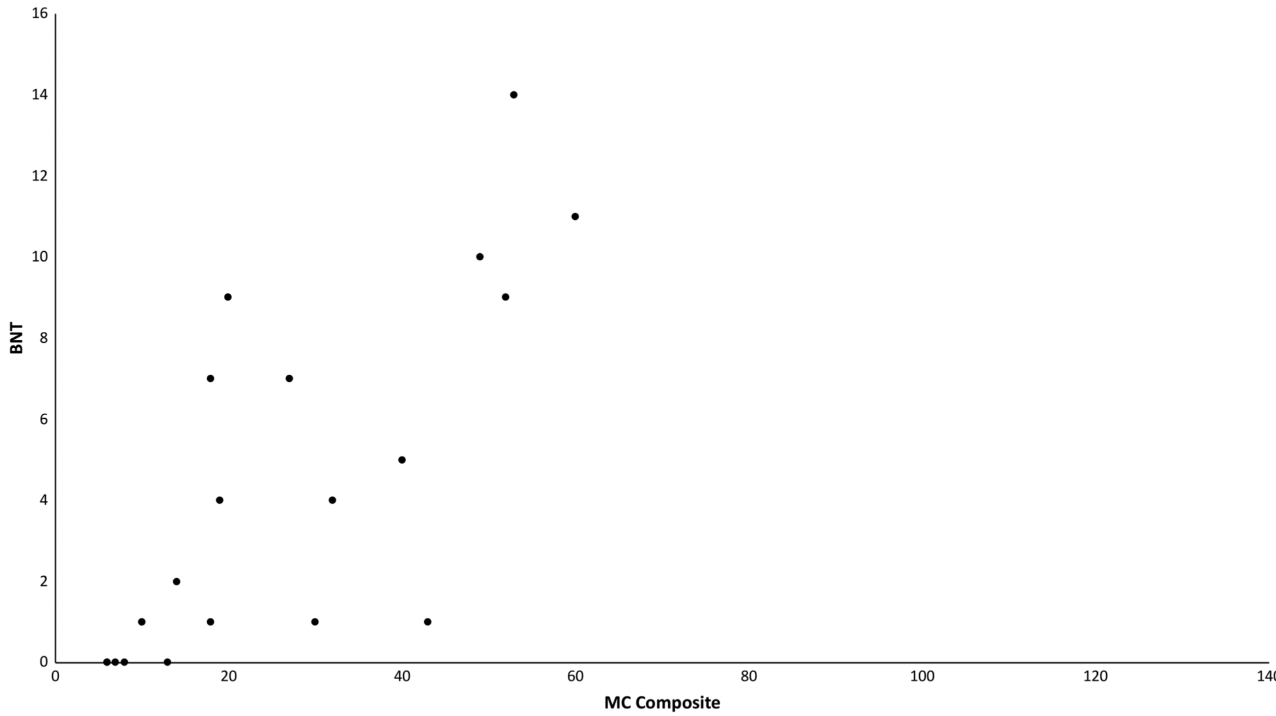


Figure 6. Correlation between performance on the Verb Naming Test (VNT) and main concept (MC) composite scores for participants with Wernicke's aphasia.

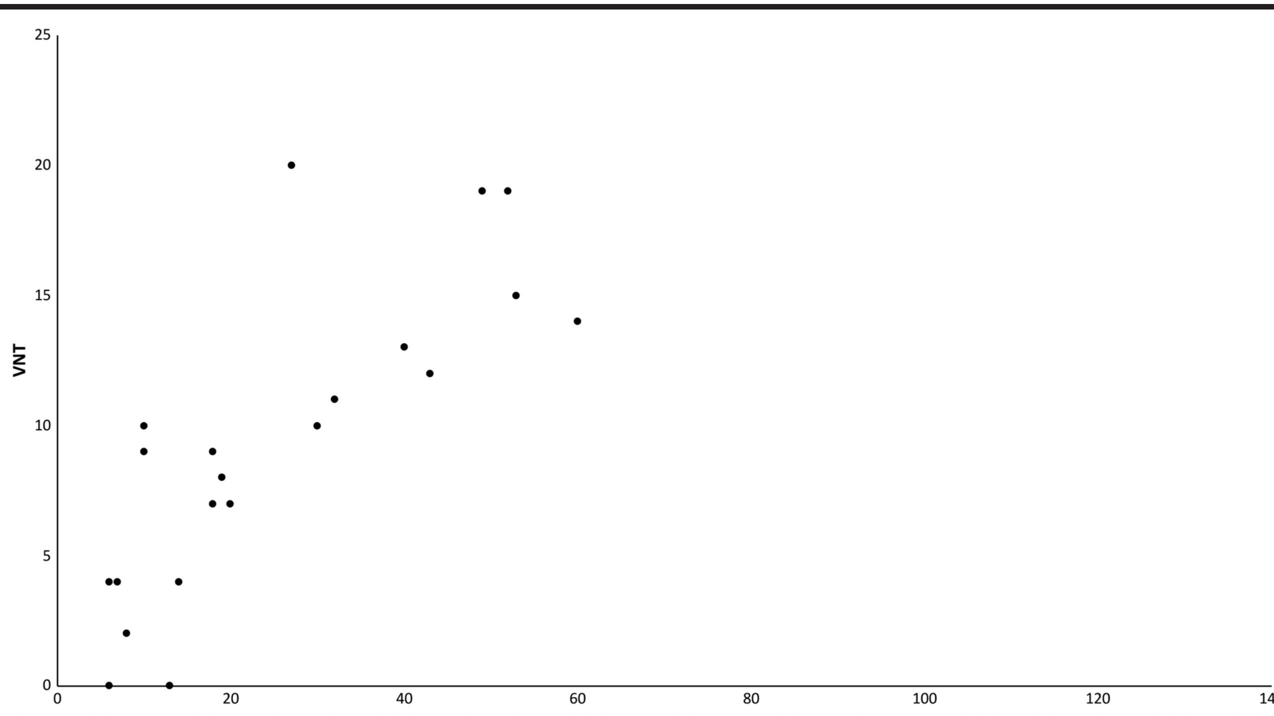


Figure 7. Correlation between performance on the Boston Naming Test (BNT) and main concept (MC) composite scores for participants with conduction aphasia.

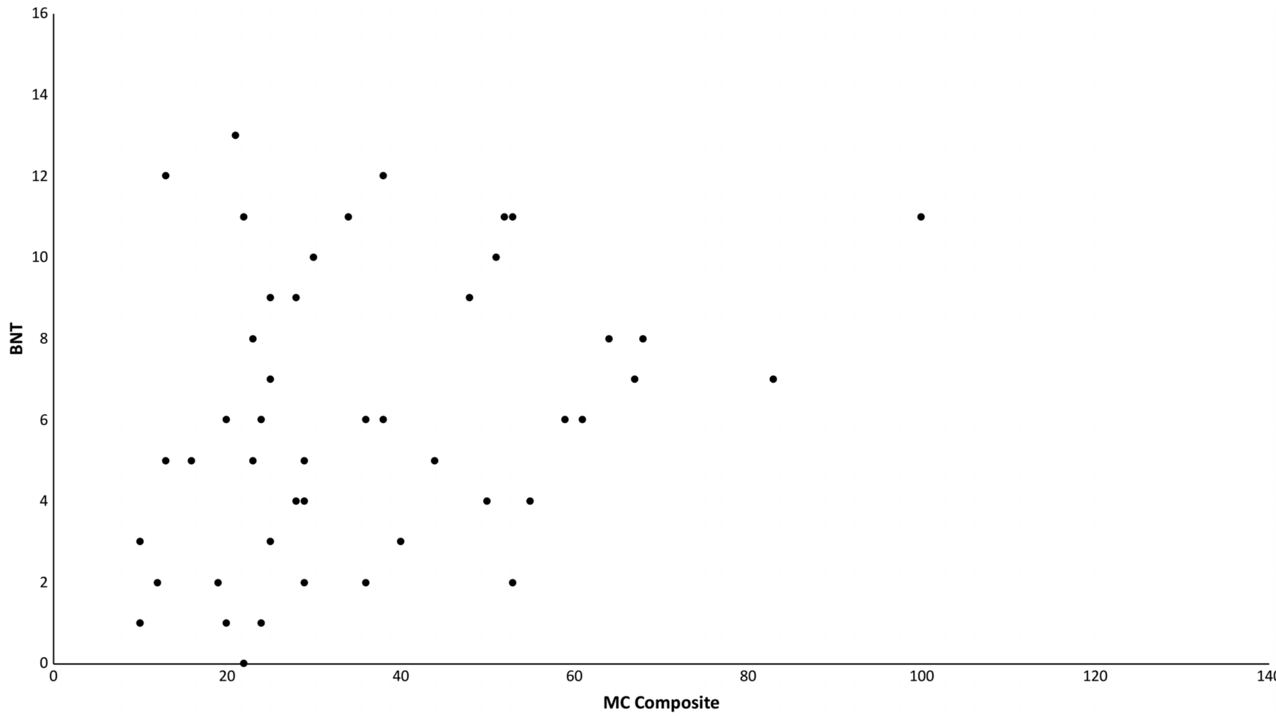


Figure 8. Correlation between performance on the Verb Naming Test (VNT) and main concept (MC) composite scores for participants with conduction aphasia.

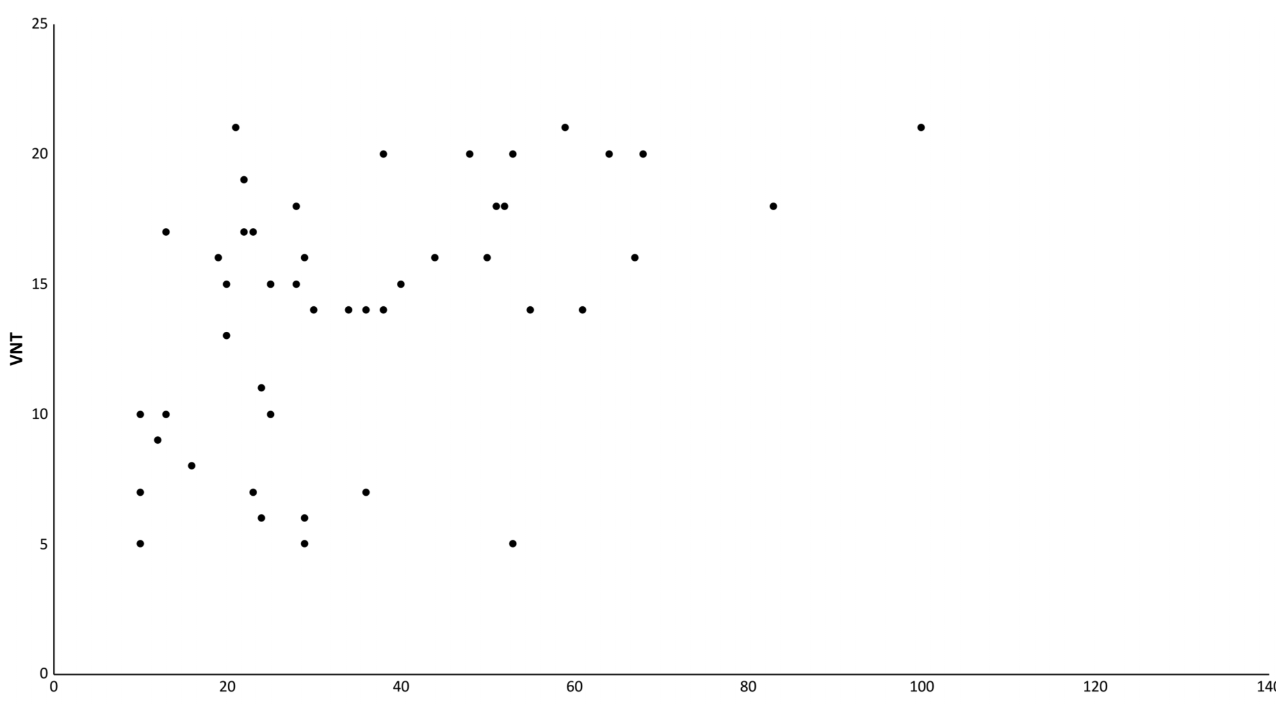


Figure 9. Correlation between performance on the Boston Naming Test (BNT) and main concept (MC) composite scores for participants with anomic aphasia.

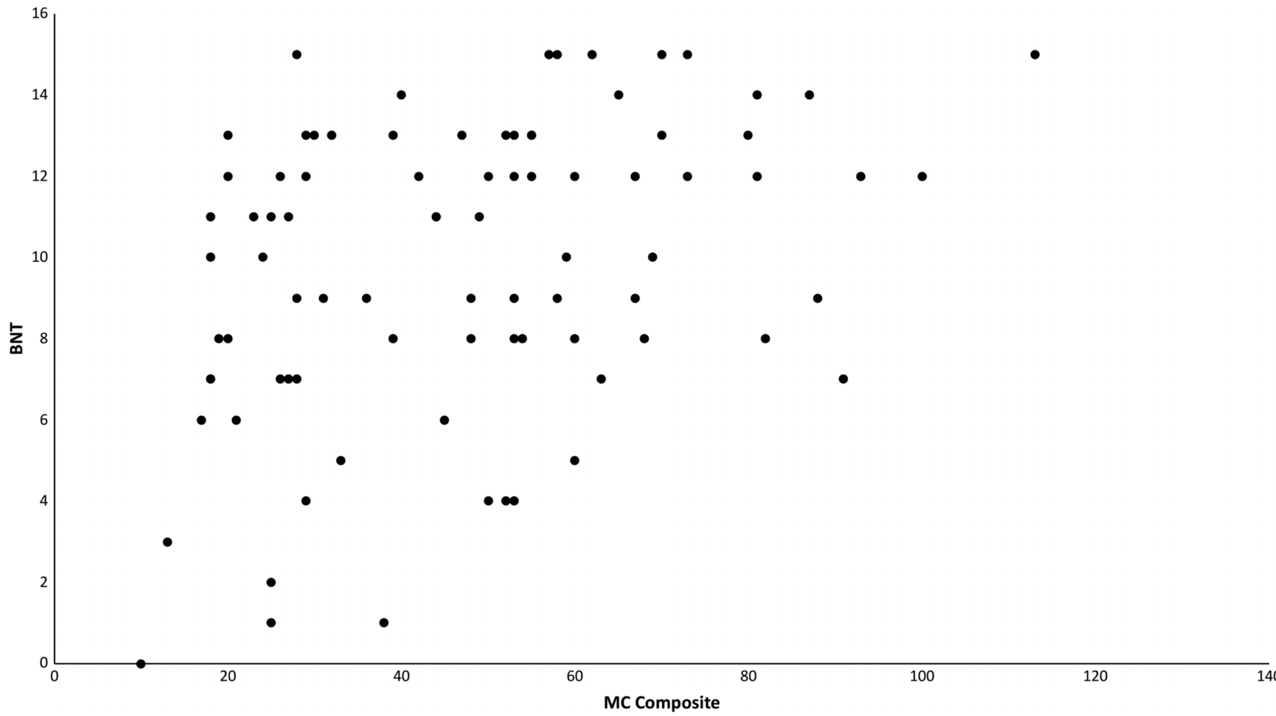


Figure 10. Correlation between performance on the Verb Naming Test (VNT) and main concept (MC) composite scores for participants with anomic aphasia.

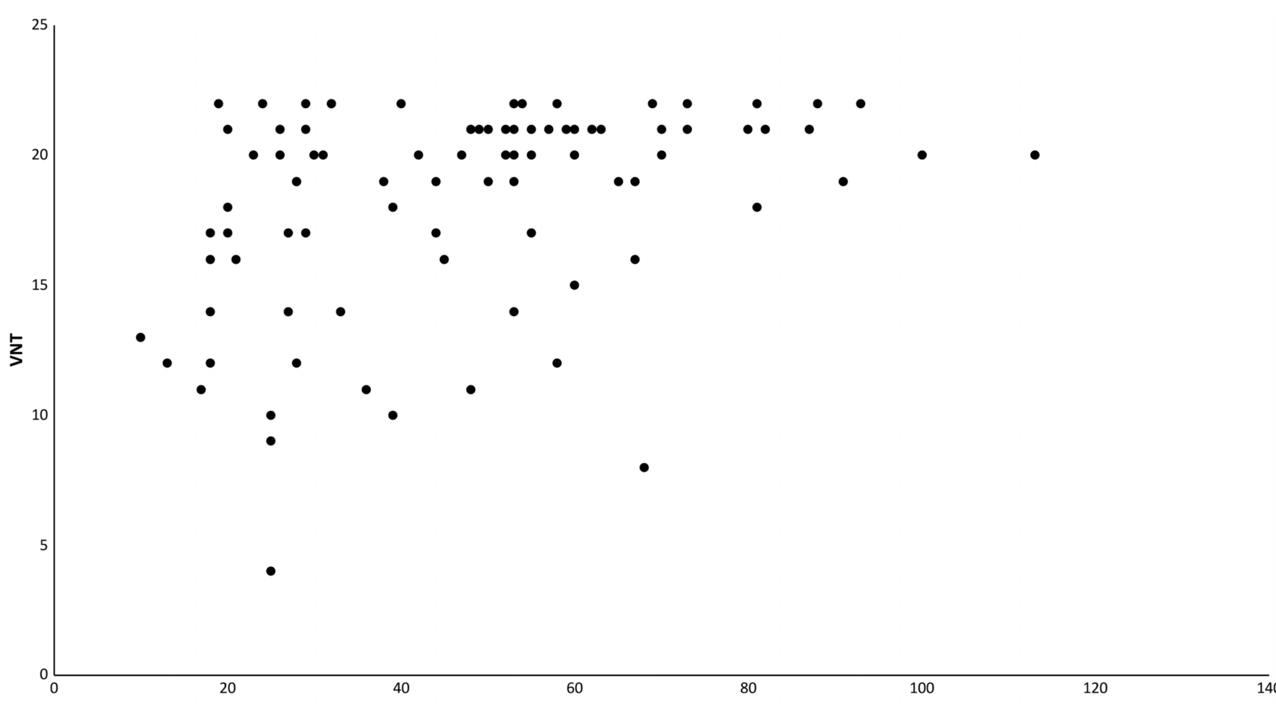


Figure 11. Scatter plot of the Boston Naming Test (BNT) and main concept (MC) composite scores for participants not aphasic by Western Aphasia Battery–Revised.

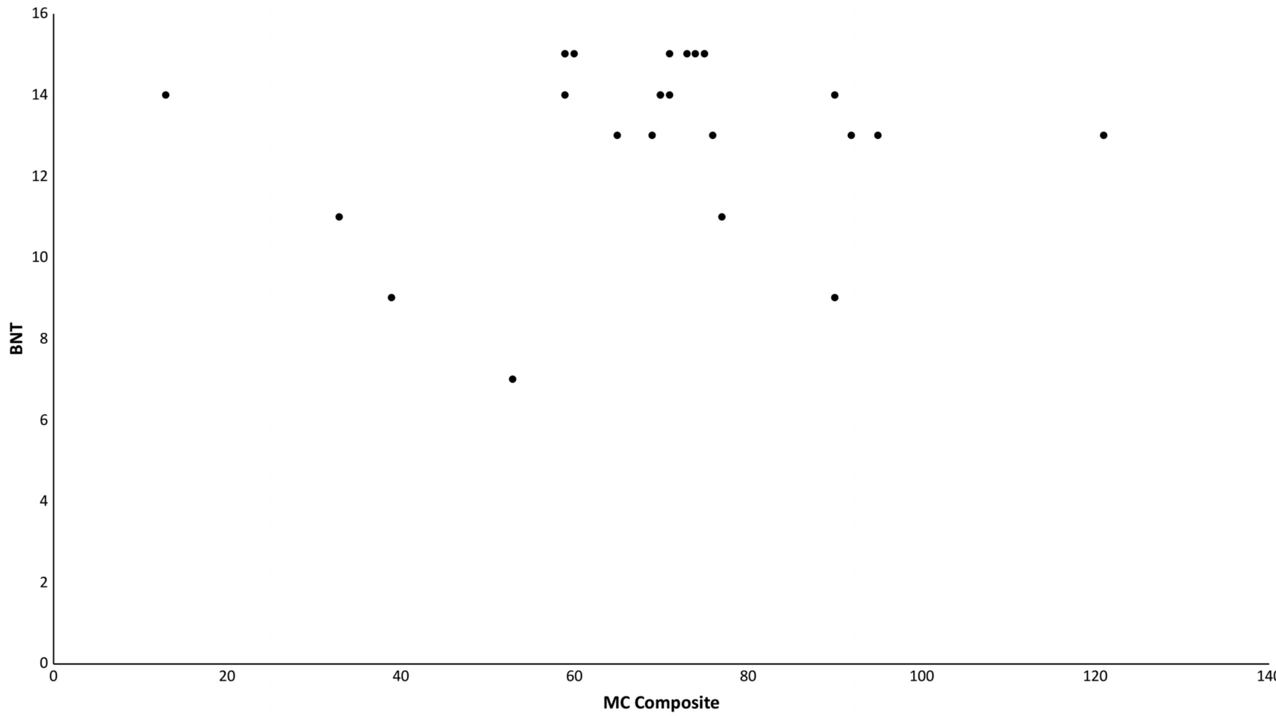
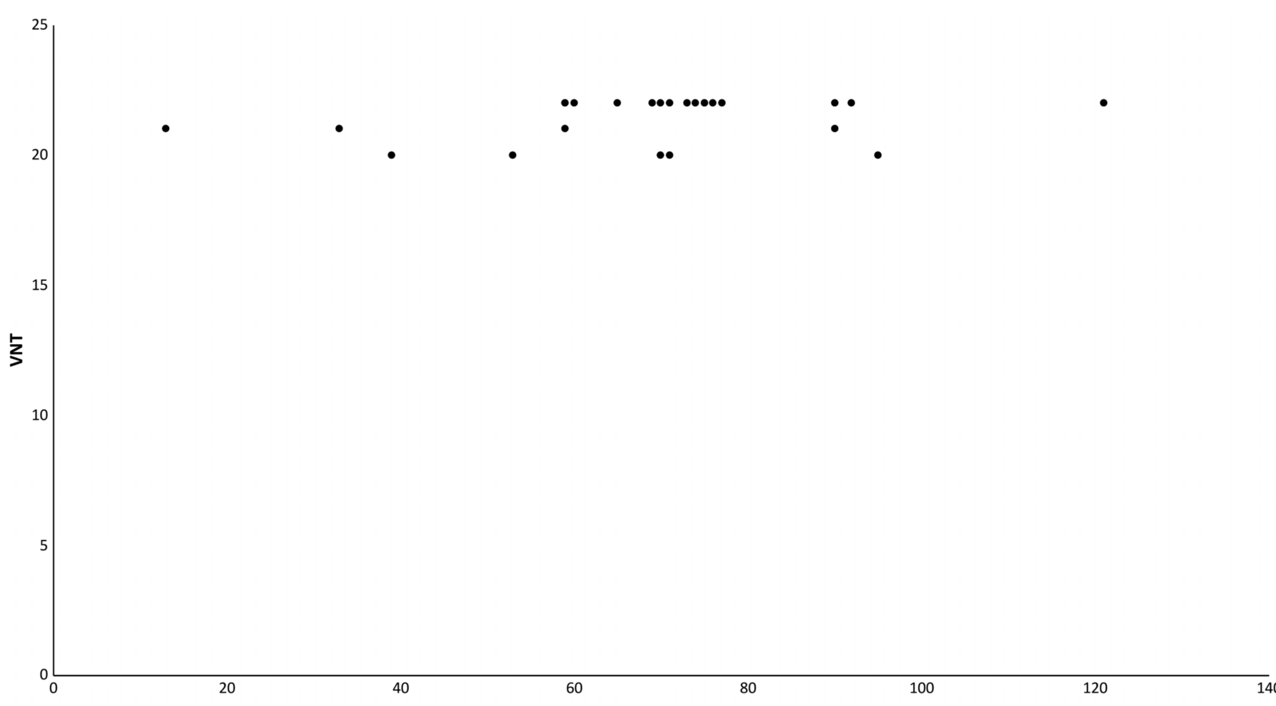


Figure 12. Scatter plot of the Verb Naming Test (VNT) and main concept (MC) composite scores for participants not aphasic by Western Aphasia Battery–Revised.



individuals NABW, because in part of ceiling effects on the naming assessments included in this investigation.

An examination of scatter plots for the WAB-R aphasia subtypes that are generally milder in severity (anomic, NABW) provides some insight into this finding. It is evident that these groups performed nearer to ceiling and in restricted ranges for naming performance, but performance was highly variable and in a broader range for MCs, often overlapping liberally with the MC scores of other, usually more severe, aphasia subtypes. It is possible that utilization of longer naming assessments might capture more nuanced and variable naming performance and thus change the significance and/or strength of these correlations; however, leveraging the AphasiaBank database both allows us to query a large and varied data set but also prescribes the data available (i.e., BNT-15 and VNT). Given the psychometric support for BNT-15 (e.g., predictive validity), it is likely that similar correlations would be observed. Still, a conservative conclusion is that lexical retrieval, as measured by picture naming using the brief naming assessments in this study, may be a relatively accurate predictor of gist for those with WAB-R-defined Broca's and Wernicke's aphasia, but not the other WAB-R subtypes examined here. The results reported here thus support both claims in the literature—first, that picture naming of objects and actions is strongly predictive of discourse performance, at least for two WAB-R aphasia subtypes, and, second, that picture naming of objects and actions is not strongly predictive of discourse performance for several other WAB-R aphasia subtypes (Bastiaanse & Jonkers, 1998; Fergadiotis & Wright, 2016; Herbert et al., 2008; Kambanaros, 2010; Mayer & Murray, 2003; Pashek & Tompkins, 2002; Ulatowska et al., 2003). More accurate relationships between these variables, especially for the groups who currently perform at ceiling on existing tests, could be more accurately characterized with the use of computerized adaptive naming assessments (e.g., Fergadiotis, Kellough, & Hula, 2015; Hula, Kellough, & Fergadiotis, 2015).

These results should be interpreted within the context of the WAB-R classification scheme, as we have emphasized, and with caution. Importantly, to determine aphasia subtype, the WAB-R utilizes a subjective fluency rating with questionable reliability (e.g., Hillis Trupe, 1984; Hula, Donovan, Kendall, & Gonzalez-Rothi, 2010), meaning that both interpretation of these findings by subtype as well as their prognostic application depends heavily on confidence in classification accuracy. In addition, the potential moderating contribution of severity to these relationships deserves further exploration. As is evident in Table 2, although average severity scores followed the expected patterns by subtype, the reality is more nuanced. Most persons with Broca's, Wernicke's, and conduction aphasia were moderate in severity, and most persons with anomic aphasia were mild in severity. However, perhaps by a function of differences in sample size and/or the WAB-R classification scheme, the range of severity within each WAB-R subtype was varied (e.g., Broca's aphasia subtype included very severe, severe, moderate, and mild, whereas Wernicke's aphasia subtype included only severe and moderate), likely contributing also to the considerable variance within each group. As AphasiaBank continues to grow, it will be possible to not only increase the sample size for each subtype (and include individuals with global and transcortical aphasias) but also increase the representation of different severities within each subtype. This will facilitate more rigorous investigations into the relationship of WAB-R aphasia subtype and severity and functional communication abilities.

Although there are advantages to database research, it also introduces disadvantages that can only partially be counteracted by the increased sample size it provides. Although these instruments are commonly utilized in research and clinical practice, there remain psychometric shortcomings (e.g., WAB-R classification discussion above); furthermore, we cannot ensure or report on assessment fidelity components such as assessor qualifications, assessor training, assessment protocol adherence, rater training, errors in coding, and so forth (Richardson, Dalton, Shafer, & Patterson, 2016) that can introduce variance to the naming assessment (BNT-15, VNT) and WAB-R results in this study. In addition, as previously mentioned, this study is restricted to measures available in the database, which shaped the approach of dividing by WAB-R subtype. Although we recognize the limitations of the WAB-R, using it as a lens through which to examine differences associated with specific constellations of deficits assists in communicating results about discourse measures to a broad audience because it is commonly used and was recently suggested as a core outcome measure for aphasia (Wallace, Worrall, Rose, & Le Dorze, 2016a). Regarding discourse measures, in previous work (Richardson & Dalton, 2016), we have found high protocol fidelity in the AphasiaBank database across locations and assessors. In addition, the accuracy of transcription is excellent, with fewer than 1% of content words spoken in the audio/video file omitted, added, or changed in the transcript files examined.

In this study, the focus was on gist, or the ability to express essential concepts, as this discourse variable seems to be well aligned with the ideas of functional communication and successful communication of messages. Recall that this discourse variable, as captured with MCA, demonstrated good ecological validity—MCA is correlated with informativeness, communication adequacy, and listener perceptions. Similarly, the CIU analysis explained only under half of the variance in confrontation naming scores in data derived by sampling the same database (Table 1; Fergadiotis & Wright, 2016). As the utility of picture naming continues to be explored in larger samples and with longer naming assessments, we suggest that its relationship to gist should continue to be explored if naming is to be an adequate surrogate for functional communication. Utilizing the findings reported here, as well as the information in Table 1, multidimensional DAs could be designed to maximize our ability to define the relationship between confrontation naming and functional communication. MCA is restricted to tasks that have shared knowledge for which MCs can be readily determined and is not suitable for investigating free speech. It would be advantageous to extend these investigations to other macrolinguistic measures in conversation (e.g., cohesion, coherence, story grammar) as well as to investigate the relationship of MCA in semi-spontaneous discourse tasks to other ecologically valid measures of conversation.

Our results may provide some helpful insights into reasonable expectations regarding treatment generalization. For decades, researchers have designed interventions to treat at the impairment level and have tested for generalization from treated to untreated items and from treated linguistic levels (e.g., single word) to untreated linguistic levels (e.g., connected speech). This second kind of generalization, across linguistic level generalization (Webster, Whitworth, & Morris, 2015), has generally remained elusive. Given the weak-to-absent correlations between naming and gist in some WAB-R aphasia subtypes, assuming classification accuracy, it may be unreasonable to expect this type of generalization, because this relationship either is not straightforward or cannot be captured with the measures utilized. Recalibrating such expectations, and basing hypotheses on information such as those presented here, will be important for improved methodology in future studies.

Acknowledgments

This work was supported by NIH NIGMS P20GM109089, awarded to Jessica D. Richardson.

The authors would like to thank AphasiaBank developers and contributors and graduate students at the University of New Mexico for their valuable contribution to this project. They would like to extend their gratitude also to the Chapman Foundation, which supported their first discourse project that eventually led to additional projects such as this.

References

Andreetta, S., Cantagallo, A., & Marini, A. (2012). Narrative discourse in anomic aphasia. *Neuropsychologia, 50*(8), 1787–1793.

Bastiaanse, R., Edwards, S., & Rispens, J. (2002). *VAST: The Verb and Sentence Test. Manual*. Suffolk, UK: Thames Valley Test Company.

Bastiaanse, R., & Jonkers, R. (1998). Verb retrieval in action naming and spontaneous speech in agrammatic and anomic aphasia. *Aphasiology, 12*(11), 951–969.

Benson, D. F. (1988). Anomia in aphasia. *Aphasiology, 2*(3/4), 229–236.

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. (2012). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews, 2012*(5), Article No. CD000425. https://doi.org/10.1002/14651858.CD000425.pub3

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, 2016*(6), Article No. CD000425. https://doi.org/10.1002/14651858.CD000425.pub4

Brookshire, R. H., & Nicholas, L. E. (1994). Speech sample-size and test-retest stability of connected speech measures for adults with aphasia. *Journal of Speech and Hearing Research, 37*(2), 399–407.

Brown, K., Worrall, L. E., Davidson, B., & Howe, T. (2012). Living successfully with aphasia: A qualitative meta-analysis of the perspectives of individuals with aphasia, family members, and speech-language pathologists. *International Journal of Speech-Language Pathology, 14*(2), 141–155.

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

Chapey, R., Duchan, J. F., Elman, R. J., Garcia, L. J., Kagan, A., Lyon, J. G., & Simmons-Mackie, N. (2001). Life participation approach to aphasia: A statement of values for the future. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders* (5th ed., pp. 279–289). Baltimore, MD: Lippincott Williams & Wilkins.

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology, 26*(10), 1250–1277.

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.

del Toro, C. M., Bislick, L. P., Comer, M., Velozo, C., Romero, S., Rothi, L. J. G., & Kendall, D. L. (2011). Development of a short form of the Boston naming test for individuals with aphasia. *Journal of Speech, Language, and Hearing Research, 54*(4), 1089–1100.

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.

Elman, R. J. (2016). Aphasia centers and the life participation approach to aphasia: A paradigm shift. *Topics in Language Disorders, 36*(2), 154–167.

Enderby, P. (2006). *Therapy outcomes measures for rehabilitation professionals*. Hoboken, NJ: Wiley.

Fergadiotis, G., Kellough, S., & Hula, W. D. (2015). Item response theory modeling of the Philadelphia Naming Test. *Journal of Speech, Language, and Hearing Research, 58*(3), 865–877.

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology, 25*(11), 1414–1430.

Fergadiotis, G., & Wright, H. H. (2016). Modelling confrontation naming and discourse performance in aphasia. *Aphasiology, 30,* 364–380.

Forbes, M., Fromm, D., & MacWhinney, B. (2012). AphasiaBank: A resource for clinicians. *Seminars in Speech and Language, 30*(3), 217–212.

Goodglass, H., & Geschwind, N. (1976). Language disorders (aphasia). *Handbook of Perception, 7,* 389–428.

Goodglass, H., & Kaplan, E. (1983). *Boston Diagnostic Aphasia Examination (BDAE)*. Philadelphia, PA: Lea & Febiger.

Goodglass, H., & Wingfield, A. (1997). *Anomia: Neuroanatomical and cognitive correlates*. San Diego, CA: Academic Press.

Grinnon, S. T., Miller, K., Marler, J. R., Lu, Y., Stout, A., Odenkirchen, J., & Kunitz, S. (2012). National Institute of Neurological Disorders and Stroke Common Data Element Project: Approach and methods. *Clinical Trials, 9*(3), 322–329.

Herbert, R., Hickin, J., Howard, D., Osborne, F., & Best, W. (2008). Do picture-naming tests provide a valid assessment of lexical retrieval in conversation in aphasia? *Aphasiology, 22*(2), 184–203.

Hillis Trupe, E. (1984). Reliability of rating spontaneous speech in the Western Aphasia Battery: Implications for classification.

In *Clinical Aphasiology: Proceedings of the Conference 1984* (pp. 55–69). Minneapolis, MN: BRK Publishers.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences*. Boston, MA: Houghton Mifflin.

Holland, A. (1980). *Communicative abilities in daily living*. Baltimore, MD: University Park Press.

Holland, A., Frattali, C., & Fromm, D. (1998). *Communication activities of daily living* (2nd ed.). Austin, TX: Pro-Ed.

Hopper, T., Holland, A., & Rewega, M. (2002). Conversational coaching: Treatment outcomes and future directions. *Aphasiology, 16*(7), 745–761.

Huber, W., Poeck, K., & Wilmes, K. (1983). The Aachen Aphasia Test. *Advances in 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.

Hula, W. D., Kellough, S., & Fergadiotis, G. (2015). Development and simulation testing of a computerized adaptive version of the Philadelphia Naming Test. *Journal of Speech, Language, and Hearing Research, 58*(3), 878–890.

Kambanaros, M. (2010). Action and object naming versus verb and noun retrieval in connected speech: Comparisons in late bilingual Greek–English anomic speakers. *Aphasiology, 24*(2), 210–230.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test*. Philadelphia, PA: Lea and Febiger.

Kertesz, A. (2007). *The Western Aphasia Battery–Revised*. San Antonio, TX: Pearson Corporation.

Kohn, S. E., & Goodglass, H. (1985). Picture-naming in aphasia. *Brain and Language, 24*(2), 266–283.

Kong, A. (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.

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The Communicative Effectiveness Index: Development and psychometric evaluation of a functional communication measure for adults. *Journal of Speech and Hearing Disorders, 54*(1), 113–124.

Mack, W. J., Freed, D. M., Williams, B. W., & Henderson, V. W. (1992). Boston Naming Test: Shortened versions for use in Alzheimer's disease. *Journal of Gerontology: Psychological Sciences, 47,* 154–158.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B., & Fromm, D. (2016). AphasiaBank as BigData. *Seminars in Speech and Language, 37*(1), 10–22.

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

Mayer, J., & Murray, L. (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. *Aphasiology, 17*(5), 481–497.

Nicholas, L. E., & Brookshire, R. H. (1993). A system for scoring main concepts in the discourse of non-brain-damaged and aphasic speakers. *Clinical Aphasiology Conference, 21,* 87–99.

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.

Pashek, G. V., & Tompkins, C. A. (2002). Context and word class influences on lexical retrieval in aphasia. *Aphasiology, 16*(3), 261–286.

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. H., Shafer, J., & Patterson, J. (2016). Assessment fidelity in aphasia research. *American Journal of Speech-Language Pathology, 25*(4S), S788–S797.

Ross, K. B., & Wertz, R. T. (1999). Comparison of impairment and disability measures for assessing severity of, and improvement in, aphasia. *Aphasiology, 13*(2), 113–124.

Sarno, M. T. (1969). *The Functional Communication Profile: Manual of directions*. New York, NY: New York Institution of Rehabilitation Medicine.

Simmons-Mackie, N., & Kagan, A. (2007). Application of the ICF in aphasia. *Seminars in Speech and Language, 28*(4), 244–253.

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

Taylor, M. L. (1965). A measurement of functional communication in aphasia. *Archives of Physical Medicine and Rehabilitation, 46,* 101–107.

Ulatowska, H., Olness, G. S., Wertz, R., Samson, A., Keebler, M., & Goins, K. (2003). Relationship between discourse and Western Aphasia Battery performance in African Americans with aphasia. *Aphasiology, 17*(5), 511–521.

Ulatowska, H. K., North, A. J., & Macaluso-Haynes, S. (1981). Production of narrative and procedural discourse in aphasia. *Brain and Language, 13*(2), 345–371.

Wallace, S. J., Worrall, L., Rose, T., & Le Dorze, G. (2016a). *Improving research outcome measurement in aphasia: Development of a core outcome set*. Platform presented at International Aphasia Rehabilitation Conference, London, England.

Wallace, S. J., Worrall, L., Rose, T., & Le Dorze, G. (2016b). Which treatment outcomes are most important to aphasia clinicians and managers? An international e-Delphi consensus study. *Aphasiology, 31,* 643–673.

Webster, J., Whitworth, A., & Morris, J. (2015). Is it time to stop "fishing"? A review of generalisation following aphasia intervention. *Aphasiology, 29*(11), 1240–1264.

World Health Organization. (2007) *International classification of functioning. Disability and health*. Geneva, Switzerland: Author.

Worrall, L., Sherratt, S., Rogers, P., Howe, T., Hersh, D., Ferguson, A., & Davidson, B. (2011). What people with aphasia want: Their goals according to the ICF. *Aphasiology, 25*(3), 309–322.

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