The cardinal deficit of people with aphasia (PWA) is anomia (Goodglass & Wingfield, 1997). In single word retrieval, as in picture naming tasks, this deficit is believed to be indicative of disruption in two cognitive processes: (i) accessing a semantic description of the target concept, and/or (ii) retrieval of a fully phonologically specified representation (e.g., Dell, 1986). During discourse, in addition to these core processes that serve word retrieval of single words, production also depends on "…factors external to the lexicon…" (p. 169, Wilshire & McCarthy, 2002). The latter processes might influence the selection of lexical items based on syntactic, structural, and/or pragmatic criteria that can be either automatic or meta-cognitive. The current study investigates the implicit assumption that performance in single-word, picture naming tasks is directly and strongly related to word retrieval performance during discourse production.

To establish a diagnosis and quantify its severity, and determine the effect of treatment in various communication disorders, speech language pathologists often use confrontation naming tests (CNTs). In CNT's, basic drawings or pictures are presented to the PWA who is asked to name its target. The results of the test are then used to determine what steps should be taken post-injury to support the client's word-production/word retrieval process and which therapeutic approach may maximize the rehabilitation outcome.

However, according to Herbert, et al (2008) CNTs may not fully take into account their non-native disposition. First, in typical conversation speakers do not name pictured objects. Also, in CNTs, examinees name bare nouns or verbs and there are no elements in which to attach these words (i.e. not conversational context). Second, the main ideas communicated in discourse may not be necessarily planned: based on Dell's model, access to word specific semantic features, retrieval of the word form, and encoding the corresponding phonemes of that word are all part of the natural steps that occur, typically without premeditated action in the healthy, non-injured brain (Martin, 2012). Based on these premises authors have argued that decontextualized tasks such as CNT's may "… [bear] little resemblance to the online, multifaceted word retrieval required during conversation" (Mayer & Murray, 2003, p. 482). This position carries significant clinical and research implications because it directly challenges the idea of using CNT's to make inferences about discourse production; and, argues that perhaps the decontextualized nature of such tests may mislead professionals when diagnosing and treating PWA if they are to rely solely upon them to make inferences about discourse production.

The specific aims of this study were:

1. To assess whether there is a relationship between performance in CNT's and the proportion of paraphasias in three different types of discourse when accounting for construct irrelevant variance (i.e. random noise and irrelevant systematic variance).
2. To determine the magnitude of the relationship between error free estimates of word retrieval at the single and discourse level.
3. To determine the relationship between observed scores in CNT's and number of paraphasias in discourse.

Method

Participants. Data from 98 PWA were retrieved from AphasiaBank (MacWhinney, Fromm, Forbes, & Holland, 2011). All participants had aphasia secondary to a left hemisphere stroke. PWA met the following inclusion criteria: (a) chronic aphasia (minimum = 6 months post onset);

(b) no reported history of psychiatric or neurodegenerative disorders; (c) aided or unaided normal hearing acuity; (d) corrected or uncorrected normal visual acuity; and (e) English as their primary language (Table 1).

Confrontation Naming Tests. All PWA were administered the Western Aphasia Battery-Revised (Kertesz, 2007), the Short Form of the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 2001), and the Verb Naming Test (Cho-Reyes & Thompson, 2012). Performance on these variables was used to form a latent factor representing the common cognitive processes that are engaged during picture naming.

Discourse Stimuli & Instructions. Language samples were collected in a single session. The responses to first three discourse tasks of the AphasiaBank Protocol were used (http://talkbank.org/APhasiaBank/protocol/list.pdf): (i) free speech, ii) picture description, (iii) story telling (Cinderella).

Transcription & Language Sample Preparation. Samples were digitally recorded and orthographically transcribed in the CHAT format that is compatible with Computerized Language Analysis (CLAN; MacWhinney, 2000). Samples were coded using word-level codes to indicate the different types of paraphasias. Following Schwartz, Dell, Martin, Gahl, & Sobel (2006) the following types of paraphasias were included in the analysis: semantic, formal/phonological, neologisms, and mixed. The number of words for each type of discourse was estimated and proportions of paraphasias in each type of discourse were calculated. Performance on these variables was used to estimate a latent factor representing the common cognitive processes that are used during discourse production. (See Table 2 for descriptive statistics and intercorrelations).

Analysis. The structural equation model (SEM) in Figure 1 was estimated in Mplus 6.1. The model isolated the systematic common variance accounted for by two unique single constructs, separating them from construct irrelevant noise and extraneous factors (e.g., test unreliability). Then, the factor scores formed based on confrontation naming tests were used to predict the factor scores that were based on discourse production. The model was estimated using the robust maximum likelihood to account for non-normality. Missing data (~2%) were accommodated using direct maximum likelihood. Based on Kline (2010) four fit indices were taken into account to examine global model fit: the Satorra-Bentler scaled $\chi 2$ statistic ; the comparative fit index (CFI); the root-mean square error of approximation (RMSEA); and, the standard root mean residual (SRMR). A good fitting model was expected to have a non-significant χ^2 at the .05 level*;* a CFI value higher than .95; an RMSEA value below .08, and an SRMR value less than .05.

Results & Discussion

The model estimation was terminated normally and demonstrated adequate fit. Fit statistics and standardized parameter estimates can be seen in Figure 2. With respect to our research questions:

(1&2) We found a statistically significant relationship between the underlying variables that determine performance in CNT's and discourse production with respect to the proportion of paraphasias produced. However, the relationship between the constructs, although robust (Γ =

.48) still suggests that there is a significant proportion of variance in how PWA deploy lexical items and produce paraphasias during discourse that is not captured by CNT's. Therefore, conclusions about expected proportion of paraphasias in discourse based on performance in CNT's are not well supported.

(3) Further, standard path analytic procedures (Bollen, 1989) employed in the model can provide estimates of the relationship between any two *observed* variables. For example, the relationship between BNT and proportion of paraphasias in story telling is equal to the product (λ_1 * Γ * λ_6) = .90 * .48 * .87 = .37. In practice, for clinicians and researchers who do not have access to large samples and techniques such SEM and rely on observed scores for a single individual at a time, inferences maybe even less justified because they do not have the resources to partial noise from the observed scores.

Results will be discussed within a theoretical framework of psychometric measurement (Bachman, 2003; Kane 2009; Mislevy, 2006) which describes how assessing psychological abilities may be viewed as a process of evidentiary reasoning, which in turn constitutes a special case of argument (Toulmin, 1958). Emphasis will be placed on implications for researchers and clinicians.

References

Bachman, L. F. (2003). Building and supporting a case for test use. *Language Assessment Quarterly*, *2*, 1–34.

Bollen, K. A. (1989). *Structural equation modeling with latent variables*. New York: Wiley.

Cho-Reyes, S. & Thompson C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology*, *iFirst*, 1–28.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283.

Dell, G.S, Schwartz M. F., Martin N., Saffran E.M., & Gagnon D.A. (1997). Lexical access in aphasic and non aphasic speakers. *Psychological Review*. *104*, 801–838.

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

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.

Kane, M. T. (2009). Validating the interpretations and uses of test scores. (pp.39- 64). In, R. W. Lissitz (Ed.), *The concept of validity. Revisions, New directions, and Applications.* Charlotte, NC: Information Age Publishing.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). Boston naming test. Philadelphia: Lea and Febiger.

Kertesz, A. (2007). Western Aphasia Battery -R. New York: Grune & Stratton.

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3^{rd} ed.). New York, NY: Guilford Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analysing talk, Volume 1: Transcription format and programs* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Mayer, J. F., & Murray, L. L. (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. *Aphasiology, 17*(5), 481-497. doi:10.1080/02687030344000148

Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 257–305). Westport, CT: American Council on Education/Praeger Publishers.

Schwartz M.F., Dell G.S., Martin N., Gahl S., Sobel P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language. 54,* 228– 264.

Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

Wilshire, C. E., & McCarthy, R. A. (2002). Evidence for a context-sensitive word retrieval disorder in a case of nonfluent aphasia. *Cognitive Neuropsychology, 19*(2), 165. doi:10.1080/02643290143000169

Tables

Table 1

Participants' Demographic Information

Characteristic	PWA (*N* = 98)
Gender Ratio	50M: 48F
Mean Age in Years	64.18 (12.72)
Ethnicity	
African-American	6
Asian	2
Hispanic	3
Other	4
White	79
Education level completed	
Some high school	5
12th grade	23
Some college	15
Bachelor's or higher	55
WAB-R Aphasia Classification	
Anomic	33
Broca	22
Conduction	22
Wernicke	14
Global	4
Transcortical Motor	3
Mean Aphasia Duration in Years	5.59 (6.23)
Mean WAB-R AQ	70.42 (17.03)

Note. Ethnicity information was unavailable for four individuals, and education information was unavailable for four individuals; WAB-R AQ = Western Aphasia Battery – Revised Aphasia Quotient; *SD*'s are shown in parentheses.

Table 2

Summary Statistics and Correlations Among Study Variables

	BNT	WAB Naming	VNT	Free Speech[a]	Picture Description[b]	Story[c]
BNT	1					
WAB	0.82	1				
VNT	0.73	0.74	1			
FREE	-0.37	-0.39	-0.46	1		
DES	-0.42	-0.45	-0.45	0.79	1	
STO	-0.28	-0.27	-0.32	0.79	0.77	1
Mean	7.09	42.76	14.53	0.02	0.04	0.03
Skewness	0.11	-0.90	-0.72	4.63	3.15	3.01
Kurtosis	-1.20	-0.30	-0.50	29.37	13.75	11.73

All correlations were statistically significant, $p < .001$.

[a, b, & c] are the proportions of paraphasias per number of words observed in each type of discourse

Figures

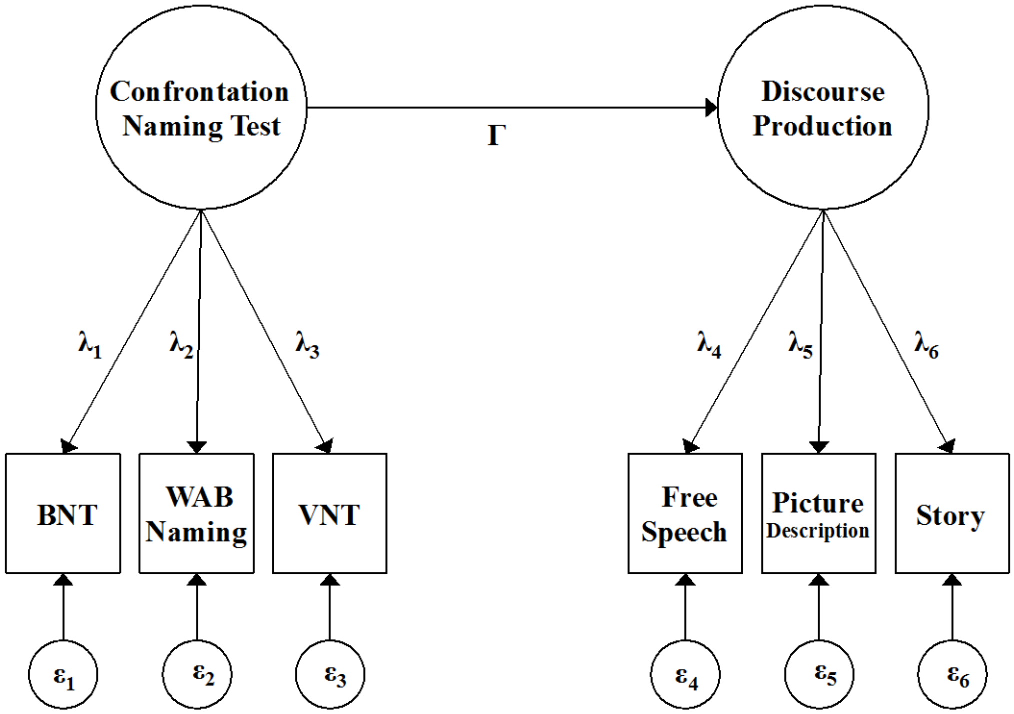

Figure 1. Two latent variables underlying the observed variables for (i) confrontational naming and (ii) word retrieval processing reflected in paraphasias during discourse production. For each each factor, the variables are conditionally independent after accounting for their respective common factor. The parameter **Γ** captures the utility of performance in confrontation naming to predict proportions of paraphasias during discourse production. The loadings (λ_j) reflect the strength of the relationship between the observed variable (e.g., BNT) and the construct it purports to measure minus any construct irrelevant variance partitioned in the error terms (i.e. ε_j).

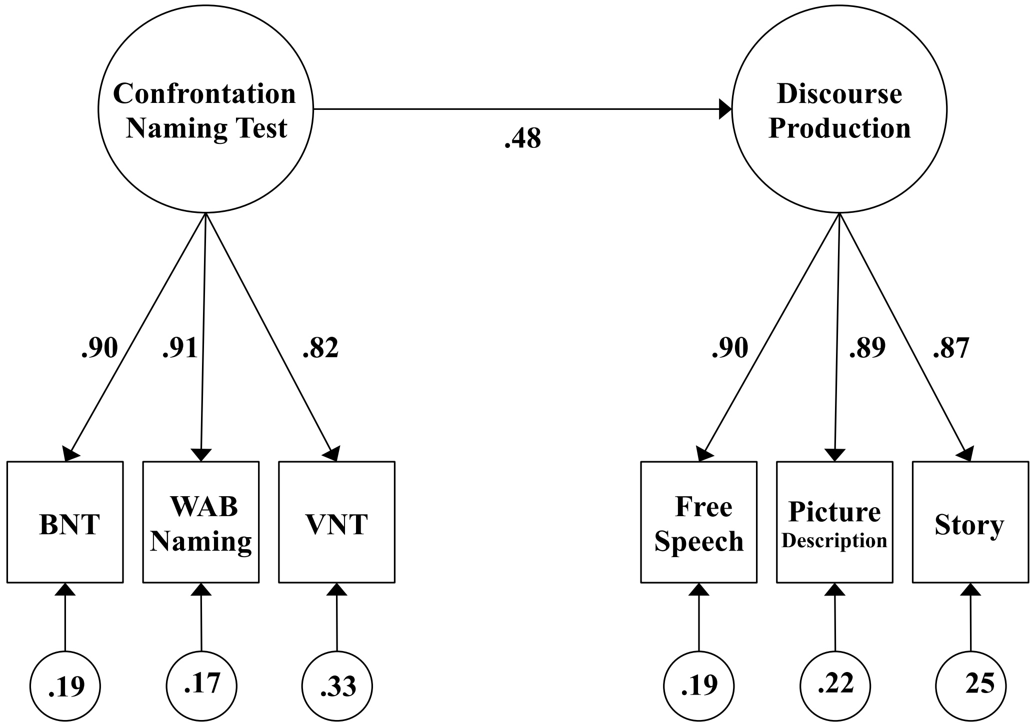

Figure 2. Fit and model parameters. χ^2 = 15.34, p = .053 ; CFI =.986; RMSEA = .097; and, SRMR = .043. For each observed variable, the variance accounted for by the common factor, R^2 = (1 – residual variance) = λ^2. For all parameter estimates, $p < .001$. Γ is a standardized regression coefficient such that Γ^2 = variance shared between the two factors.