

**Title:** The thorny issue of choosing outcome measures for aphasia therapy trials: A comparison of discourse and aphasia battery outcomes following Multi-modality and Constraint Induced aphasia therapy.

**Background and aim:** Constraint induced aphasia treatments are efficacious on omnibus measures of aphasia severity, and on global proxy ratings of communication activity (Cherney et al., 2009). Recently, multi-modality aphasia treatment was found to be equally efficacious on similar measures (Rose et al., 2013). Conversation is a frequent genre in human communication, and people with aphasia have indicated that successful conversation is a major rehabilitation goal for them (Worrall et al., 2011). While treatment efficacy has typically been measured on picture naming tasks and aphasia batteries, performance on these measures may not predict performance in connected speech (Carragher et al., 2012; Mayer & Murray, 2003). Thus, valid treatment efficacy comparisons should include the impact of treatments on discourse. However, it remains unclear which discourse genres provide the most valid and reliable outcome measure. Therefore, in this study we aimed to investigate the comparative impact of M-MAT and CIAT Plus on three discourse genres and compared these to outcomes on four standardised measures.

**Methods:** Eleven people with chronic aphasia participated in 2 weeks of intensive CIAT Plus and 2 weeks of M-MAT (30 hours over 2 weeks), with a one-week break between the two intensive treatment blocks (Rose et al., 2013). Treatment stimuli were 160 nouns and verbs depicted in line drawings on cards. Table 1 describes the participant characteristics. Five participants undertook CIAT Plus first, while six undertook M-MAT first. Discourse samples from three different conditions were collected from each participant (picture description from the Western Aphasia Battery, WAB (kertesz, 1982); Cinderella story retell; semi-structured conversation with one of the investigators) at three time points (pre-treatment, post CIAT Plus and post M-MAT). The first 16 minutes of each conversation sample was analysed. All

samples were video-recorded and orthographically transcribed according the Systematic Analysis of Language Transcript conventions (Miller & Iglesias, 2010). In addition to the discourse measures, participants were assessed on the WAB, the Boston Naming Test (BNT) (Kaplan et al., 2000), the Scenario Test (van der Meulen et al., 2010), and the Communicative Effectiveness Index (CETI) (Lomas et al., 1989) at all three time points.

*Data Analysis:* We analysed the number of tokens of substantive nouns and verbs in each discourse condition. Substantive nouns include all proper and common nouns (Halliday & Matthiessen, 2004). Substantive verbs exclude auxiliaries, all forms of the verbs *be*, *have*, *do*, and any light verbs (Huddleston & Pullman, 2002). Point-to-point inter-rater agreement was calculated on 20% of randomly selected transcripts. High inter-rate agreement was demonstrated: Tokens of substantive nouns: 96.5%; Tokens of substantive verbs: 95.7%.

**Results:** Figure 1 provides an overview of the verb and noun counts respectively for all three discourse genres for the participants who received CIAT Plus first and Figure 2 for participants who received M-MAT first. *There was considerable variability in results within and across participants and also across the three discourse conditions;* although every participant demonstrated positive change in either nouns or verbs in at least one discourse condition. In the conversation condition, 8 participants showed improvements on nouns and 3 on verbs, in the story retell, 8 participants showed improvements on nouns and 9 on verbs, while in picture description 6 participants showed improvements on nouns and 5 on verbs (11 improvements overall). Overall, there was no clear discourse advantage for M-MAT or CIAT Plus.

The results on the WAB are described in Table 2 for the participants who received M-MAT first and Table 3 for those who received CIAT Plus first. In this study, we chose a somewhat conservative AQ change score of 3 points overall and/or a 1 point change on either the

fluency or information rating score (Spontaneous Speech section) to reflect treatment responsiveness (Hula et al., 2010).. Eight of the eleven participants demonstrated improvement immediately post treatment on the WAB AQ for at least one time point. A comparison of WAB AQ immediately following M-MAT as compared to immediately following CIAT Plus revealed four participants favoured M-MAT (> 2 point WAB AQ difference between M-MAT and CIAT Plus) and five participants favoured CIAT Plus. Order effects are likely to have played a significant role: seven participants achieved greater WAB AQ change scores following the first treatment than following the second treatment phase (compared to mid-phase scores).

The results for the Scenario Test, CETI, and BNT are in Tables 4 and 5. All participants showed an improvement in BNT scores on at least one time point. A SEM of 1.02 has been indicated for the BNT (Flanagan & Jackson, 1997), which was used to indicate meaningful change. Comparing the change scores immediately post M-MAT and post CIAT Plus indicated comparatively equal number of participants improved post M-MAT (8) and post CIAT Plus (7). Again, the results might reflect order effects: seven participants achieved greater BNT change scores following the first treatment than following the second treatment phase, and 4 following the second phase.

Fewer participants showed change on the Scenario Test Scores. Using the psychometric measures in van der Meulen et al. (2010) a SEM of 2.13 was calculated using the formula  $SEM = SD (\sqrt{1-ICC})$ . Only 3 participants showed an increase in score of 2 or more at any one point. Of these, 2 showed increases after both M-MAT and CIAT Plus, and one post CIAT Plus only. The SEM for the CETI is 5.87 (Lomas et al., 1989). Nine of the 11 participants demonstrated an increase in scores of  $\geq 6$  for at least one time point: 5 participants following M-MAT and 5 participants following CIAT Plus.

*Comparing improvements following each treatment for each participant, there was little consistency between the measures.* Table 6 provides a summary of improvements across all measures post M-MAT and CIAT Plus. This time, a meaningful improvement in a discourse genre was indicated by a >3 point increase in either nouns or verbs from the immediately previous count (Post 1<sup>st</sup> treatment-Baseline; Post 2<sup>nd</sup> treatment-Post 1<sup>st</sup> treatment). Comparing change scores between the WAB and the different discourse genres, a consistent indication of improvement (or lack of) was demonstrated for 8 (of the possible 22) data points for picture description, 10 for the story retell, and 12 for the conversation. As for change scores associated with the BNT, only 8 data points for picture description, 14 for the story retell, and 12 for the conversation matched the improvement indicators. Only 12 data points for picture description, 9 for the story retell, and 9 for the conversation matched the Scenario test findings. For the CETI, 11 data points were consistent with findings from picture description, 7 for the story retell, and 8 for the conversation.

**Discussion:** WAB-AQ, BNT, and CETI change scores suggest M-MAT and CIAT Plus are equally efficacious. However, participants varied in their individual responsiveness to and preference for each treatment. Positive change was demonstrated on discourse measures following both treatments, with no clear advantage of treatment type. There was little consistency between standardised and discourse measures, or between the three discourse measures for each individual. These inconsistent results suggest caution should be taken when interpreting outcomes from different measures. It is timely, as the field moves toward running greater numbers of RCT's requiring a single primary outcome measure, for a robust discussion and consensus on optimal outcome measures.

## Reference

- Carragher, M., Conroy, P., Sage, K., & Wilkinson, R. (2012). Can impairment-focused therapy change the everyday conversations of people with aphasia? A review of the literature and future directions.
- Flanagan, J. L., & Jackson, S. T. (1997). Test-retest reliability of three aphasia tests: Performance of non-brain-damaged older adults. *Journal of communication disorders, 30*(1), 33–43.
- Halliday, M. A. K., & Matthiessen, C. M. I. (2004). *An introduction to functional grammar*. New York: Hodder Arnold.
- Huddleston, R., & Pullum, G. K. (2002). *The Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- Kaplan, E., Goodglass, H., Weintraub, S., Segal, O., & van Loon-Vervoorn, A. (2001). *Boston naming test*. Pro-ed.
- Kertesz, A. (1982). *Western aphasia battery test manual*. Psychological Corp.
- 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 adult aphasia. *The Journal of speech and hearing disorders, 54*(1), 113–124.
- Mayer, J., & Murray, L. (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. *Aphasiology, 17*, 481–497.
- Miller, J., & Iglesias, A. (2010). *Systematic Analysis of Language Transcripts (SALT)*. Madison, WI: University of Wisconsin-Madison.
- Rose, M. L., Attard, M. C., Mok, Z., Lanyon, L. E., & Foster, A. M. (2013). Multi-modality aphasia therapy is as efficacious as a constraint-induced aphasia therapy for chronic aphasia: A phase 1 study. *Aphasiology, 27*(8), 938–971.

Van der Meulen, I., van de Sandt-Koenderman, W. M., Duivenvoorden, H. J., & Ribbers, G.

M. (2010). Measuring verbal and non-verbal communication in aphasia: reliability, validity, and sensitivity to change of the Scenario Test. *International Journal of Language & Communication Disorders*, 45(4), 424–435.

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, 309–322. doi:10.1080/02687038.2010.508530

Table 1. Summary of Participant Characteristics

Participant	Age	Gender	Education (years)	Stroke Type/ Lesion side	MPO	PreTx WAB AQ	Aphasia Type	Limb apraxia*	Apraxia of Speech <sup>+</sup>	Hemi- paresis	Handed- ness
RW	49	F	15	Left ischemic	77	92.8	Anomic	Absent	Absent	Right	Right
SS	59	F	16	Left (type n/a)	25	91.2	Anomic	Absent	Mild	None	Right
LV	69	M	15	Left (type n/a)	34	85.6	Anomic	None	Mild-Moderate	None	Right
JP	64	F	13	Left hemorrhagic	22	77.2	Anomic	Moderate	Very Mild	Right	Right
BH	39	M	15	Left ischemic	88	63.8	Broca's	Mild	Mild	Right	Right
ST	46	M	16	Left SAH	22	61.5	Broca's	Mild	Mild-Moderate	Right	Right
AC	64	F	17	Left ischemic	40	57.4	Conduction	Mild	Moderate-Severe	None	Right
JB	53	M	15	Left ischemic	17	56.8	Broca's	Mild- Moderate	Mild-Moderate	Right	Right
LM	74	F	15	Left ischemic	79	51.9	Broca's	Moderate	Moderate	None	Right
PD	56	M	19	Left ischemic	22	50.6	Broca's	Moderate	Mild	Right	Right
PK	66	M	10	Left ischemic	58	36.2	Broca's	None	Moderate-Severe	None	Right

Note: MPO: months post-onset; (type NA): type not available; \* Test of Limb Apraxia (Helm-Estabrooks, 1992); <sup>+</sup> Apraxia Battery for Adults (Dabul, 2000); SAH: sub-arachnoid haemorrhage

Table 2. Results of WAB-R at baseline, immediate post-treatment (M-MAT first)

Assessment	BH			RW			SS			LV			PK			JP		
	Pre Tx	Post M	Post C+	Pre Tx	Post M	Post C+	Pre Tx	Post M	Post C+	Pre Tx	Post M	Post C+	Pre Tx	Post M	Post C+	Pre Tx	Post M	Post C+
<b>WAB-R</b>																		
Aphasia quotient	63.8	66.2	<b>72.3</b>	92.8	91.9	<b>96.1</b>	91.1	<b>95.2</b>	92.2	85.6	87.6	<b>88.9</b>	36.2	<b>45.2</b>	<b>52.9</b>	77.2	<b>81.1</b>	<b>81.6</b>
Spontaneous speech																		
<i>Information content /10</i>	5	5	<b>7</b>	9	8	<b>10</b>	10	10	10	10	10	10	3	<b>5</b>	<b>7</b>	8	<b>9</b>	<b>9</b>
<i>Fluency /10</i>	5	<b>6</b>	<b>6</b>	9	9	9	9	9	9	9	9	9	4	4	4	6	6	6
Auditory verbal comprehension /10	8.3	8.0	7.6	10	9.7	10	9.2	9.3	9.2	8.0	9.2	8.75	5.7	7.1	8.05	8.7	8.85	9
Repetition/10	6.1	7.3	7.2	10	9.7	10	8.6	9.3	8.6	9.3	8.9	9.4	3.2	2.7	2.8	9.2	9.8	9.8
Naming and word finding																		
Total score /10	7.5	6.8	7.2	8.4	9.4	9.1	8.9	9.3	9.3	6.1	6.7	7.3	2.2	3.8	4.6	6.7	6.9	7
<i>Object naming/60</i>	49	44	49	57	59	58	57	59	56	42	43	50	12	24	27	46	46	42
<i>Word fluency /20</i>	8	12	11	7	15	13	16	14	17	2	6	4	6	5	8	5	7	9
<i>Sentence completion /10</i>	10	8	10	10	10	10	9	10	10	7	10	9	0	3	4	8	8	10
<i>Responsive speech /10</i>	8	4	2	10	10	10	7	10	10	10	8	10	4	6	7	8	8	9

Note: Pre Tx = Pre-treatment; Post M = Post M-MAT; Post C+ = Post CIAT Plus. WAB-R = Western Aphasia Battery–Revised; Bolded figures: treatment responsiveness (see text for definition).



Table 3. Results of WAB-R at baseline, immediate post-treatment (CIAT Plus first)

Assessment	JB			ST			LM			AC			PD		
	Pre Tx	Post C+	Post M	Pre Tx	Post C+	Post M	Pre Tx	Post C+	Post M	Pre Tx	Post C+	Post M	Pre Tx	Post C+	Post M
<b>WAB-R</b>															
Aphasia quotient	56.8	55	53.8	65.5	<b>70</b>	65.8	51.9	<b>59.1</b>	<b>60.9</b>	57.4	56.3	56.9	50.6	<b>54.2</b>	<b>54.8</b>
Spontaneous speech															
<i>Information content /10</i>	7	5	5	8	8	8	6	<b>8</b>	<b>8</b>	5	5	5	4	<b>6</b>	<b>5</b>
<i>Fluency /10</i>	4	4	4	4	4	4	4	4	4	6	5	5	3	3	<b>4</b>
Auditory verbal comprehension /10	7.5	7.2	6.6	7.55	8.4	7.9	4.8	8.15	7.75	8.1	8.95	7.75	7.6	7.3	6.4
Repetition/10	5.9	5.7	4.8	5.8	6.8	5.9	4.0	4.6	5.8	4.9	3.4	5.2	3.2	7.9	7.7
Naming and word finding															
Total score /10	6.0	5.6	6.5	7.4	7.7	7.1	4	4.8	4.9	4.7	5.8	5.5	3.2	2.9	3.3
<i>Object naming/60</i>	42	35	42	50	50	55	29	24	29	28	37	38	20	12	16
<i>Word fluency /20</i>	6	5	7	7	9	6	1	6	8	7	4	3	2	5	3
<i>Sentence completion /10</i>	5	8	8	7	10	6	4	8	6	4	7	7	6	8	6
<i>Responsive speech /10</i>	7	8	8	10	8	4	6	10	6	8	10	7	4	4	8

Note: Pre Tx = Pre-treatment; Post M = Post M-MAT; Post C+ = Post CIAT Plus. WAB-R = Western Aphasia Battery–Revised; Bolded figures: treatment responsiveness (see text for definition).

Table 4. Results of BNT, Scenario Test, and CETI at baseline, immediate post-treatment (M-MAT first)

	BH			RW			SS			LV			PK			JP		
	Pre Tx	Post M	Post C+	Pre Tx	Post M	Post C+	Pre Tx	Post M	Post C+	Pre Tx	Post M	Post C+	Pre Tx	Post M	Post C+	Pre Tx	Post M	Post C+
<b>BNT</b> /60	21	<b>27</b>	<b>42</b>	44	<b>59</b>	53	51	<b>56</b>	53	28	<b>40</b>	39	3	<b>5</b>	<b>9</b>	18	<b>30</b>	<b>32</b>
<b>Scenario Test</b> /54	32	<b>41</b>	<b>44</b>	54	54	54	54	54	53	54	54	51	38	38	38	48	45	48
<b>CETI</b> /100	79	<b>87</b>	86	60	60	<b>73</b>	68	-	<b>78</b>	93	98	<b>99</b>	28	31	28	29	<b>41</b>	43

Note: Pre Tx = Pre-treatment; Post M = Post M-MAT; Post C+ = Post CIAT Plus; BNT = Boston Naming Test; CETI = Communicative Effectiveness Index. Bolded figures: treatment responsiveness (see text for definition).

Table 5. Results of BNT, Scenario Test, and CETI at baseline, immediate post-treatment (CIAT Plus first)

	JB			ST			LM			AC			PD		
	Pre Tx	Post C+	Post M	Pre Tx	Post C+	Post M	Pre Tx	Post C+	Post M	Pre Tx	Post C+	Post M	Pre Tx	Post C+	Post M
<b>BNT</b> /60	7	<b>9</b>	<b>15</b>	28	<b>40</b>	31	10	<b>14</b>	7	10	9	<b>13</b>	2	<b>4</b>	0
<b>Scenario Test</b> /54	48	38	38	42	<b>44</b>	43	33	<b>40</b>	<b>42</b>	46	47	46	30	29	31
<b>CETI</b> /100	42	41	39	29	-	<b>62</b>	46	46	<b>52</b>	54	<b>60</b>	60	31	<b>44</b>	<b>64</b>

Note: Pre Tx = Pre-treatment; Post M = Post M-MAT; Post C+ = Post CIAT Plus; BNT = Boston Naming Test; CETI = Communicative Effectiveness Index. Bolded figures: treatment responsiveness (see text for definition).

Table 6. Summary of improvements across formal and discourse (combined nouns + verbs) measures

Participant	WAB	BNT	Scenario	CETI	Discourse			# Discourse Improve	
					Pic	Story	Conv	MMAT	CIAT
BH <sub>M</sub>	+	+	+	+	=	+ <sup>W,B,S,C</sup>	=	1	
BH <sub>C</sub>	+	+	+	=	+ <sup>W,B,S,C</sup>	+ <sup>W,B,S,C</sup>	na		2
RW <sub>M</sub>	=	+	=	=	= <sup>W,S,C</sup>	na	+ <sup>B</sup>	1	
RW <sub>C</sub>	+	=	=	+	+ <sup>W,C</sup>	+	= <sup>B,S</sup>		2
SS <sub>M</sub>	+	+	=	na	= <sup>S</sup>	+ <sup>W,B</sup>	+ <sup>W,B</sup>	2	
SS <sub>C</sub>	=	=	=	+	= <sup>W,B,S</sup>	= <sup>W,B,S</sup>	= <sup>W,B,S</sup>		0
LV <sub>M</sub>	=	+	=	=	+ <sup>B</sup>	+ <sup>B</sup>	= <sup>W,S,C</sup>	2	
LV <sub>C</sub>	=	=	=	+	+ <sup>C</sup>	= <sup>W,B,S</sup>	+ <sup>C</sup>		2
JP <sub>M</sub>	+	+	=	+	+ <sup>W,B,C</sup>	+ <sup>W,B,C</sup>	+ <sup>W,B,C</sup>	3	
JP <sub>C</sub>	=	+	=	=	= <sup>W,C,S</sup>	= <sup>W,C,S</sup>	= <sup>W,C,S</sup>		0
PK <sub>M</sub>	+	+	=	=	= <sup>S,C</sup>	+ <sup>W,B</sup>	+ <sup>W,B</sup>	2	
PK <sub>C</sub>	+	+	=	=	= <sup>S,C</sup>	= <sup>S,C</sup>	+ <sup>W,B</sup>		1
LM <sub>C</sub>	+	+	+	=	= <sup>C</sup>	+ <sup>W,B,S</sup>	+ <sup>W,B,S</sup>		2
LM <sub>M</sub>	=	=	+	+	+ <sup>S,C</sup>	= <sup>W,B</sup>	+ <sup>S,C</sup>	2	
AC <sub>C</sub>	=	=	=	+	+ <sup>C</sup>	+ <sup>C</sup>	= <sup>W,B,S</sup>		2
AC <sub>M</sub>	=	+	=	=	= <sup>W,S,C</sup>	+ <sup>B</sup>	+ <sup>B</sup>	2	
PD <sub>C</sub>	+	+	=	+	= <sup>S</sup>	= <sup>S</sup>	na		0
PD <sub>M</sub>	+	=	=	+	= <sup>B,S</sup>	= <sup>B,S</sup>	+ <sup>W,C</sup>	1	
ST <sub>C</sub>	+	+	+	na	=	na	+ <sup>W,B,S</sup>		1
ST <sub>M</sub>	=	=	=	+	= <sup>W,B,S</sup>	+ <sup>C</sup>	+ <sup>C</sup>	2	
JB <sub>C</sub>	=	+	=	=	+ <sup>B</sup>	+ <sup>B</sup>	+ <sup>B</sup>		3
JB <sub>M</sub>	=	+	=	=	+ <sup>B</sup>	+ <sup>B</sup>	= <sup>W,S,C</sup>	2	
#Imp. cond	11	15	5	10	9	13	13		
#Part. Imp	8	11	3	9	7	10	10		
#Consistent	-	-	-	-	40	40	41		
# Dis Co I	-	-	-	-	-	-	-	19	15

Note: <sub>M</sub> M-MAT change. <sub>C</sub> CIAT Plus change. + indicates improvement; = indicate no improvement. <sup>W</sup> Indicates consistency with WAB; <sup>B</sup> Indicates consistency with BNT; <sup>S</sup> Indicates consistency with Scenario Test. <sup>C</sup> Indicates consistency with CETI. WAB= Western Aphasia Battery; BNT = Boston Naming Test; CETI = Communicative Effectiveness Index. na = score not available. Bolded text: all three discourse condition change results matched standard test change result. #Imp. Cond: Total number of conditions where improvement seen (/22); #Part. Imp: Total number of participants showing improvement (/11); #Consistent: Total number of test change results consistent with the discourse change results (/88).



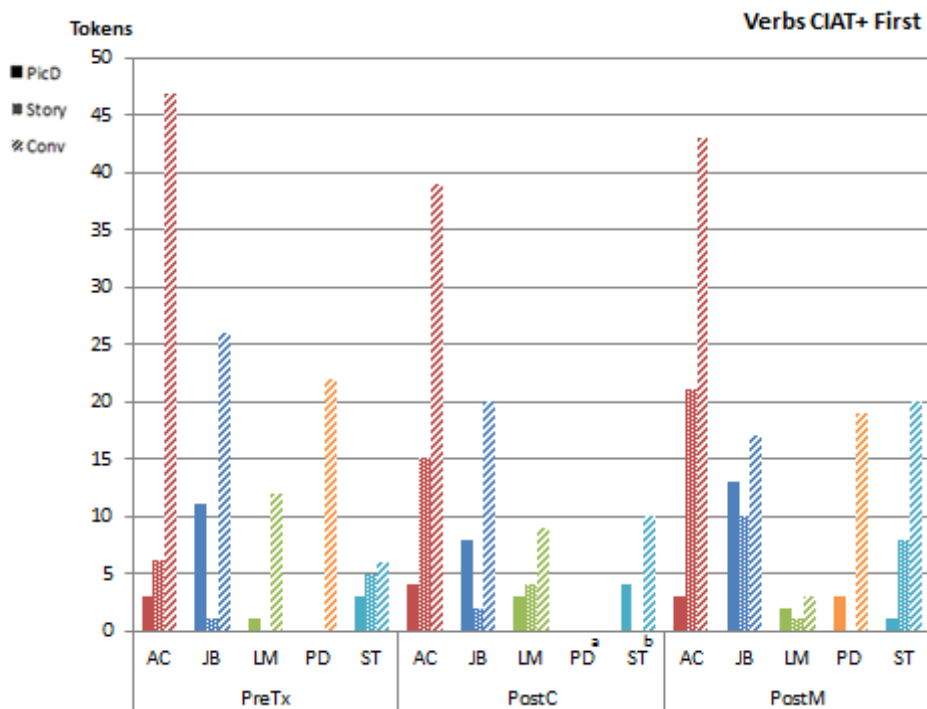
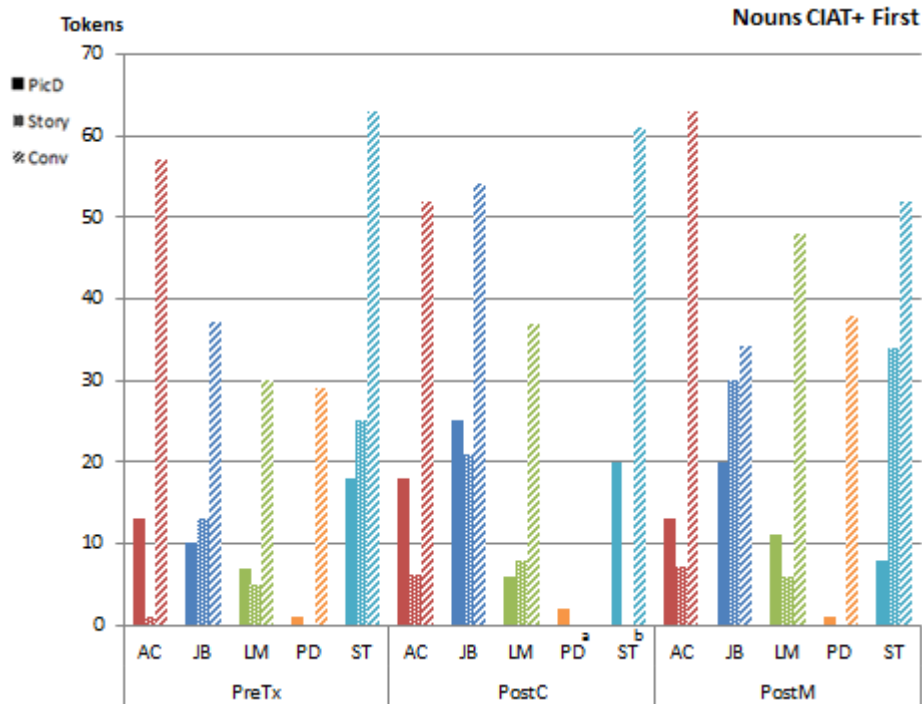


Figure 1. Substantive noun verb tokens across all three genres (CIATplus first).

*Notes.* PD did not produce any output for the story retell condition at any data point. <sup>a</sup> There is no data for the post CIAT+ conversation condition for PD. <sup>b</sup> There is no data for the post CIAT+ story retell condition for ST.

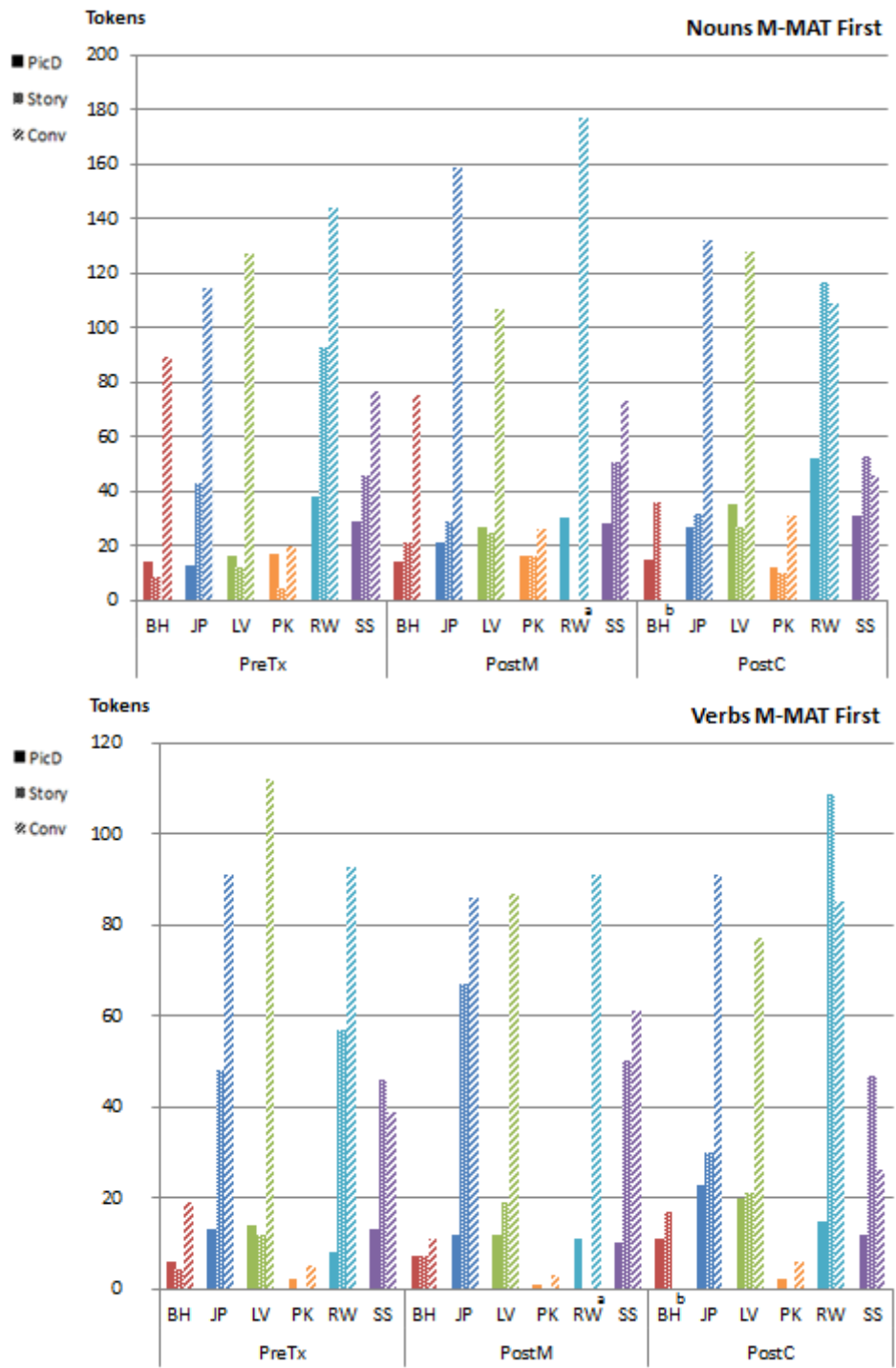


Figure 2. Substantive noun verb tokens across all three genres (M-MAT first)

Notes. <sup>a</sup> There is no data for the post M-MAT story retell condition for RW. <sup>b</sup> There is no data for the post CIAT+ conversation condition for BH.