

Stability of a Functional Measure of Word Retrieval in Narrative Discourse Proposal for CAC 2012

Interest in the ability of people with aphasia to communicate in natural contexts has resulted in examination of their discourse production. There is an expectation that changes in discourse production may reflect meaningful treatment-related changes. In a recent special issue of *Aphasiology*, Wright (2011) stated that it is clinically important for researchers to establish clinician friendly, reliable, and valid methods for analyzing discourse in order to quantify meaningful changes. An important part of this endeavor requires assessment of the stability of such methods. This is important for a number of reasons. Currently, clinical researchers often analyze discourse production only once prior to treatment and again at the end of treatment. This is a risky practice unless the stability of the discourse measurement is known *a priori*. If the stability is not known, there is no way that change in discourse can be attributed to the treatment with any degree of confidence unless multiple pre-treatment measurements are collected and analyzed. Additionally, if the discourse measurements are not relatively stable before treatment, a relatively large treatment effect may be necessary to demonstrate that change is related to treatment and not to the normal variability of the measurement. Because this may prove to be challenging, a more strategic approach might be to use a measurement that is relatively stable before treatment. Finally, apart from treatment, the issue of stability of discourse measurements is also important if those measurements are used to describe and analyze aspects of an individual's language impairment. If a measurement is not reasonably stable from session to session, it will probably not contribute to a valid assessment of the individual's impairment.

In 2003, Mayer and Murray proposed a measurement of word retrieval in connected speech that they described as functional, simple, and easily quantifiable. The measure, Percent Word Retrieval (%WR), was the percentage of noun and verb retrieval attempts that were successful. Mayer and Murray used %WR to measure word retrieval in two connected speech tasks: composite description and conversational discourse. They reported strong interjudge and intrajudge reliability of the scoring system. Mayer and Murray noted that the initial testing and subsequent analyses were completed "in the context of feasible clinical time demands" and suggested that its simplicity might make on-line data measurement possible. However, they did not provide information about the session-to-session stability of %WR. The purpose of this paper is to examine the stability of %WR across three sessions with no intervening treatment.

Method

Participants

The participants were 12 right-handed native-English aphasic speakers with anomia as a prominent characteristic in connected speech. None had other history of neurologic impairment and none received concomitant speech-language treatment. Table 1 contains demographic information and test results.

Procedures

Using the stimuli and procedures described by Nicholas and Brookshire (1993), discourses were elicited in three sessions separated by 2 to 7 days. All sessions were audiotaped then orthographically transcribed by a graduate student. The author independently checked the transcriptions. Disagreements were resolved prior to scoring. For purposes of this investigation, only narrative discourses (stories about picture sequences and complex scenes) were analyzed.

This differs from the stimuli used by Mayer & Murray (2003), who elicited composite descriptions of author-created picture sequences that depicted a series of events, each including multiple characters and activities. The Nicholas & Brookshire stimuli were chosen for this study because they are readily available and have been used in a number of aphasia treatment studies (e.g., Antonucci, 2009; Boyle, 2004; Wambaugh & Ferguson, 2007). Narrative discourses rather than composite descriptions were elicited because they seem more ecologically valid in terms of adult communication activities. Nicholas and Brookshire's (1993) procedures were used to count words. Mayer & Murray's (2003) procedures to score accurate and error noun and verb productions and to compute %WR were used. Following Mayer & Murray, the first 300 words of the narrative discourses were scored, and for those participants whose total output was fewer than 300 words, the entire narrative transcripts were scored.

Results & Discussion

Point-to-point interjudge and intrajudge agreement for %WR exceeded 85%. To assess the extent to which participants' %WR scores in Session 1 were related to their scores in subsequent sessions, Pearson product-moment correlations were calculated (Table 2 and Figures 1 and 2). Correlations for the percentage of nouns that were successfully retrieved ranged from .20 to .68; for verbs they ranged from .46 to .72. These scores represent weak to moderate correlations, suggesting that the stability of %WR is not ideal. The differences in %WR scores for nouns between sessions ranged from 0.57% to 24.45% (Table 3), meaning that for at least one participant the ability to retrieve nouns would have to improve by more than 25% before the change could be attributed to treatment. The differences in %WR for verbs between sessions ranged from 0% to 36.65%, meaning that for at least one participant, the ability to retrieve verbs would have to improve by more than 37% before the change could be attributed to treatment.

In light of these disappointing results, a second analysis of the data combining the noun and verb productions was done to see whether this would improve the stability of the scores across sessions. Pearson product-moment correlations for this analysis (Table 2 and Figure 3) ranged from .32 to .79, with differences between sessions ranging from 0.56% to 30.62% (Table 3). These results were essentially no different from those in the separate noun and verb analyses.

Although the %WR measurement may be simple and relatively quick to calculate once the discourse has been transcribed, its session-to-session variability makes it a questionable choice for measuring word retrieval in discourse or for measuring treatment-related changes. There is no obvious participant characteristic that an investigator can use *a priori* to determine whether a single administration with the measurement will produce a valid assessment of the participant's noun and verb retrieval. Mildly impaired participants (P1, P10) were as likely to produce scores that differed by more than 10% between sessions as were moderately impaired participants (P4, P12). While most participants with anomic aphasia had differences of less than 10% between sessions (P5, P7, P8), one (P6) had one difference score of more than 10%. Therefore, if an investigator wishes to use %WR as a measure of word retrieval in discourse, it should be assessed during at least two sessions in order to establish its stability. If %WR is to be used to assess treatment-related change, at least 3 baseline measurements prior to treatment should be collected in order to assess whether and how it changes in the absence of treatment.

There are several discourse-level measurements that provide information about word retrieval and that have demonstrated acceptable session-to-session stability. These include the measurements associated with correct information units that were developed by Nicholas and

Brookshire (1993) and a measurement of word finding difficulty (Boyle, 2010). Neither of these measurements could be described as clinician friendly, however, since they require meticulous transcription and precise coding, both of which are time-intensive endeavors. Percent Word Retrieval, requiring only that successful and unsuccessful noun and verb production attempts be counted, appeared to be a promising candidate for a reliable, clinician-friendly measurement. Unfortunately, the results of this study suggest that such a measurement remains elusive.

References

- Antonucci, S.M. (2009). Use of semantic feature analysis in group aphasia treatment. *Aphasiology*, 23, 854-866.
- Boyle, M. (2004). Semantic Feature Analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-Language Pathology*, 13, 236-249.
- Boyle, M. (2010). Variability of word-finding behavior and informativeness during discourse production in aphasia. Paper presented at the 2010 International Aphasia Rehabilitation Conference, Montreal, Canada 6/27/10.
- German, D.J. (1990). *The Test of Adolescent/Adult Word Finding*. Austin, TX: Pro-Ed.
- Goodglass, H. & Kaplan, E. (1983). *The Boston Diagnostic Aphasia Examination, 2nd Edition*. Boston: Lea & Febiger.
- Kertesz, A. (1982). *Western Aphasia Battery*. New York: The Psychological Corporation.
- Mayer, J.F. & Murray, L.L. (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. *Aphasiology*, 17, 481-497.
- Nicholas, L.E. & Brookshire, R.H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, 36, 338-350.
- Wambaugh, J.L. & Ferguson, M. (2007). Application of semantic feature analysis to retrieval of action names in aphasia. *Journal of Rehabilitation Research & Development*, 44, 381-394.
- Wright, H.H. (2011). Discourse in aphasia: An introduction to current research and future directions. *Aphasiology*, 25, 1283-1285.

Table 1. Participants' demographic information and test results.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Age	57	67	61	61	70	87	65	38	52	50	80	51
Gender	M	F	M	M	M	M	M	F	F	M	M	M
Education (years)	12	12	12	12	18	12	16	12	12	12	19	16
MPO	65	15	59	14	15	36	37	38	64	7	14	72
WAB AQ	82	*	67	54.5	90.6	72.2	86.6	67.4	70.2	82	61.2	46.3
Aphasia type	BA	BA	BA	BA	AA	AA	AA	AA	CA	CA	WA	WA
TAWF(106)	67	82	68	28	84	63.0	53	56	57	84	28	8
Standard Score	63	88	76	<70	90	72.0	<70	<52	<58	78	<70	<58
Percentile Rank	0.2	19	4	<1	23	2	<1	<0.1	<0.1	6	<1	<0.1
Etiology	LCVA	L CVA	L CVA	L CVA	L CVA	L CVA	L CVA	TBI	L CVA	L CVA	L CVA	L CVA

MPO = months post onset

WAB AQ= *Western Aphasia Battery* (Kertesz, 1982) Aphasia Quotient

* Severity Rating = 3, *Boston Diagnostic Aphasia Examination*, 2nd Edition (Goodglass & Kaplan, 1983)

BA = Broca's aphasia; AA = anomic aphasia; CA = conduction aphasia; WA = Wernicke's aphasia

TAWF = *Test of Adolescent/Adult Word Finding* (German, 1990)

Table 2. Pearson Product-Moment Correlation coefficients (r) for participants' performance between Sessions 1 and 2, Sessions 2 and 3, and Sessions 1 and 3 on %WR (the percentage of nouns and verbs that were accurately retrieved) during narrative discourse production tasks. %WR Nouns = the between-sessions correlation for the number of nouns that were retrieved accurately in a 300-word sample from narrative discourse tasks. %WR Verbs = the between-sessions correlation for the number of verbs that were retrieved accurately in a 300-word sample from narrative discourse tasks. %WR Nouns+Verbs = the between-sessions correlations for the combined nouns and verbs that were retrieved accurately in a 300-word sample from narrative discourse tasks.

	Session 1 – Session 2	Session 2- Session 3	Session 1 – Session 3
%WR Nouns	0.68	0.38	0.20
%WR Verbs	0.50	0.72	0.46
%WR Nouns+Verbs	0.79	0.63	0.32

Table 3. Differences in percentages of accurate word retrieval between sessions for each participant.

	Difference Scores %WR Nouns			Difference Scores %WR Verbs			Difference Scores %WR N+V		
	Sessions 1-2	Sessions 2-3	Sessions 1-3	Sessions 1-2	Sessions 2-3	Sessions 1-3	Sessions 1-2	Sessions 2-3	Sessions 1-3
P1	-10.88	-7.67	-18.55	-36.65	0	-36.65	-25.52	-5.1	-30.62
P2	4.1	15.15	19.25	15.48	16.28	31.76	9.37	14.43	23.8
P3	-12.05	-9.81	-21.86	33.33	-19.23	14.1	-5.52	-11.58	-17.1
P4	-0.57	10	9.43	-10.6	18.33	7.73	-3.33	13.33	10
P5	-8.99	-0.04	-9.03	10.76	9.07	19.83	-0.43	3.3	2.87
P6	1.29	8.73	10.02	-12.28	14.52	2.24	-5.71	12.32	6.61
P7	-0.79	-17.39	-18.18	-7.48	10.52	3.04	-4.81	-1.24	-6.05
P8	-15.48	11.17	-4.31	6.31	-14.66	-8.35	-7.03	0.83	-6.2
P9	1.74	22.71	24.45	-16.77	-4.23	-21	-4.2	16.7	12.5
P10	-13.49	-7.7	-21.19	-18.04	2.86	-15.18	-12.82	-4.76	-17.58
P11	5.42	6.25	11.67	-3.33	-6.67	-10	-0.56	-1.19	-1.75
P12	-22.26	23.4	1.14	-3.94	-2.83	-6.77	-16.84	9.88	-6.96

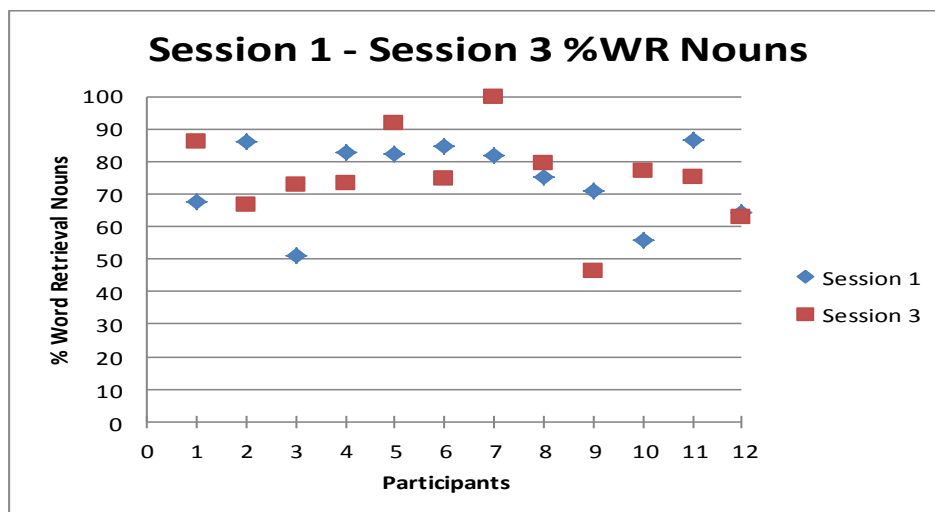
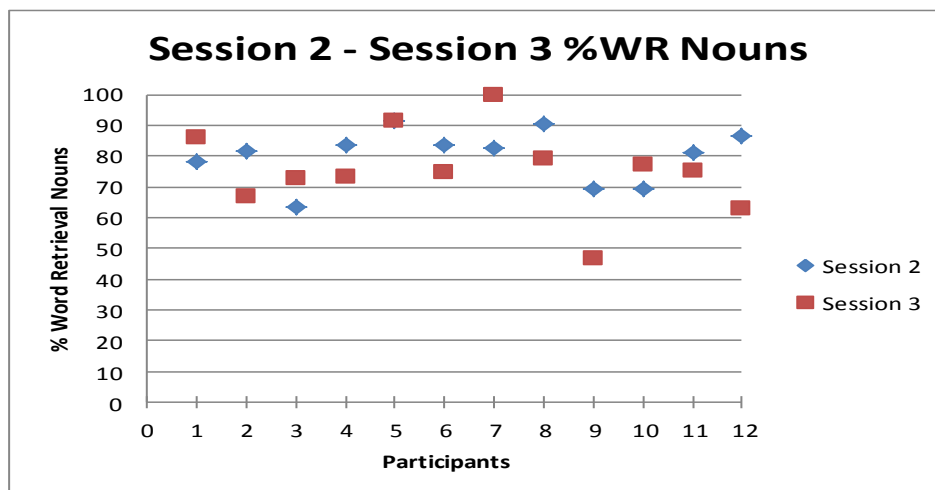
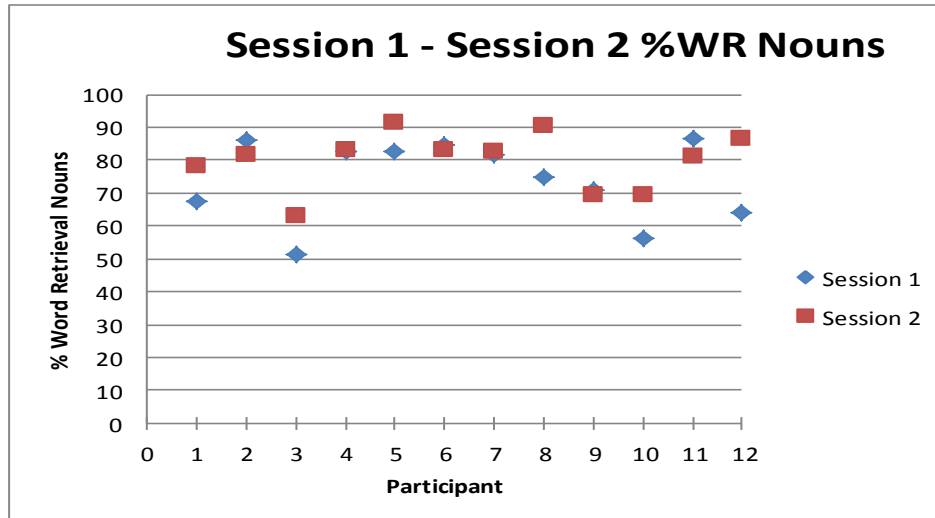


Figure 1. Percent Word Retrieval (%WR) scores for nouns for each participant illustrating the relationship of scores in Session 1 to Session 2 (top chart), Session 2 to Session 3 (middle chart), and Session 1 to Session 3 (bottom chart).

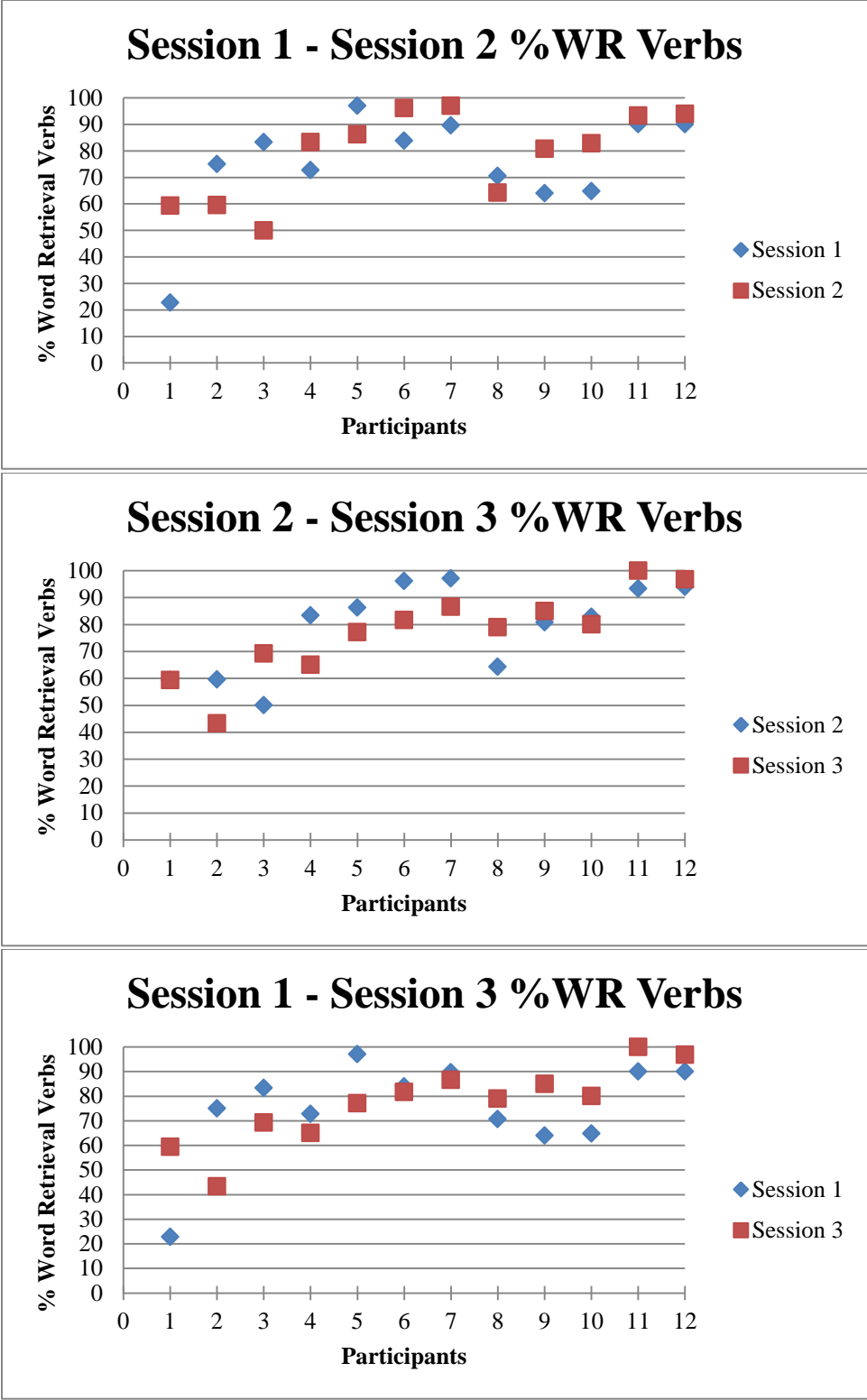


Figure 2. Percent Word Retrieval (%WR) scores for verbs for each participant illustrating the relationship of scores in Session 1 to Session 2 (top chart), Session 2 to Session 3 (middle chart), and Session 1 to Session 3 (bottom chart).

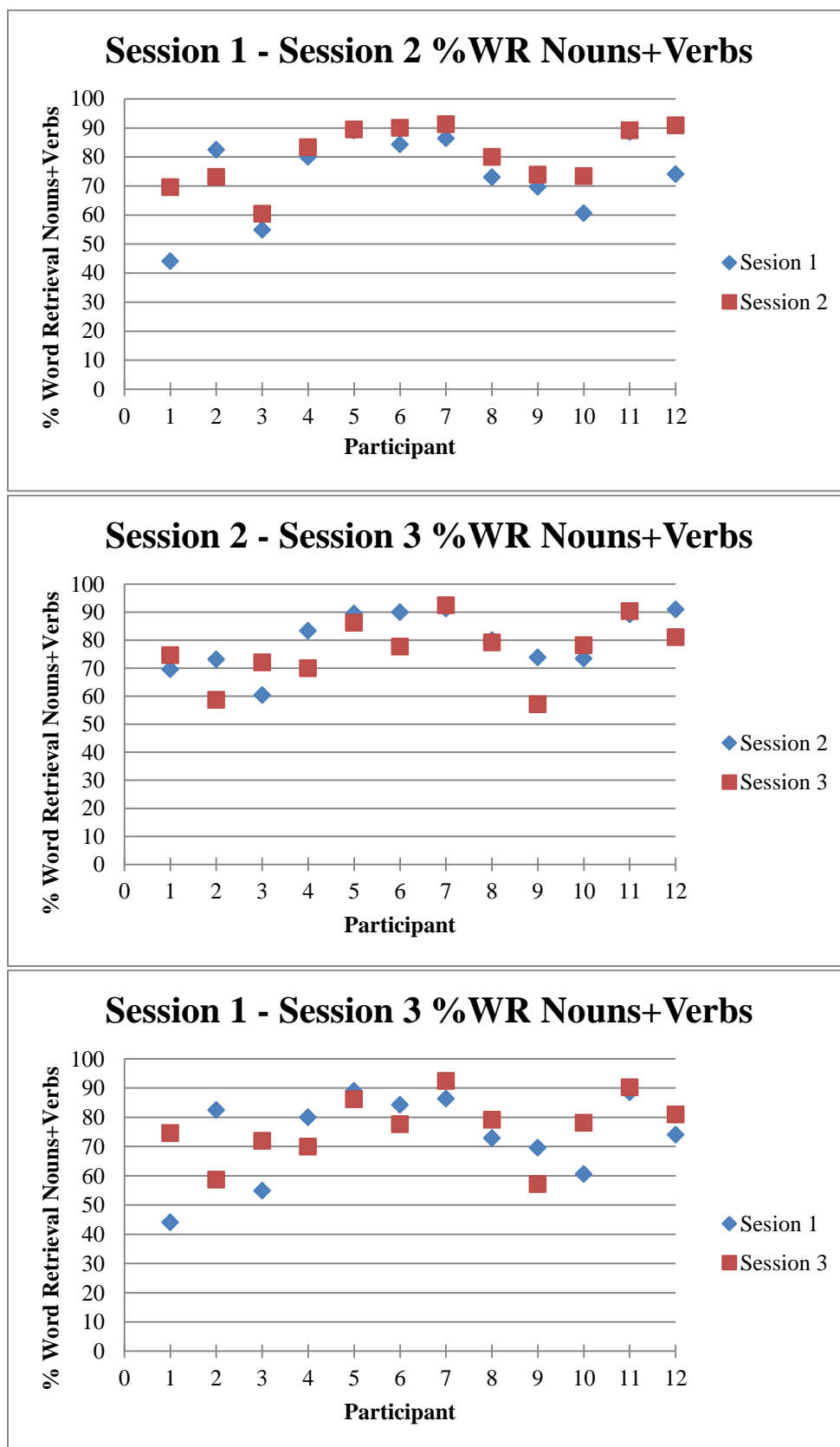


Figure 3. Percent Word Retrieval (% WR) scores for combined nouns and verbs for each participant illustrating the relationship of scores in Session 1 to Session 2 (top chart), Session 2 to Session 3 (middle chart), and Session 1 to Session 3 (bottom chart).

