APHASIOLOGY
AN INTERNATIONAL, INTERDISCIPLINARY JOURNAL

Ψ Psychology Press

Aphasiology

Comparing connected language elicitation procedures in persons with aphasia: Concurrent validation of the Story Retell Procedure

PLEASE SCROLL DOWN FOR ARTICLE

Ψ **Psychology Press**
Taylor & Francis Group

Comparing connected language elicitation procedures in persons with aphasia: Concurrent validation of the Story Retell Procedure

Malcolm R. McNeil, Jee Eun Sung, Dorothy Yang, Sheila R. Pratt, Tepanta R. D. Fossett and Patrick J. Doyle

VA Pittsburgh Healthcare System and University of Pittsburgh, PA, USA

Stacey Pavelko

University of Pittsburgh, PA, USA

Background: The Story Retell Procedure (SRP) (Doyle et al., 1998) is a well-described method for eliciting connected language samples in persons with aphasia (PWA). However, the stimuli and task demands of the SRP are fundamentally different from commonly employed picture description, narrative, and procedural description tasks reported in the aphasia literature. As such, the extent to which measures of linguistic performance derived from the SRP may be associated with those obtained from picture description, narrative, and procedural description tasks is unknown.
Aims: To assess the concurrent validity of linguistic performance measures obtained from the SRP with those obtained from picture description, narrative, and procedural description tasks by examining the correlations and the magnitude differences across the linguistic variables among the elicitation tasks. Secondarily, we examined the relationship of the percentage of information units per minute (%IU/Min) to other linguistic variables within the SRP and across the other elicitation tasks.
Methods and Procedures: This study compared the SRP to six different, frequently used sampling procedures (three sets of picture descriptions, one fairytale generation, one set of narratives, and one set of procedural description tasks) from which the same five verbal productivity, four information content, two grammatical, and two verbal disruption measures were computed. Language samples were elicited from 20 PWA, spanning the aphasia comprehension severity range. Tests of association and difference were calculated for each measure between the SRP and the other sampling methods.
Outcomes & Results: Significant and strong associations were obtained between the SRP and the other elicitation tasks for most linguistic measures. The SRP produced either no significant or significantly greater instances of the dependent variable except for the type–token ratio, which yielded a significantly lower value than the other sampling procedures.

Conclusions: The findings are interpreted as support for the concurrent validity of the SRP and as evidence that a single form of the SRP will yield a language sample that is generally equivalent in distribution to other sampling procedures, and one that is generally greater in quantity to those typically used to assess connected spoken language in PWA. Additionally, it was found that the %IU/Min metric predicted highly the information content linguistic measures on the SRP as well as on the other elicitation procedures. However, it did not predict well measures of verbal productivity, grammaticality, or verbal disruptions.

Among the elicited and observed procedures used to describe, classify, diagnose, measure change, quantify severity, and plan intervention for persons with aphasia (PWA), the measurement of connected spoken language has become a stable and valued procedure for many of these purposes. While being recognised as an important component in the assessment process, the most valid, reliable, and efficient methods for sampling connected spoken language have received relatively little experimental attention from clinical and experimental aphasiologists. Among the various methods used to elicit connected spoken language (e.g., picture description, story generation, personal or procedural narratives, video narration, etc.), the recently developed SRP derived from the stimuli used in the Discourse Comprehension Test (DCT) (Brookshire & Nicholas, 1997) has proven valid, reliable, and experimentally useful. Because the DCT was composed of 10 stories (plus 2 practice stories) that were equated on a number of important discourse-level linguistic variables—number of words, number of sentences, mean sentence length, number of subordinate clauses, number of T-units, ratio of clauses to T-units, listening difficulty, number of unfamiliar words, number of stated main ideas (propositions), number of implied main ideas (propositions), number of stated details, and the number of implied details—it has been used as a stimulus to elicit connected spoken language samples from non-impaired individuals and from PWA. In the SRP task, participants listen to each of the three predetermined stories constituting one of four story forms derived from the DCT stories. Participants are then instructed to retell each story in their own words, without picture support, immediately following its presentation. Participants' productions are subjected to various forms of linguistic description and quantification.

To date, research on the SRP has validated four equivalent forms (three stories each) based on linguistic variables (Doyle et al., 2000); investigated the value of picture-supported comprehension and retelling (Doyle et al., 1998); established the validity and reliability of the *information unit* scoring convention (McNeil, Doyle, Fossett, Park, & Goda, 2001); evaluated the scoring sensitivity of the percent information unit per minute efficiency measure (McNeil, Doyle, Park, Fossett, & Brodsky, 2002); investigated the inter-rater reliability of the SRP (Hula, McNeil, Doyle, Rubinsky, & Fossett, 2003); and investigated its auditory memory requirements (Brodsky et al., 2003). The SRP has the measurement advantage of having more specific predetermined targets for the retold stories than those in the other elicitation procedures, thus increasing the accuracy of the connected sample measurement. However, SRP performance reflects both comprehension and production processing and may be conditioned by the verbal memory demands of the task.

While the above-outlined psychometric properties of the test and of the scoring methods for the SRP have been investigated, concurrent validation of this procedure

with other established connected spoken language sampling procedures that do not rely on comprehension or memorial factors has not. This study sought to compare several aspects of the language generated from the SRP with other published procedures for eliciting spoken language that do not rely on comprehension and memory in PWA. These were composed of two frequently used picture descriptions from published aphasia tests—the BDAE *Cookie Theft* picture (Goodglass, Kaplan, & Barresi, 2001) and the WAB *Picnic* picture (Kertesz, 1982)—two story-like picture descriptions (*Cat in Tree* and *Birthday Party*), and two sequenced picture descriptions (*The Argument* and *Directions*). PWA described "what was happening" in each picture with the pictures presented throughout the description. Two procedural description tasks (*Tell me how you would go about doing dishes by hand* and *Tell me how you would go about writing and sending a letter*), two personal narratives (*Tell me what you do on Sundays* and *Tell me where you live*), and the Cinderella Story (Berndt, Wayland, Rochon, Saffran, & Schwartz, 2000) were compared with the SRP.

The specific aims of the current study are: (a) to establish the equivalency of the SRP and other elicitation tasks by analysing the correlations among the elicitation tasks for each linguistic measure and differences in magnitude for the linguistic measures among the tasks (see Appendix for linguistic measures); (b) to establish the degree of association between the percent information units per minute (%IU/Min) as a single overall SRP metric and the other linguistic measures; and (c) to establish the associations between the %IU/Min from the SRP and the other linguistic measures from the other elicitation tasks.

METHOD AND PROCEDURES

Selection criteria

A total of 20 pre-morbidly right-handed PWA met the following selection criteria: pure tone hearing screening at 35 dB HL in at least one ear at 0.5, 1, 2, and 3 KHz; 20/40 or better visual acuity (with correction if necessary) measured with the reduced Snellen chart; performance > 8.35 (greater than 1%ile) on the 55-item Revised Token Test (RTT) (McNeil & Prescott, 1978); performance that yielded a ratio (the delayed recall/immediate recall $\times$ 100) greater than .70 on the delayed retell compared to the initial retell on the Story Retelling Test of the Arizona Battery for Communication Disorders in Dementia (ABCD) (Bayles & Tomoeda, 1993); and performance at or above 7.83 (which equates to the 20th percentile on the 180-item PICA) for individuals with left hemisphere damage on the "Two-Item" Shortened Porch Index of Communicative Ability (SPICA) (DiSimoni, Keith, & Darley, 1980).

Participants

The participants were 8 male and 12 female native speakers of English with demonstrated language performance consistent with the McNeil and Pratt (2001) definition of aphasia as determined by their performance on the selection measures. The participants had a mean age of 63 years and ranged from 43 to 82 years old ($SD = 10$), a mean of 14 years of education with a range from 11 to 23 years ($SD = 3$), and were an average of 7 months post onset with a range from 3 to 312 months ($SD = 78$). The mean RTT overall percentile score was 56 and ranged from 2

to 94 ($SD = 26$ with two participants performing within the 1st decile; one at the 2nd and 3rd deciles; three at the 4th, 5th, 6th, and 7th deciles and two at the 8th, 9th, and 10th deciles). The mean SPICA overall percentile score (estimated from the norms from the full 180-item PICA; Porch, 1981) was 78 and ranged from 52 to 92 ($SD = 11$). The mean ABCD immediate and delayed story retell ratio score was 1.02 and ranged from .75 to 1.33 ($SD = 0.14$). Descriptive information for each participant is summarised in Table 1.

Sampling methods

Connected language samples were elicited from the PWA using the six experimental tasks described above and the SRP: one of the four forms of the SRP, the Cinderella Story (Berndt et al., 2000), and the five different elicitation procedures (with two samples of each) published by Nicholas and Brookshire (1993). SRP stories were presented without picture support during presentation and retell. Participants were instructed to listen to each of the three preselected stories that make up one form of the SRP and then retell each story in their own words, as completely as possible, immediately following its presentation. The Cinderella Story was presented from its published booklet containing pictures from the "Cinderella" fairytale. After the

TABLE 1
Participant biographical data and descriptive performance measures

Participant	Gender	Age	Education Level (Yrs.)	[a]MPO	[b]RTT Percentile	[c]Estimated PICA OA Percentile	[d]ABCD Ratio
1	F	52	18	74	68	80	1.00
2	F	66	11	63	3	52	1.33
3	F	75	12	61	22	58	1.00
4	F	63	12	73	45	81	1.13
5	F	49	14	30	50	69	.75
6	F	49	17	121	63	86	.82
7	M	82	13	312	40	85	1.11
8	M	43	14	66	51	85	1.00
9	M	55	16	96	82	88	1.00
10	F	72	14	24	50	89	1.00
11	M	61	14	30	77	79	.92
12	F	60	12	91	2	76	1.00
13	M	73	14	3	39	64	1.00
14	M	61	23	27	86	66	1.00
15	F	65	12	16	91	92	1.07
16	F	75	12	17	46	83	.86
17	M	65	12	191	72	81	1.00
18	F	72	16	20	63	88	1.33
19	M	64	12	192	94	91	1.00
20	F	61	14	8	67	72	1.00
Mean	(12F;8M)	63.15	14.10	75.75	55.55	78.25	1.01
SD		10.07	2.90	77.81	26.20	11.31	.14

[a]MPO = Months Post Onset; [b]RTT = *Revised Token Test* (McNeil & Prescott, 1978); [c]PICA = *Porch Index of Communicative Ability* (Porch, 1981); [d]ABCD Ratio = *Arizona Battery for Communication Disorders of Dementia* (Bayles & Tomoeda, 1993), determined by number of delayed recall items/number of immediate recall items × 100.

participant had finished looking at the pictures, and after the pictures were removed from view, they were instructed to tell the story of Cinderella in their own words. The Nicholas and Brookshire tasks are described above and involved picture descriptions and two narrative tasks. Thus, sampling tasks were conducted according to their respective published instructions, were administered in random order across participants, and were audio-recorded for subsequent orthographic transcription and analysis.

Dependent measures

Five measures of verbal productivity (*number of utterances*, *number of words*, *number of words per minute*, *mean length of utterance*, and *type–token ratio*), four measures of information content (*percent of story propositions*, *number of correct information units*, *percent of correct information units*, *number of correct information units per minute*; and *percent of information units per minute* for the SRP only), two measures of grammaticality (*number of conjunctions* and *percent of grammatically well-formed sentences*), and two measures of verbal disruptions (*percent of mazes* and *number of abandoned sentences*) were computed for each elicitation procedure. The %IU/Min for the SRP task was calculated because that efficiency measure has been found to be more sensitive for detecting pathology than information units alone (McNeil et al., 2002).

Three speech-language pathology students trained in both language transcription and SALT (Miller & Chapman, 1998) analysis transcribed, coded, and analysed the recorded samples. Of the samples, 10% (data from two randomly selected PWA) were re-transcribed by two of the three transcribers. From this, inter- and intra-transcriber agreement was calculated for each of the language-sampling tasks and each of the dependent measures. Both within-rater and between-rater scores exceeded 90% point-to-point agreement for all dependent measures on each of the elicitation procedures.

Each of the different sampling tasks (one form from the SRP, the Cinderella Story, and the two samples from each of the five Nicholas and Brookshire elicitation tasks) were analysed separately. An SRP form is composed of three individual stories, and data from these stories were combined to compose the SRP form data. Similarly, data from both samples of each of the five different Nicholas and Brookshire elicitation procedures were combined. That is, the data were combined from each of the two aphasia test pictures (the BDAE *Cookie Theft* picture and the WAB *Picnic* picture), two sequenced picture descriptions (*Argument* and *Directions*), two story-like picture descriptions (*Cat in Tree* and *Birthday Party*), two procedural descriptions (*Doing dishes by hand* and *Writing and Sending a Letter*), and two personal narratives (*What you do on Sundays* and *Where you live*), in order to compare across task demands and to increase the corpus size for each procedure to a level more comparable to the three-story SRP task.

Because Nicholas and Brookshire (1993) suggested that a stable and representative language sample might be derived by using a single task from each of the five elicitation procedures (aphasia test picture description, story-like and sequenced picture description, procedural descriptions, and personal narrative), the same dependent measures were calculated with the data combined into two sets (Sets A and B) of five stimuli each. Set A consisted of the BDAE *Cookie Theft*, *Birthday Party*, *Argument*, *What you do on Sundays*, and *Doing dishes by hand* tasks. Set B consisted of the WAB *Picnic*, *Cat in Tree*, *Directions*, *Where you live*, and *Writing*

and Sending a Letter tasks. These data, from sets A and B, and A + B combined, were then compared to the SRP, and the Cinderella Story.

Because the primary goal of the study was to establish the concurrent validity of the SRP, it was compared to each of the other sampling procedures across each of the linguistic measures. In order to determine the degree of association, Pearson Product Moment correlation coefficients were calculated for each dependent measure across each of the designated sampling procedures. In order to determine if the magnitude of any differences that exist among sampling procedures could be attributed to chance, a one-way repeated ANOVA, with Bonferroni adjustment for post-hoc comparisons, was likewise calculated. The predetermined alpha level of $p \leqslant .05$ was adjusted for multiple comparisons and set at $p \leqslant .01$ for all statistical comparisons, and a criterion of .70 was established as a substantive and statistically significant correlation between variables.

The linguistic variables of number of utterances, number of words, number of conjunctions, and number of abandoned utterances were summed for both samples (e.g. "what you do on Sundays" and "where you live") from each of the five Nicholas and Brookshire tasks, and for the three SRP stories and the single Cinderella Story. The remaining linguistic variables were averaged for each of the Nicholas and Brookshire elicitation tasks, the SRP, and for the single production of the Cinderella Story.

RESULTS

Correlations across linguistic variables between the SRP, Cinderella Story, and each of the five Nicholas and Brookshire tasks

Correlation coefficients between the SRP and each of the other elicitation procedures, for each of the 13 linguistic measures are summarised in Table 2. The SRP was correlated positively (with one exception) and significantly with the other procedures across the majority (80%; 61/77) of the 13 linguistic measures. Of these correlation coefficients, 53% (41/77) reached the criterion of .70, accounting for approximately 50% of the variance between the two variables. The great majority (85%: 40/47) of the correlations across the measures and elicitation procedures reached or exceeded the .70 criterion for the number of words per minute (6 of 6 correlations), mean length utterance (6/6), percent story propositions (3/5; story propositions were not calculated for the personal narratives), percent correct information units (5/6), number of correct information units per minute (5/6), percent grammatically well-formed sentences (5/6), percent mazes (5/6), and number of aborted sentences (5/6). Consistently lower correlation coefficients, with many non-significant, and few (7%; 2/30) reaching the .70 criterion, were derived across the elicitation procedures for the number of utterances (0/6), number of words (0/6), type–token-ratio (0/6), number of correct information units (1/6), and number of conjunctions (1/6).

While the great majority of the correlation coefficients between the SRP and the other elicitation procedures were significant ($p \leqslant .05$) across the 13 linguistic variables, the greatest number to reach the .70 criterion was found between the SRP productions and those of the Cinderella Story and the story-like pictures (69% each). The numbers of these correlations that reached criterion decreased from that of the aphasia test picture descriptions (54%), to the sequenced picture descriptions (46%), to the personal narratives (42%), and finally to the procedural descriptions (39%).

TABLE 2
Correlation coefficients (r) between the SRP and the other elicitation tasks across
all linguistic measures

Linguistic Measures	CIND	BDAE/WAB PICT.	SEQ. PICT.	NOVEL PICT.	PROCED	PERSONAL
Number of Utterances	.30	.56*	.19	.43	.25	.42
Number of Words	.47*	.46*	.38	.62*	.51*	.62*
Number of Words per Minute	**.76***	**.82***	**.83***	**.80***	**.87***	**.85***
Mean Length Utterance	**.70***	**.81***	.69*	**.80***	**.80***	**.86***
Type-Token Ratio	.35	.33	.31	.41	.23	.37
Percent of Story Propositions	**.71***	.69*	**.81***	**.82***	.64*	N/A
Number of Correct Information Unit	.62*	.58*	.59*	**.71***	.54*	.44
Percent of Correct Information Unit	**.86***	**.88***	**.75***	**.83***	**.75***	.41
Number of Correct Information Unit per Minute	**.81***	**.91***	.69*	**.83***	**.82***	**.75***
Number of Conjunctions	**.70***	−.11	.63*	.05	.52*	.45*
Percent of Grammatical Wellformedness	**.76***	**.89***	**.84***	**.86***	.68*	**.70***
Percent of Mazes	**.87***	**.93***	**.88***	**.89***	**.90***	.58*
Number of Abandoned Utterances	**.81***	**.78***	**.78***	**.70***	.47*	**.82***

Note. Correlation coefficients with an asterisk are statistically significant, $p \leq .05$. Shaded correlation coefficients in bold font meet or exceed the predetermined .70 criteria for a substantively high correlation coefficient.

Correlations across linguistic variables between SRP and aggregated Nicholas and Brookshire tasks

Values from each linguistic variable for the SRP and those same summed linguistic values from the five single tasks (one from each of the distinct procedures) derived from the Nicholas and Brookshire procedures, were also correlated. As described above, this yielded two forms of the procedure (Set A and B), as well as a combination of the two. These correlation coefficients are summarised in Table 3. In general, it can be seen that the same linguistic variables (number of utterances, number of words, and number of correct information units) that did not meet the .70 correlation criterion for the tasks separated by unique elicitation task demands (Figure 1) also did not meet this criterion when the tasks were combined. However, correlation coefficients for the type–token-ratio and number of conjunctions were substantively increased by this aggregation of tasks. The average correlation coefficients for those variables that did meet the criterion were .86 for the aphasia test picture descriptions, .81 for the sequenced picture descriptions, .80 for the story-like picture descriptions, .82 for the procedural descriptions, .79 for the personal narratives, 79 for set A, .81 for set B, and .80 for set A + B.

Correlations of the SRP %IU/Min metric with the SRP and the six other elicitation procedures across linguistic variables

It can be seen from Table 4 that the number of words per minute, percent story propositions, number correct information units, percent correct information units,

TABLE 3
Correlation coefficients (r) between the SRP and all linguistic measures across elicitation
Sets A, B, and A+B

Linguistic Measures	Set A	Set B	Set A+B
Number of Utterances	.30	.57*	.46*
Number of Words	.53*	.59*	.57*
Number of Words per Minute	**.73***	**.83***	**.81***
Mean Length Utterance	**.84***	**.75***	**.78***
Type -Token Ratio	**.75***	**.85***	**.87***
Percent of Story Propositions	**.87***	**.80***	**.78***
Number of Correct Information Unit	.63*	.69*	.67*
Percent of Correct Information Unit	**.83***	**.81***	**.83***
Number of Correct Information Unit per Minute	**.84***	**.92***	**.90***
Number of Conjunctions	**.76***	**.72***	**.77***
Percent of Grammatical Wellformedness	**.71***	**.74***	**.72***
Percent of Mazes	**.74***	**.84***	**.80***
Number of Abandoned Utterances	**.79***	**.79***	**.79***

Note. Correlation coefficients with an asterisk are statistically significant, $p \leq .05$. Shaded correlation coefficients in bold font meet or exceed the predetermined .70 criteria for a substantively high correlation coefficient.

and number correct information units per minute correlated positively, significantly, and at criterion levels with the SRP %IU/Min metric. None of the other correlation coefficients was statistically significant for the SRP. The correlation coefficients for the number of words per minute also reached criterion for the aphasia test picture
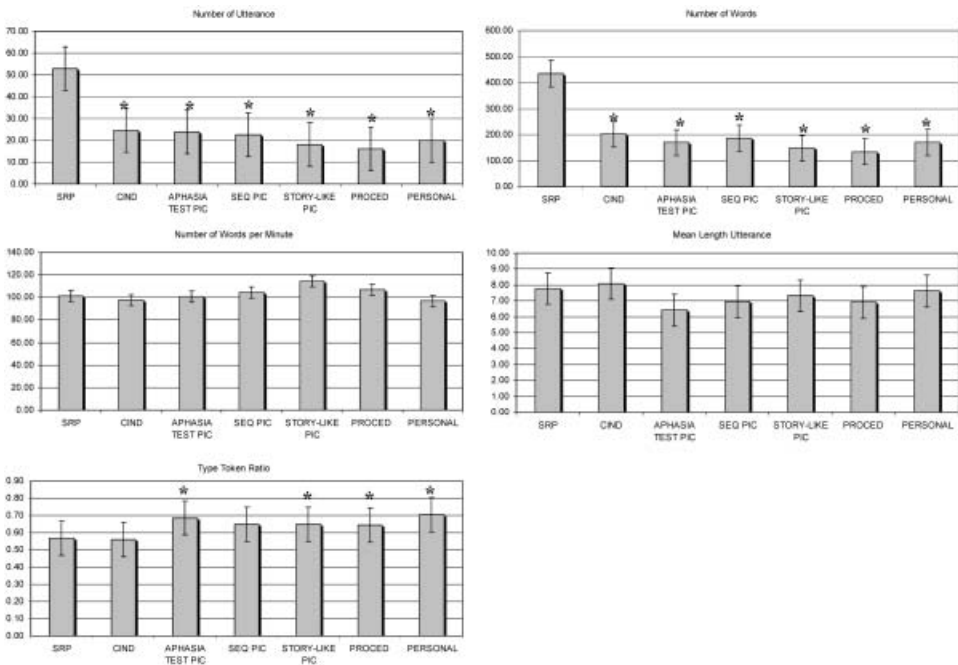

Figure 1. Verbal productivity measures across elicitation procedures. *Significant mean differences between the Story Retell Procedure (SRP) and the elicitation procedure.

TABLE 4

Correlation coefficients (r) between the SRP %IU's/Min and all linguistic measures across all elicitation tasks

Linguistic Measures	SRP	CIND	BDAE/WAB PICT.	SEQ. PICT.	NOVEL PICT.	PROCED	PERSONAL
Number of Utterances	.23	−.39	−.23	−.09	−.37	−.13	−.36
Number of Words	.15	−.15	.20	.21	−.06	.14	−.13
Number of Words per Minute	**.81***	.60*	**.72***	.61*	.69*	**.73***	**.73***
Mean Length Utterance	.52	.40	**.70***	67*	.65*	.60*	.46*
Type-Token Ratio	.15	.18	.33	.09	.14	.04	−.08
Percent of Story Propositions	**.72***	.43	.61*	.61*	.51*	.65*	N/A
Number of Correct Information Unit	**.73***	.39	.60*	.49*	.53*	.32	.22
Percent of Correct Information Unit	**.70***	**.70***	**.76***	.55*	**.71***	.56*	.50*
Number of Correct Information Unit per Minute	**.93***	**.76***	**.87***	.67*	**.78***	**.72***	.63*
Number of Conjunctions	.28	.06	.40	.42	.22	.49*	.04
Percent of Grammatical Wellformedness	.42	.42	.53*	.48*	.48*	.65*	.24
Percent of Mazes	−.39	−.54*	−.53*	−.48*	−.41	−.50*	−.49*
Number of Abandoned Utterances	−.43	−.55*	−.27	−.30	−.31	−.46*	−.41

Note. Correlation coefficients with an asterisk are statistically significant, $p \leq .05$. Shaded correlation coefficients in bold font meet or exceed the predetermined .70 criteria for a substantively high correlation coefficient.

descriptions, the procedural descriptions, and the personal narratives. Likewise, the mean length of utterance reached correlation criterion for the aphasia test picture descriptions. The percent correct information units and number of correct information units per minute also reached correlation criterion for the Cinderella Story, the aphasia test picture descriptions, and the story-like picture descriptions. The number of correct information units per minute also reached criterion for the procedural description tasks. Of the correlations from the grammatical category only one was significant for the number of conjunctions and none reached criterion, while several were significant but none reached criterion for the percent grammatically well-formed sentences. Half of the correlations were statistically significant, but none of them reached criterion for verbal disruptions.

The correlations between the SRP %IU/Min and Brookshire and Nicholas' sets A, B, and A + B for each linguistic variable, shown in Table 5, generally paralleled those where the procedures were subdivided. Only the measures of number of words per minute, percent correct information units, and number correct information units per minute reached correlation criterion. The pattern of correlation coefficients among elicitation tasks was not systematic.

Magnitude of differences

The five panels of bar graphs in Figure 1 summarise data from each of the verbal productivity linguistic measures for each sampling procedures. From left to right and top to bottom, the figures represent the total number of utterances, total number of words, total number words per minute, mean length of utterance, and type–token-ratio for each elicitation procedure. The asterisk above the frequency bars indicates that the value derived from the SRP was significantly different from the value for that elicitation procedure with the multiple comparison adjusted alpha ($p \leqslant .01$). It can be

TABLE 5
Correlations (r) between the SRP %IU's/Min and all linguistic measures across elicitation Set A, Set B, Set A+B

Linguistic Measures	Set A	Set B	Set A+B
Number of Utterances	−.45*	37	−.43
Number of Words	−.02	.14	.06
Number of Words per Minute	.65*	**.81***	**.77***
Mean Length Utterance	.61*	.60*	.49*
Type -Token Ratio	.24	.25	.20
Percent of Story Propositions	.67*	.62*	.53*
Number of Correct Information Unit	.38	.43	.41
Percent of Correct Information Unit	**.74***	.63*	.69*
Number of Correct Information Unit per Minute	**.81***	**.88***	**.86***
Number of Conjunctions	.24	.19	.19
Percent of Grammatical Wellformedness	.52*	.58*	.55*
Percent of Mazes	−.49*	−.49*	−.49*
Number of Abandoned Utterances	−.35	−.38	−.37

Note. Correlation coefficients with an asterisk are statistically significant, $p \leq .05$. Shaded correlation coefficients in bold font meet or exceed the predetermined .70 criteria for a substantively high correlation coefficient
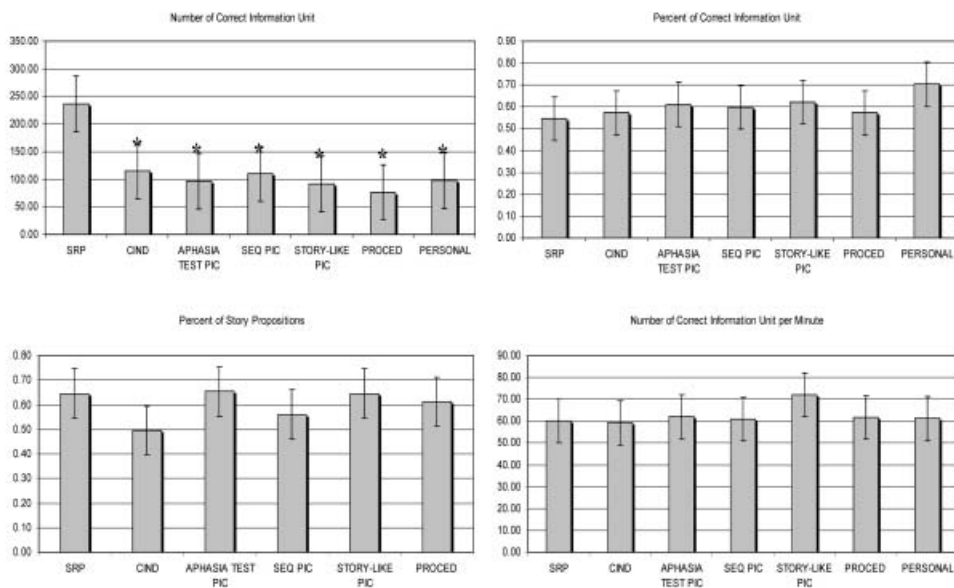
Figure 2. Information content measures across elicitation procedures. *Significant mean differences between the Story Retell Procedure (SRP) and the elicitation procedure.
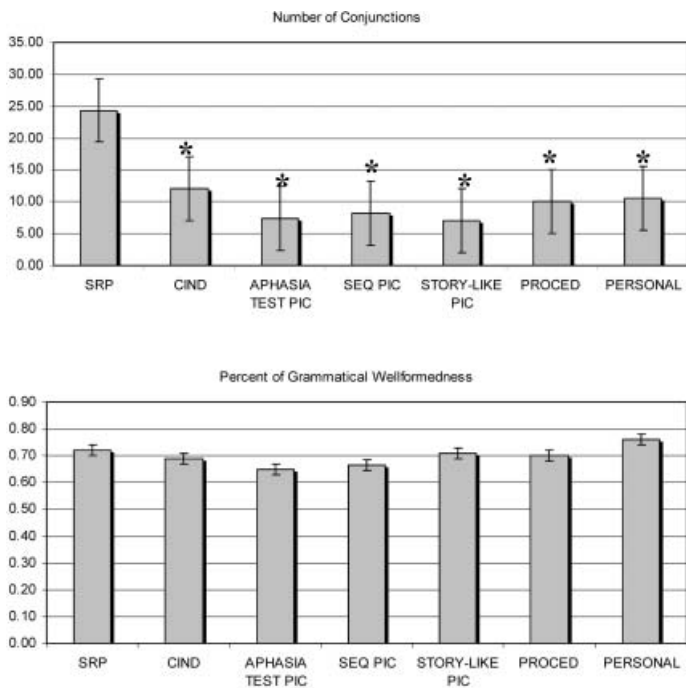

Figure 3. Grammaticality measures across elicitation procedures. *Significant mean differences between the Story Retell Procedure (SRP) and the elicitation procedure.
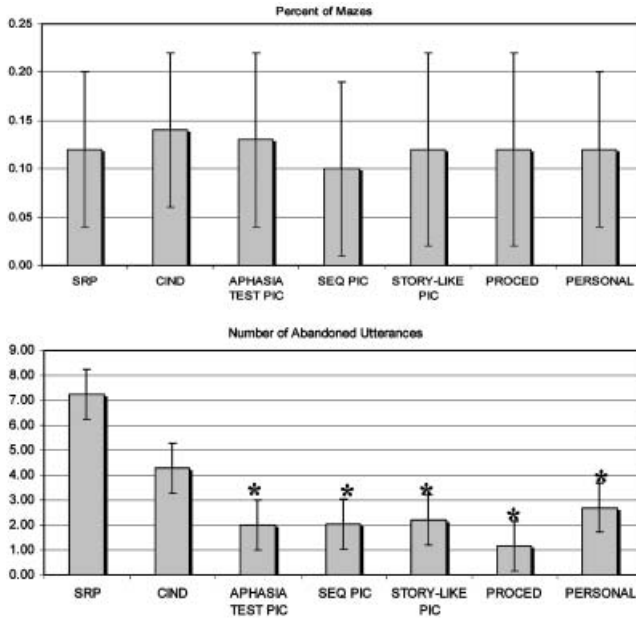
Figure 4. Verbal disruption measures across elicitation procedures. *Significant mean differences between the Story Retell Procedure (SRP) and the elicitation procedure.

seen that the SRP produced a significantly larger number of utterances ($p \leqslant .01$) and words ($p \leqslant .01$). No significant difference in magnitude of performance was evident for the number of words per minute ($p > .01$) or the mean length of utterance ($p > .01$). The SRP generated a significantly smaller type–token ratio than the other elicitation procedures ($p \leqslant .01$), except the Cinderella Story and the story-like picture descriptions, which were non-significantly different from the SRP ($p > .01$).

Figure 2 summarises the data for the information content measures composed of the total number of correct information units, percent correct information units, percent of story propositions, and number of correct information units per minute. As indicated by the asterisks, the SRP produced significantly ($p \leqslant .01$) more correct information units than each of the other elicitation procedures. No significant differences were found for any of the other information content measures for any of the elicitation procedures (all $p \geqslant .01$).

Figure 3 summarises the data for the two grammaticality measures. It can be seen in the first panel that the SRP generated significantly ($p \leqslant .01$) more conjunctions that created a compound sentence than each of the other elicitation procedures. No significant difference ($p \geqslant .01$) in the number of grammatically well-formed sentences was observed between the SRP and any of the other elicitation procedures.

Figure 4 summarises the data for the two measures selected to quantify verbal disruptions. While there was no significant difference between the percentage of mazes produced between the SRP and the other elicitation procedures ($p > .01$), the SRP did yield significantly more abandoned utterances than all other elicitation procedures (all $p \leqslant .01$) except the Cinderella Story, which was not significantly different from the SRP ($p > .01$).

DISCUSSION

A primary concern with the interpretation of the SRP and other retelling tasks as elicited, connected, spoken language sampling procedures is the fact that they require the individual to comprehend and remember the stimulus for later retelling. As such, the quality and/or the quantity of the sample could be influenced by what is comprehended and remembered, making it a procedure that is limited in its ability to provide an optimal or perhaps even valid estimate of connected spoken language. These potential limitations are minimised with connected spoken language elicitation procedures that require picture descriptions, fairytale generations, and personal narratives or procedural descriptions such as those with which the SRP was compared in this study. As such, the degree to which the SRP correlates with those elicitation procedures that do not tax the comprehension and memory systems would support its validity. This would be especially true if the magnitude of the behaviours were equivalent or greater for the SRP compared to the elicitation procedures whose comprehension and memory demands are minimised. The data derived from the SRP across the majority of the linguistic variables calculated in this investigation provide support for the concurrent validity of this retell procedure as one yielding language samples whose values are predicted from one elicitation procedure to the other. That the values derived from the SRP correlated positively and significantly across all elicitation procedures—Cinderella Story (85%), aphasia test picture descriptions (92%), sequenced picture descriptions (85%), story-like picture descriptions (77%), procedural descriptions (85%), and personal narratives (85%)—is interpreted as support for this contention. Support is also strengthened by the fact that the majority of the correlation coefficients derived across the measures reached or exceeded the .70 criterion for the Cinderella Story (69%), the story-like picture descriptions (69%), and the aphasia test picture descriptions (53%). These relationships were found for the majority (62%) of the linguistic measures and included those for the verbal productivity, information content, grammaticality, and verbal disruption domains of linguistic description generated from the SRP. To the degree that these relationships that cross elicitation procedures and linguistic measures provide a valid and reliable index of the connected spoken language of PWA, the SRP has established reasonable comparability and concurrent validity.

The average correlation coefficients for those variables that did meet the criterion were largely unchanged between the separated procedures (average correlation coefficients equalled .84 for the aphasia test picture descriptions, .78 for the sequenced picture descriptions, .82 for the story-like picture descriptions, .74 for the procedural descriptions, and .65 for the personal narratives) and those same procedures when they were aggregated (.79 for set A, .81 for set B, and .80 for sets A + B). No substantive changes in the average correlation coefficients that reached criterion were found by aggregating all ten tasks beyond those yielded by the five.

The linguistic variables that also correlated positively, significantly, and highly between the SRP and the other elicitation procedures when they were aggregated (sets A, B, and A + B) were generally the same variables that also correlated at similar levels when they were separated. The type–token ratio was the only linguistic variable in which the correlation coefficients increased substantively when the SRP was correlated with the aggregated elicitation tasks compared to when it was correlated with the individual procedures (with two samples each). This was likely

due to the sensitivity of the type–token ratio to sample size (Wachal & Spreen, 1973; Wright, Silverman & Newhoff, 2003).

The SRP generated a language sample that was greater or equal in quantity to the other elicitation procedures investigated in this study for the great majority of the linguistic variables computed. That is, the SRP yielded a significantly larger number of behaviours for the number of utterances. Obtaining a larger number of behaviours may present an advantage for the purposes of analysis due to additional measurement stability/reliability. A larger number of behaviours provides the examiner more opportunities to observe linguistic complexity (e.g., morphosyntactic and discourse) as well as more and different types of errors. Non-significant differences between the SRP and the other elicitation procedures were found for words per minute, mean length of utterances, percent correct information units, number of correct information units per minute, percent story propositions, percent of grammatically well-formed sentences, and percent mazes. Therefore, together with the high correlation coefficients across most of the independent and dependent variables, we interpret the non-significant differences across elicitation procedures and the significantly larger number of behaviours for all other variables as evidence that the SRP is equivalent or superior, in terms of more opportunity to observe both positive and negative relevant behaviours, compared to the other sampling procedures as they are typically employed, and using these linguistic variables. If sample sizes are equivalent or larger, as when elicitation tasks are aggregated such as those derived in Forms A and B of the Nicholas and Brookshire tasks, this superiority may not be apparent.

It is interesting to note that when there were no significant differences across the language-sampling procedures for a particular linguistic variable (e.g., percent correct information units), the behaviours elicited on the SRP correlated highly across those procedures, with the overall average correlations across procedures and measures achieving .84 for those that reached the .70 criterion. Conversely, when the SRP yielded significantly greater linguistic behaviours compared to the other elicitation tasks, the correlations tended to fall below the criterion.

Derived from a series of studies, the %IU/Min has been advocated as a single metric for the quantification of the SRP (Brodsky et al., 2003; Hula et al., 2003; McNeil et al., 2001, 2002). As a secondary question, correlation coefficients were calculated between the %IU/Min measure and the other linguistic measures for the SRP only in order to assess the degree to which it predicts these other behaviours. The results confirmed that the measures of information content are particularly well predicted by this single metric for the SRP. The interpretation of this is that the single %IU/Min metric does not predict other relevant linguistic behaviours in PWA. Thus, linguistic measures in addition to the %IU/Min appear to be necessary in order to accurately assess the range of linguistic behaviours typically evaluated in PWA.

In summary, the data from this study support the concurrent validity of the SRP as a spoken language elicitation procedure. The high correlation coefficients among language elicitation procedures across several linguistic variables is consistent with the interpretation that the comprehension and memorial demands of the SRP, compared to the other connected spoken language sampling procedures used in this study, do not restrict the quantity or the nature of language derived from it. In addition, the well-established psychometric properties of the SRP, along with its four equivalent forms and pre-specified production targets, appear to offer a sufficient

advantage to the accuracy and efficiency of transcribing and scoring the language sample to warrant its consideration over other language elicitation procedures. Finally, while the %IU/Min metric predicts information content measures on the SRP, it does not appear to be robust or general enough to predict measures of verbal productivity, grammaticality, or verbal disruptions.

REFERENCES

Bayles, K., & Tomoeda, C. (1993). *Arizona Battery for Communication Disorders of Dementia*. Tucson, AZ: Canyonlands Publishing.

Berndt, R. S., Wayland, S., Rochon, E., Saffran, E., & Schwartz, M. (2000). *Quantitative production analysis* (*QPA*). Philadelphia, PA: Psychology Press.

Brodsky, M. B., McNeil, M. R., Doyle, P. J., Fossett, T. R. D., Timm, N. H., & Park, G. H. (2003). Auditory serial position effects in story retelling for normal adult subjects and person with aphasia. *Journal of Speech-Language-Hearing Research*, 46(5), 1124–1137.

Brookshire, R. H., & Nicholas, L. E. (1997). *Discourse Comprehension Test*. Minneapolis, MN: BRK Publishers.

DiSimoni, F. G., Keith, R. L., & Darley, F. L. (1980). Prediction of PICA overall score by short versions of the test. *Journal of Speech and Hearing Research*, 23, 511–516.

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

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

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *Boston Diagnostic Aphasia Examination*. (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Hula, W. D., McNeil, M. R., Doyle, P. J., Rubinsky, H. J., & Fossett, T. R. D. (2003). The inter-rater reliability of the Story Retell Procedure. *Aphasiology*, 17, 523–528.

Kertesz, A. (1982). *Western Aphasia Battery*. New York: The Psychological Corporation, Harcourt Brace Jovanovich, Inc.

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

McNeil, M. R., Doyle, P. J., Park, G. H., Fossett, T. R. D., & Brodsky, M. B. (2002). Increasing the sensitivity of the Story Retell Procedure for the discrimination of normal elderly subjects from persons with aphasia. *Aphasiology*, 16, 815–822.

McNeil, M. R., & Pratt, S. R. (2001). A standard definition of aphasia: Toward a general theory of aphasia. *Aphasiology*, 15(10/11), 901–911.

McNeil, M. R., & Prescott, T. E. (1978). *The Revised Token Test*. Austin, TX: Pro-Ed.

Miller, J., & Chapman, R. (1998). *SALT: Systematic analysis of language transcripts*. [Windows versions 1.0–5.0]. Madison, WI: Language Analysis Laboratory, Waisman Center, University of Wisconsin.

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.

Porch, B. E. (1981). *Porch Index of Communicative Ability*. Palo Alto, CA: Consulting Psychologists Press.

Wachal, R. S., & Spreen, O. (1973). Some measures of lexical diversity in aphasic and normal language performance. *Language and Speech*, 16, 169–181.

Wright, H. H., Silverman, S. W., & Newhoff, M. (2003). Measures of lexical diversity in aphasia. *Aphasiology*, 17, 443–452.

APPENDIX: DEFINITIONS OF LINGUISTIC MEASURES

Verbal productivity

- Number of utterances: Total number of speaker attempts – included all utterances.
- Number of words: Total number of completed words (main body and mazes).
- Number of words per minute: Total completed words/elapsed time in minutes (main body and mazes).
- Mean length of utterance: Mean length of utterance in words—each word counts as one word for this calculation regardless of how many bound morphemes it may contain. Words found in mazes and omitted words are not included.
- Type–token ratio: Different words/total words.

Information content

- Percent of story propositions: Story propositions accurately and completely contain all "essential" information. Information is considered essential as long as the general meaning of the story is preserved.
- Number of Correct Information Units: Total number of words that are intelligible in context and accurately convey information relevant to the eliciting stimulus.
- Percent of Correct Information Units: The total number of Correct Information Units/the total number of words.
- Number per minute of Correct Information Units: The total number of Correct Information Units/elapsed time in minutes.
- Percent Information Units per minute: Percent of Correct Information Units/ elapsed time in minutes.

Grammatical well-formedness and syntactic complexity

- Number of conjunctions: Any of the following words (AFTER, AND, AS, BECAUSE, BUT, IF, OR, SINCE, SO, THEN, UNTIL, WHILE) is considered a conjunction if used in an utterance in some meaningful way (i.e., to connect two conjoining sentences). The preceding words, however, are not considered conjunctions if they serve no functional purpose in a sentence (i.e., in cases where the speaker uses them as a filler).
- Percent of grammatically well-formed sentences: The number of accurate and complete clauses (including independent clauses, dependent clauses, and prepositional phrases)/the total number of clauses and phrases.

Verbal disruptions

- Percent of mazes: Percent of total words that are in mazes (filled pauses, false starts, repetitions, and reformulations in an utterance).
- Number of abandoned sentences: Total number of sentences where the speaker voluntarily stops before completing it.