APHASIOLOGY, 2001, *15* (10/11), 991–1006

Reliability and concurrent validity of the information unit scoring metric for the story retelling procedure

Malcolm R. McNeil, Patrick J. Doyle, Tepanta R.D. Fossett, and Grace H. Park

VA Pittsburgh Health Care System and University of Pittsburgh, PA, USA

Amy J. Goda

VA Pittsburgh Health Care System, PA, USA

This study reports the reliability and concurrent validity of the information units (IU) metric as an efficient method for quantifying the amount of information comprehended and reproduced on the Story Retelling Procedure (SRP) (Doyle et al., 2000). Subjects were 31 normal adults and 15 adults with aphasia. Significant and moderately high correlation coefficients were obtained for subjects with aphasia between %IUs and most linguistic measures including the correct information unit (Nicholas & Brookshire 1993, 1995) while low and non-significant correlations were found for many measures of language productivity, efficiency, and disruption. The %IUs among the four SRP forms within group was non-significant ($p > .05$) and correlations were significant and high. Normal speakers produced significantly greater %IUs than aphasic speakers. Standard error of measurement was low across forms for both groups (3–4%) and the range of individual subjects' performance overlapped between 20 and 27% for the group with aphasia and between 36 and 55% for the normal group.

These results support the conclusion that %IU is a reliable and valid measure and differentiates aphasic from normal individuals better than normal individuals from persons with aphasia.

Despite the established importance of assessing the connected spoken language of persons with aphasia (Doyle, Goda, & Spencer, 1995; Doyle et al., 1998; Doyle et al., 1994; Nicholas & Brookshire, 1995; Ulatowska, Macaluso-Haynes, & North, 1980; Wambaugh, Thompson, Doyle, & Camarata, 1991), the most valid and reliable methods and procedures for doing so remain a matter of debate and experimentation. Valid sampling of connected language is necessary to provide an accurate prediction of language competence in relevant contexts other than those in which the sampling occurs (Doyle, Tsironas, Goda, & Kalinyak, 1996). Additionally, established reliable sampling procedures are necessary to have confidence that the abbreviated sample will accurately estimate the individual's performance on repeated trials and across different performance evaluators (testers).

Address correspondence to: Malcolm R. McNeil, PhD, Department of Communication Science and Disorders, University of Pittsburgh, 4033 Forbes Tower, Pittsburgh, PA 15260, USA. Email: mcneil@csd.upmc.edu

This study was supported by the Department of Veterans Affairs Rehabilitation Research and Development Service Project No B-2265R, awarded to Malcolm R. McNeil and Patrick J. Doyle

http://www.tandf.co.uk/journals/pp/02687038.html DOI:10.1080/02687040143000348

Several procedures have been developed to manage the inevitable trade-off between reliability and validity in connected spoken language sampling and scoring. Direct observation of language usage in the contexts of interest (e.g., during interpersonal discourse, listening to a news broadcast, or listening to a lecture and reformulating for transmission to another person or a class) has the greatest overall validity. However, the number of uncontrolled variables inherent in these sampling procedures and the time required to acquire them, make reliability and practicality difficult or impossible to achieve. For example, listener familiarity (Bottenberg & Lemme, 1991; Doyle et al., 1994; Li, Williams, & Volpe, 1995) and task demands (Doyle et al., Spencer, 1995; Shadden, Burnette, Eikenberry, & DiBrezzo, 1991) have been demonstrated to affect performance in discourse tasks. The speaker's familiarity with the topic, audience size, and emotional state can also affect performance. These and other similar variables are potential sources of variance and make assessments difficult to replicate from trial to trial and patient to patient in their varied contexts. In order to constrain these sources of variance and reduce the pragmatic difficulties associated with the measurement of connected language production, a number of elicitation procedures have been developed. These include: (a) conversational observation (Oelschlaeger & Thorne, 1999), (b) scripted interviews (Goodglass & Kaplan, 1983), (c) on-line video narration (McNeil, Small, Masterson, & Fossett, 1995), (d) off-line video scene description (Chafe, 1980), (e) fable generation/story telling (Berndt et al., 2000; Ulatowska, Chapman, Highley, & Prince, 1998), (f) picture description procedures (Nicholas & Brookshire, 1993, 1995; Yorkston & Beukelman, 1980), procedural description (Nicholas & Brookshire, 1993, 1995), and (g) story retellings (Doyle et al., 1998; Stout, Yorkston, & Pimentel, 2000). Each has advantages and disadvantages for specific populations and for specific clinical or experimental purposes. Doyle et al. (1998) and Doyle et al. (2000) used a story retelling method with concurrent picture presentation and have argued for the advantages of this method over several others. These advantages include the production of a known and constrained language sample from which a connected language analysis can quickly and reliably be completed. The use of picture-supported stimulus presentations also generally adds to the facilitation of comprehension, the reformulation of stimulus content, and a reduction in memory demands over those that involve off-line scene description. Doyle et al. (1998) and Doyle et al. (2000) have presented preliminary validity (in the form of sampling procedures) and reliability (in the form of equivalent story forms) data on a standardised SRP. These studies have demonstrated that picture-supported presentations and retellings of stories, derived from the *Discourse Comprehension Test* (Brookshire & Nicholas, 1997), when grouped into specific story combinations, yield four equivalent forms when indexed by information content, verbal productivity, verbal disruption, and grammaticality measures.

Because linguistic analyses of the SRP are time and effort intensive and beyond the clinical time available for most patient management, other valid and reliable but simplified scoring measures and procedures are needed. Responding to this need, Yorkston and Beukelman (1980) developed the *Content Unit* measure and it has been used frequently (Cherney, Drimmer, & Halper, 1997; Menn, Ramsberger, & Helm-Estabrooks, 1994). Kearns (1985) also developed the *Content Words* measure. Nicholas and Brookshire (1993) extended this type of measure with their conceptual and psychometric development of the *Correct Information Unit* (CIU) as a means of capturing the semantic essence of the communiqué. The virtue of the CIU metric is that

through application of extensive scoring rules, it can be applied to language samples about which the expected content is unknown. The necessity of using such a metric represents the typical situation for most language sampling procedures (e.g., interviews, natural discourse observation, picture descriptions, story and fable generation, etc). The liabilities of the CIU metric are that the necessary transcription of the language sample, along with the application of the CIU rules, is difficult to achieve reliably (Oelschlaeger & Thorne, 1999) and the time required for these procedures is generally clinically and experimentally prohibitive. In order to minimise these liabilities, a similar but simplified metric was developed and assessed on story retellings from which the target connected spoken language sample is known.

The purpose of this investigation was to assess the reliability and concurrent validity of the *Percent Information Unit* (%IU), a newly devised and simplified (from the earlier devised CIU metric from which it was spawned) scoring procedure for quantifying the informativeness of connected language in persons with aphasia. Answers to the following experimental questions were sought: (1) Is the %IU point-to-point inter-rater reliability above 0.90? (2) Is there a significant difference in the %IUs produced between normal subjects and subjects with aphasia? (3) Are the differences between SRP forms nonsignificant and the correlations among the forms high and positive? (4) Are there significant ($p < .05$) and high ($r > 0.70$) correlation coefficients between %IUs and other measures of information content? (5) Are there low and nonsignificant correlation coefficients between %IUs and measures of verbal productivity, verbal disruptions, and grammatical well-formedness? (6) Do the %IUs reliably discriminate between aphasic and normal control subjects using ± 2 standard deviations (SD) for cutoff scores and ± 1 standard error of measurement (SEM), from individual subject means? (7) Does the %IU correlate positively and highly with the *RTT* & *PICA* standardised aphasia tests?

Additionally, validity and reliability data for the SRP have been collected only from persons with aphasia. Reference data from individuals without impairment are needed in order to establish sensitivity and specificity data for deficit detection and eventually for differential diagnosis. In order to accomplish this, assessments of validity and reliability for the language sampling procedure and its scoring methods are required for the relevant comparison groups (e.g., normal population) as well.

METHOD

Participants

Subjects were 15 persons with aphasia, whose diagnosis was consistent with the definition of McNeil (1988), and 31 normal individuals. Each of the normal subjects was between 22 and 80 years of age ($M = 43$; $SD = 18$) and without self-reported history of neurological disease. All subjects achieved vision and hearing inclusion criteria and obtained no greater than 2 points difference between the Immediate and Delayed Story Recall Tasks from the *Arizona Battery of Communication Disorders of Dementia* (Bayles & Tomoeda, 1993). The subjects with aphasia ranged in age from 47 to 79 years ($M = 64$, $SD = 11$). They had an average aphasia severity as measured by the *Porch Index of Communicative Ability* (*PICA*) (Porch, 1981) and indexed by the overall percentile of 78 (range = 43–94, $SD = 15$) and an average overall *Revised Token Test* (RTT) (McNeil & Prescott, 1978) percentile of 57 (range = 4–96; $SD = 32$). Biographical and descriptive data for each subject with aphasia are presented in Table 1.

TABLE 1
Biographical and descriptive performance data for subjects with aphasia (N=15)

Subjects	*Age*	*MPO*	*RTT percentile*	*ABCD ratio*	*RCPM*	*PICA OA percentile*	*PICA VRB percentile*
1	62	11	73	85	34	92	78
2	77	44	19	118	24	59	63
3	47	11	4	100	24	65	54
4	51	77	53	133	29	87	60
5	79	13	77	233	20	75	77
6	56	84	95	100	32	87	89
7	74	71	96	86	27	94	97
8	55	30	63	100	32	75	71
9	66	33	80	100	27	89	76
10	57	85	58	125	27	86	75
11	64	252	14	91	24	93	68
12	71	94	4	100	22	43	37
13	52	17	92	100	36	87	91
14	73	23	66	117	21	76	70
15	74	11	54	100	18	63	54
M	63.87	57.07	56.53	113	26.47	78.07	70.67
SD	10.45	62.12	32.10	36	5.33	14.90	15.71

MPO = months post onset; RTT = *Revised Token Test* (McNeil & Prescott, 1978) percentile compared to adults with left-hemisphere damage; ABCD ratio = *Arizona Battery for Communication Disorders of Dementia* (Bayles & Tomoeda, 1993) ratio, determined by the number of delayed recall items divided by the number of immediate recall items and multiplied by 10; RCPM = *Raven's Coloured Progressive Matrices* (Raven, 1976) raw score out of a possible 36; PICA = *Porch Index of Communicative Ability* (Porch, 1981) percentile compared to adults with left-hemisphere damage: OA = overall percentile and VRB = verbal percentile.

Development of the IU measure

The IU is defined as an identified word, phrase, or acceptable alternative from the story stimulus that is intelligible and informative and that conveys accurate and relevant information about the story (see Appendix A for word and phrase inclusionary and exclusionary criteria). The stimulus-bound constraint allows for a more precise measure of information content, excluding non-story specific information, which can inflate measurement of information content. An example of the difference between the informativeness and specificity of the IU and CIU (Nicholas & Brookshire, 1993, 1995) metric is presented in Appendix B. In addition, pre-selected IUs permit more efficient identification and selection because they do not have to be orthographically transcribed, as do CIUs, and because the content, and all reasonable alternatives, are known a-priori and scoring can potentially be done on-line.

Procedure

Data collection. All language samples for the persons with aphasia were those collected in the Doyle et al. (2000) study. The samples collected and scores assigned from the 31 normal subjects in this study were collected under identical procedures to those of persons with aphasia.

Reliability. Prior to the identification and quantification of IUs, four research assistants were trained to correctly code IUs according to the established criteria and written definition. IU accuracy was defined by three of four judges' agreement. Two separate measures were obtained and point-to-point reliability was calculated as the percentage of agreements/agreements + disagreements for each rater for each measure. First, each of the four research assistants was evaluated on their ability to correctly identify all IUs in each of the 12 model stories using a point-to-point agreement measure. Second, inter-judge reliability was assessed for both normal and aphasic subjects' retellings. One of the 12 story-retellings from each subject was randomly selected, such that each of the 12 stories was represented an equivalent number of times. All four judges identified the IUs for each story from the model transcript (not from the subject's transcript).

Validation. Connected language samples were obtained from subjects' retellings of all four SRP forms under experimental conditions specified in Doyle et al. (1998). Retellings were orthographically transcribed and the number of IUs in each story retell was determined. As the number of possible IUs is known from the stimulus stories and because the number of possible IUs is not equivalent across stories and story forms, the number produced was divided by the number possible, to derive the %IUs. For purposes of establishing concurrent validity of the %IU measure, CIUs (Nicholas & Brookshire, 1993, 1995) and accurate and complete story propositions (Doyle et al., 1998) were counted for both the normal and aphasic groups using the scoring rules developed by the respective investigators. Measures of efficiency (CIU and %IU over time) were calculated for both groups. Additionally, the descriptive linguistic measures of verbal productivity, disruption, grammatical well-formedness, and sound production errors reported previously for the same aphasic subjects (Doyle et al., 2000) were used in this investigation and compared to the %IU measure. All correlation coefficients were computed using Pearson Product Moment correlations unless the normalcy assumption was not met, in which case the Spearman Rank Order correlation coefficient was computed.

In order to evaluate the IU metric's relationship to aphasic language processing, the IU metric was correlated with the *PICA* (Porch, 1981) and the *RTT* (McNeil & Prescott, 1978). Each individual subtest, grouped-modality subtests, and the overall score of the PICA were correlated with %IU from each of the four SRP forms. Only the overall RTT score was used for these correlations.

RESULTS

Scoring reliability

The percent agreement among the four research assistants' identification of IUs in the model stories averaged 99% (range = 96–100%). Percent agreement for inter-judge reliability averaged 96% (range = 81–100%) and 96% (range = 92–97%) for coding normal and aphasic subjects' retells, respectively.

SRP form reliability

Table 2 summarises the %IU means and standard deviations across SRP forms for both subject groups. Tests of normality and equal variance were passed for all forms ($p > .05$). A one-way repeated measures ANOVA comparing %IUs among the four forms was

TABLE 2
Means and standard deviations for %IUs for subjects with aphasia and normal control group for each SRP form

	%IUs				
Subjects	*Form A*	*Form B*	*Form C*	*Form D*	*Average*
Aphasia ($N = 15$)					
M	22.04	22.57	22.61	21.62	22.21
SD	11.12	10.05	11.67	13.81	11.66
Normal ($N = 31$)					
M	48.75	50.59	48.44	54.30	50.52
SD	7.49	8.83	7.23	7.95	7.88

nonsignificant ($df = 59$, $F = .041$, $p < .05$) for the aphasic group. However, a significant difference among forms ($df = 123$, $F = 14.01$, $p > .05$) was found for the normal group. Post-hoc testing revealed that the normal group produced significantly more %IUs on form D than on the other three forms, which were not significantly different from one another.

Table 3 summarises the correlation coefficients derived from %IUs calculated for the subjects with aphasia, among the SRP forms. All coefficients were significant and high and ranged from 0.91 to 0.96. The correlations among forms for the normal subjects were all significant, though lower than for the aphasic subjects. The correlations for form A with forms B, C, and D were 0.72, 0.75, and 0.77, respectively. The lowest correlation was between form B and form C at 0.66, while forms B and D correlated at 0.75. Forms C and D yielded the highest correlation at 0.84.

As summarised in Table 4, the Standard Error of Measurement (SEM), calculated across SRP forms, was low and ranged from 2.55 to 3.57%IUs for the aphasic group and from 3.61 to 4.76%IUs for the normal group.

TABLE 3
Pearson product moment correlation coefficients (r) and correlation coefficients squared (r^2) for %IUs for subjects with aphasia and normal control groups for each SRP form

	SRP form					
	B		*C*		*D*	
	r	r^2(%)	*r*	r^2(%)	*r*	r^2(%)
Subjects with aphasia ($N = 15$)						
A	0.93	87	0.94	88	0.91	83
B	—	—	0.94	88	0.93	87
C			—	—	0.96	92
Normal subjects ($N = 31$)						
A	0.72	52	0.75	56	0.77	59
B	—	—	0.66	44	0.75	56
C			—	—	0.84	71

TABLE 4
Standard errors of measurement (SEM) for the %IUs for subjects with aphasia and normal control group for each SRP form

	SEM (%IU)				
Subjects	*Form A*	*Form B*	*Form C*	*Form D*	*Average*
Aphasia ($N = 15$)	3.21	2.55	2.66	3.57	3.00
Normal ($N = 31$)	3.78	4.76	3.61	3.69	3.96

Concurrent validation

Information content/efficiency measures

Aphasic subject group. Obtained correlations between the %IU and the other information measures were high and significant ($p < .05$) with averaged correlation coefficients of .87 (range = 0.75–0.96) for Percent Correct Information Units (%CIUs) (Table 5). As shown in Table 6, %IUs and Percent Accurate and Complete Story Propositions (%ACSP) correlation coefficients averaged 0.91, and ranged from 0.79 to 0.96. Finally, the %IUs Per Minute and %CIUs Per Minute correlation coefficients ranged from 0.53 to 0.89 with an average of 0.73 (Table 7).

Normal subject group. Correlation coefficients were computed across the SRP forms between the %IUs and other measures of information content for the normal subject group. These were generally low to moderate and only about one-quarter to two-thirds were significant ($p < .05$) across SRP forms. The significant correlation coefficients averaged 0.41 (range = 0.39–0.43) for %CIU (Table 5). %IUs and %ACSP had an

TABLE 5
Spearman rank order correlation coefficients for %IUs with %CIUs for subjects with aphasia and normal subjects for each SRP form

	%CIU/Form			
%IU/Form	*A*	*B*	*C*	*D*
Subjects with aphasia ($N = 15$)				
A	0.86*	0.75*	0.84*	0.78*
B	0.96*	0.88*	0.91*	0.90*
C	0.87*	0.81*	0.93*	0.90*
D	0.88*	0.79*	0.89*	0.89*
Normal subjects ($N = 31$)				
A	0.43*	0.42*	0.30	0.39*
B	0.12	0.23	0.02	0.16
C	0.19	0.14	0.24	0.18
D	0.16	0.10	0.17	0.10

%IU = total IUs produced divided by total possible IUs. %CIU = total number of CIUs produced divided by total words produced in the story. Correlation coefficients that are statistically significant ($p < .05$) are indicated with an asterisk (*).

TABLE 6
Spearman rank order correlation coefficients for %IU with %accurate and complete story propositions (%ACSP) for subjects with aphasia and normal control subjects

	%ACSP/Form			
%IU/Form	*A*	*B*	*C*	*D*
Subjects with aphasia (*N* = 15)				
A	0.93*	0.79*	0.91*	0.92*
B	0.90*	0.89*	0.96*	0.94*
C	0.89*	0.79*	0.96*	0.91*
D	0.93*	0.87*	0.95*	0.94*
Normal subjects (*N* = 31)				
A	0.43*	0.32	0.27	0.42*
B	0.31	0.38*	0.11	0.33
C	0.26	0.13	0.24	0.38*
D	0.25	0.11	0.40*	0.48*

%ACSP = total number of accurate and complete story propositions divided by the total number of propositions. Pearson product moment correlation coefficients were computed for the subjects with aphasia; Spearman rank order correlation coefficients were computed for normal subjects. Correlation coefficients that are statistically significant ($p < .05$) are indicated with an asterisk (*).

TABLE 7
Spearman rank order correlation coefficients for %IUs/minute with %CIUs/minute for subjects with aphasia and the normal control group for each SRP form

	%CIU/minute/form			
%IU/minute/form	*A*	*B*	*C*	*D*
Aphasic subjects (*N* = 15)				
A	0.68	0.60	0.59	0.53
B	0.87	0.81	0.71	0.69
C	0.78	0.68	0.89	0.85
D	0.77	0.61	0.85	0.82
Normal subjects (*N* = 31)				
A	0.74	0.66	0.81	0.78
B	0.63	0.69	0.71	0.69
C	0.67	0.58	0.82	0.69
D	0.66	0.60	0.74	0.76

All correlation coefficients are significant ($p < .05$). %IU/minute = %IUs divided by total time of retell. %CIU/minute = %CIUs divided by total time of retell.

average correlation coefficient of 0.41 (range = 0.38–0.48) (Table 6). The %IUs Per Minute correlated with the %CIUs Per Minute yielded an average coefficient of 0.70 (range = 0.58–0.82, Table 7).

Verbal productivity and disruption, grammatical well-formedness, and sound production error measures

Correlation coefficients were computed between the %IUs and measures of verbal productivity and verbal disruptions for each SRP form for the subjects with aphasia only. The normal subjects were not evaluated on these variables because the frequency of occurrence of errors on these measures was too low for analysis. These measures were total number of words, number of words per minute, number of utterances, mean length of utterance in words, number of utterances with mazes, total duration of silent pauses in seconds, grammatical well-formedness, and sound production errors (see Doyle et al. 2000, for operational definitions). Performance on these measures was predicted to correlate more poorly with the %IU metric than the information content measures.

None of the correlation coefficients was significantly different from zero ($p > .05$) for Type-Token-Ratio, Total Number of Words, Mean Length of Utterance, Number of Pauses Per Minute (≥ 2 seconds per pause), or for the number of Mazes Per Minute. For the Words Per Minute measure, two of the computed correlation coefficients were significant ($p < .05$) and both were moderately low (0.56). Statistical significance was achieved by half of the correlation coefficients between %IUs and mean length of pause (for forms C and D only), and between %IUs and sound production errors. In addition, all correlations between %IUs and sound production errors were negative. For both comparisons, the significant correlation coefficients were of modest magnitude with an average of 0.59 (range = 0.55–0.62) for mean length of pause, and -0.57 (range = -0.54–-0.62) for sound production errors. Unexpectedly, all correlation coefficients between %IUs and grammatical well-formedness were significant, with an average of 0.60 (range = 0.53–0.74).

Normal/aphasic IU comparisons

A one-way repeated measure ANOVA comparing %IUs between groups yielded a significantly greater %IUs for the normal group ($df = 183$, $F = 83.92$, $p \leq .001$).

Standardised measures

%IU/PICA and RTT comparisons

Correlation coefficients were computed between the %IUs and performances on the *PICA* and *RTT* for the subjects with aphasia only, due to the narrow range of performance by the normal subjects on these measures. Spearman rank order correlation coefficients between %IUs for all four SRP forms and the combined PICA verbal subtests (I, IV, IX, XII) were significant ($p < .05$) and high (average = 0.80, range = 0.76–0.91). See Table 8 for a summary of these and other correlational data.

Table 8 also summarises the %IU and overall *PICA* and overall *RTT* correlations. The overall *PICA* score correlated with %IUs significantly ($r = .59$) only for SRP form B. High (ranging from 0.80 to 0.85) and significant correlation coefficients were obtained between %IUs and Overall *RTT* scores for all SRP forms.

TABLE 8
Spearman rank order correlation coefficients for %IUs for each SRP form correlated with PICA subtest scores combined by various language functions, overall (O.A.) PICA, and overall RTT scores from subjects with aphasia

	%IUs/SRP form			
PICA subtests combinations	*A*	*B*	*C*	*D*
I, IV, IX, XII	0.76*	0.91*	0.76*	0.77*
II and III	0.28	0.29	0.24	0.02
V and VII	0.46	0.45	0.34	0.24
VI and X	0.45	0.53*	0.42	0.39
VIII and XI	0.46	0.53	0.42	0.39
A, B, C, and D	0.38	0.36	0.23	0.48
A, B, C, D, and E	0.34	0.33	0.21	0.41
O.A. PICA	0.36	0.59*	0.49	0.46
O.A. RTT	0.82*	0.82*	0.80*	0.85*

An asterisk (*) indicates all correlation coefficients that are significant ($p < .05$). *Verbal subtests:* I, IV, IX, XII; *Gestural subtests:* II and III; *Reading subtests:* V, VII; *Auditory subtests:* VI and X; VIII and XI, *Writing/copying subtests:* A, B, C, D, E, F.

Subject group discrimination

As shown in Table 9, the number of subjects whose average performance plus or minus one SEM, fell within two standard deviations of the group's mean was calculated across each form and for the average across forms. On the average across forms, slightly less than one-quarter of the aphasic subjects fell within two standard deviations of the normal group's performance. On the average across forms, slightly less than half of the normal subjects fell within the two SD range of aphasic scores.

DISCUSSION

The purpose of this study was to assess the reliability and validity of a measure developed to quantify the informativeness of connected language in persons with aphasia. The reliability and validity of the procedure used to elicit the language samples to which the

TABLE 9
Intergroup comparisons

	SRP form				
Subject group	*A*	*B*	*C*	*D*	*Average*
Aphasia					
M %IU +2 standard deviations	45.50	42.74	45.94	49.24	45.86
Number (percent) misclassified as normal	3 (20%)	4 (27%)	4 (27%)	3 (20%)	14 (23%)
Normal					
M %IU −2 standard deviations	33.77	32.93	33.98	38.40	34.77
Number (percent) misclassified as aphasic	17 (55%)	11 (36%)	17 (55%)	12 (39%)	14.25 (46%)

Cutoff scores for intergroup comparisons using the average %IUs +2 standard deviations for the group with aphasia and the average %IUs −2 standard deviations for the normal group and the number and percent of subjects misclassified for each group using individual subjects' %IU + 1SEM for the subjects with aphasia and − 1 SEM for the normal subjects, for each SRP form.

%IU measure was applied, has been previously established (Doyle et al., 2000; Doyle et al., 1998) and has some similarities with other connected language sampling procedures (Berndt et al., 2000; Ulatowska et al., 1980) used with individuals with aphasia. The story-retell procedure allows language formulation and production operations to occur within a constrained context and provides a standardised and efficient sampling procedure. Several variables suggest that the %IU metric and scoring procedure are more efficient than other measures (e.g., CIUs) for quantifying information content. For example, IUs and their alternatives have been pre-determined, eliminating the need for a great deal of decision making when scoring. As IUs and alternatives are identified on score sheet checklists and it is unnecessary to transcribe story reproductions, most scoring can be done on-line, once the examiner is familiar with the stories and checklists. Additionally, the story model provides a known stimulus context, which may facilitate interpretation of ambiguous patient/subject productions.

Support for the %IU metric and the scoring procedure's high reliability and validity were found in this study. High inter-rater reliability, averaging above 95% point-to-point agreement, was achieved for IU identification in the original stimulus stories and in both normal and aphasic language samples.

One question asked whether there was a significant difference between the percentage of IUs produced by the normal subjects and the subjects with aphasia. The significant difference between the two groups, without an interaction across SRP forms confirms the positive answer to this question. The %IUs produced by the aphasic subjects was less than half of that produced by the normal subjects and evidenced more variability. Question number 6 queried the discriminative accuracy of the SRP/%IU measure using plus (for the aphasic group) and minus (for the normal group) two SDs and SEMs added or subtracted from individual subject scores as cut-off measures. Using this analysis, overall, 23% of the subjects with aphasia were misclassified as normal, and the misclassifications ranged from 20 to 27% across the SRP forms. Normal subjects were misclassified as aphasic 46% of the time overall across forms, with a range of 36 to 55%. Although the number of subjects sampled in this investigation is too small to generate high confidence in the IU's power of aphasia versus normal subject differentiation, these data do provide motivation to pursue this line of research.

It is reasonable to assume that multiple exposures to a particular story might result in learning. It was therefore desirable to develop several equivalent forms of the sampling procedure for test–retest purposes. Doyle et al. (2000) have previously demonstrated linguistic equivalence of each of the four forms used in the SRP. However, this equivalence had not been demonstrated for the %IU metric. Across-form reliability was demonstrated with small and equivalent standard deviations and low and relatively equivalent SEM for both subject groups. There were no differences in %IU magnitude across the forms for the subjects with aphasia, and the differences that did exist across forms for the normal subjects were small and nonsignificant except for form D. The equivalent performance for both subject groups across the four forms for the subjects with aphasia and the three forms for the normal subjects provides further support for the substitution of equivalent form reliability for traditional test–retest reliability. The reliability of %IU performance to multiple exposures of the same stories and retell procedures could be evaluated. However, the equivalence of the multiple forms, within the limits of the SEM, provide the opportunity for at least four consecutive administrations of the SRP/%IU to the same subject/patient with confidence that if change occurs in performance, it can be attributed to a change in the patient and not the instability of the sampling or scoring procedures.

Question number 3 asked whether the %IU measure was concurrently validated by the acquisition of high correlation coefficients between the %IU and other language information and efficiency measures (e.g., %CIUs, %ACSP, %IUs/Minute with %CIUs/Minute) for persons with aphasia and for normal subjects. Correlations between %IU and %CIU yielded high correlation coefficients, which averaged 0.87 and ranged from 0.75 to 0.96 across the forms. These results, as well as correlation coefficients ranging from 0.79 to 0.96 for %IUs correlated with accurate and complete story propositions (average 0.91), validate the %IU measure for individuals with aphasia. Likewise, the relatively high correlation coefficients for the subjects with aphasia (ranging from 0.53 to 0.89 and averaging 0.74) and for the normal subjects (ranging from 0.58 to 0.82 and averaging 0.70) on the information efficiency measures (percent information units per minute correlated with the percent correct information units per minute) add confidence that the IU is measuring or predicts the same underlying mechanisms affecting the performance on more traditional and well standardised language measures.

The few significant and low to moderate magnitude of the correlation coefficients obtained with the normal subjects on the language information measures (differentially from the language efficiency measures) does not provide concurrent validation for normal subjects. The reason for this disparity between correlation coefficients in the normal and aphasic subjects and between the information measures and these same measures computed as a function of the time taken to produce them is not readily apparent. A reduced range of scores for the normal subjects likely contributed to the reduced correlation coefficients found in these analyses. Nonetheless, because the normal subjects' %IU performance does not predict other more traditional language content variables as well as it does for the persons with aphasia, it is most prudent to recommend caution in the use of the SRP/%IU procedure and measure as a method of normal language sampling and measurement until additional experimentation can be completed.

Question number 5 was motivated by the prediction that the correlation coefficients between %IU and measures of verbal productivity (e.g., total number of words, type–token–ratio, etc.), verbal disruptions (e.g., number of utterances with mazes, total duration of silent pauses, etc.) or grammaticality (e.g., grammatical well-formedness) would be low and nonsignificant. This prediction was based in the belief that the information unit is more accurately considered a measure of lexical/semantic access than of processes related to phonological, syntactic, or verbal disruptions. Because these pathological behaviours did not occur with sufficient frequency to be analysed in the normal group they were analysed only for the group with aphasia. With very few exceptions and unpredictably across SRP forms, these correlations conformed to our predictions. The generally low and non-significant correlations between the %IU and these non-information-specific linguistic measures suggest that if a complete assay of language performance is desired, measures in addition to the %IU will have to be used to describe and quantify them. These low correlation coefficients also add concurrent validity to the %IU measure as a metric that is psycholinguistically loaded on the lexical/semantic (and perhaps discourse) level of language processing more than on the phonological, morphosyntactic, and phonological/phonetic/motoric speech production levels of aphasic performance.

The significant and positive correlations averaging 0.60 between %IU and grammatical well-formedness challenge the assumption that performance on this syntactic measure by persons with aphasia is independent of lexical-semantic access deficits. Indeed this may be a reasonable challenge given that most current syntactic production and comprehension accounts readily accommodate lexical/semantic deficits

as one mechanism for the generation of agrammatic comprehension or agrammatic production deficits.

In order to determine whether the SRP/%IU quantification of connected spoken language predicts overall or components of more traditional aphasia measures such as the PICA and the RTT, correlation coefficients were computed between these %IUs and these two measures for the group of subjects with aphasia. These correlations were also computed in order to get a first-pass look at the task requirements on the SRP indexed using the %IU. That is, if correlations were positive and high between %IU and RTT overall scores and not between the spoken language production measures of the PICA (i.e., subtests I, IV, IX, XII), it might be speculated that the SRP task is more heavily reliant on deficits of comprehension of the story to be retold than on deficits or limitations of the language formulation and production system. It is apparent from the results of these comparisons that *both* language comprehension, as indexed by the overall RTT, and language production, as indexed by the combined spoken language production subtests of the PICA, underpin the SRP procedure and the %IU measure. As such, it is tempting to speculate that this new measure may provide a valid overall measure of aphasia severity. However, the poor and generally nonsignificant correlations across the SRP Forms, between the %IU and the well-validated and reliable PICA overall score, gives pause to this temptation. Additional research comparing the SRP/%IU with a much larger sample of persons with aphasia is necessary before succumbing to the alternative temptation of accepting this null hypothesis.

The precise cognitive/linguistic demands imposed on the comprehension and production by the SRP are not fully explicated by the analyses presented here. Likewise, the precise cognitive/linguistic limitations revealed by the persons with aphasia in this study on the SRP are also not fully explicated by these analyses. Nonetheless, some speculation about both problems is possible. One likely candidate shared by both problems is the working and intermediate-term memory requirements of the story retell procedure. Each story in the SRP spans about 2 minutes of connected speech that has to be comprehended, recalled, formulated, and produced. There are apparent working memory, intermediate-term and perhaps ''long-term'' memory requirements in this task. As only about 50% of the stated IUs from the stories were retold by the normal subjects, it is reasonable to assume that some or perhaps most of the 50% reduction can be accounted for by the memory requirements of the task, as these subjects would be expected to comprehend smaller units of connected language near perfectly and to produce connected language samples with near perfect semantic, syntactic, phonologic, and morphological integrity. Additionally, it will be recalled that there was a higher correlation coefficient between the %IU and the RTT than between the %IU and subtests VI and X of the PICA. One possible explanation for this perhaps unexpected finding may be found in the different working memory requirements of the two aphasia auditory comprehension tests. The RTT requires the storage and manipulation of linguistic information that varies in the number of critical linguistic elements, placing varying demands on working memory capacity and activation duration. Contrarily, the comprehension of subtests VI and X of the PICA places relatively fewer demands on working memory buffer size or activation duration. If we assume that there are working and intermediate-term memory demands of the SRP, and if we assume that memory limitations are a source of performance differences on the SRP/%IU measure in persons with aphasia, it is reasonable to speculate that performance would correlate more highly with the RTT than with the PICA. What is not clear is whether the SRP and the %IU metric would either correlate more highly with the RTT than other connected language

sampling procedures or with other metrics such as the CIU. Additional studies are in progress that will explore the memory demands of the SRP/%IU.

Regardless of the memory demands of the SRP or the memory deficits of persons with aphasia revealed by the SRP, it is important to note that the memory requirements of the SRP task and %IU scoring procedure do not pose a threat to the validity of the tool as a measure of informativeness of language production. There is general agreement that CIUs represent a valid measure of the informativeness of a spoken language sample. Because IUs occur less frequently than CIUs, it is tempting to conclude that the IU metric is not a good measure of informativeness. However, it must be remembered that the %CIUs scored in a story is calculated relative to the number of words produced in the story that are not informative or relevant to the story. The %IUs in a story retell is, however, calculated relative to the target words in the story (those told to the subject), not the proportion of the IUs that were correct, informative, and relevant to the story compared to those that were produced that were not correct, informative, and relevant to the story. It should also be remembered that the percent IUs and CIUs (as well as %IUs and %CIUs) correlated highly and positively in the context of the SRP in this investigation. Both IUs and CIUs are considered to be ''informative'' and the %IU metric cannot be equated to a general measure of informativeness as defined by %CIU.

In summary, the %IU metric used in the context of the specific SRP employed in this investigation can be scored reliably and administered at least four times reliably (using the four equivalent SRP forms). The procedures that foster its reliability of scoring also make it a relatively short and efficient method for acquiring a great deal of language performance data about persons with aphasia. The %IU's positive and high correlation coefficients with other measures of information content of connected spoken language, and its low correlation coefficients with measures more remotely related to information content, support its concurrent validity. Additionally, the SRP/%IU's high and positive correlation with the combined spoken verbal subtests of the PICA (and low and nonsignificant correlations with other language tasks on the PICA) and with the overall RTT score, support its validity as a measure relevant to aphasia as well as a measure reliant on both comprehension and production of connected spoken language. The %IUs produced (relative to those in the stimulus story) may be limited by the working or short-term memory factors in the normal population and by these and other language comprehension and production deficits in the persons with aphasia. Finally, the ability of the SRP/%IU measure to separate only about one-half of the normal subjects in this study from persons with aphasia and about 25% of the persons with aphasia from the normal subjects, suggests that the tool lacks the power to discriminate and is not likely to contribute to the detection of aphasia or its separation from normal performance if used alone.

REFERENCES

Bayles, K.A., & Tomoeda, M.S. (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: Psychology Press.

Bottenberg, D., & Lemme, M.L. (1991). Effect of shared and unshared listener knowledge on narratives of normal and aphasic adults. *Clinical Aphasiology, 19*, 109–116.

Brookshire, R.H., & Nicholas, L.E. (1997). *Discourse comprehension test.* Minneapolis, Minnesota: BRK Publishers.

Chafe, W.L. (1980). The development of consciousness in the production of a narrative. In R.O. Freedle (Series Ed.) & W.L. Chafe (Vol. Ed.), *Volume III in the series Advances in discourse processes. The Pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (pp. 9–50). Norwood, NJ: Ablex Publishing Corp.

Cherney, L.R., Drimmer, D.P., & Halper, A.S. (1997). Informational content and unilateral neglect: A longitudinal investigation of five subjects with right hemisphere damage. *Aphasiology*, *11*, 351–363.

Doyle, P.J., Goda, A.J., & Spencer, K.A. (1995). The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. *American Journal of Speech-Language Pathology*, *4*, 130–134.

Doyle, P.J., McNeil, M.R., Park, G., Goda, A., Spencer, K., Lustig, A., & Carroll, B. (2000). Linguistic validation of four parallel forms of a story retelling procedure. *Aphasiology*, *14*, 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.

Doyle, P.J., Thompson, C.K., Oleyar, K.S., Wambaugh, J.L., & Jackson, A.V. (1994). The effects of setting variables on conversational discourse in normal and aphasic adults. *Clinical Aphasiology*, *22*, 135–144.

Doyle, P.J., Tsironas, D., Goda, A.J., & Kalinyak, M. (1996). The relationship between objective measures and listeners' judgments of the communicative informativeness of the connected discourse of adults with aphasia. *American Journal of Speech-Language Pathology*, *5*, 53–60.

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders*. Philadelphia: Lea & Febiger.

Kearns, K.P. (1985). Response elaboration training for patient initiated utterances. *Clinical Aphasiology*, *15*, 196–204.

Li, E.C., Williams, S.E., & Della Volpe, A. (1995). The effects of topic and listener familiarity on discourse variables in procedural and narrative discourse tasks. *Journal of Communication Disorders*, *28*, 39–55.

McNeil, M.R. (1988). Aphasia in the adult. In N.J. Lass, L.V. McReynolds, J. Northern, & D.E. Yoder (Eds.), *Handbook of speech-language pathology and audiology* (pp. 738–786). Toronto: B.C. Decker Inc.

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

McNeil, M.R., Small, S.L., Masterson, R.J., & Fossett, T.R.D. (1995). Behavioral and pharmacological treatment of lexical-semantic deficits in a single patient with primary progressive aphasia. *American Journal of Speech-Language Pathology*, *4*, 76–87.

Menn, L., Ramsberger, G., & Helm-Estabrooks, N. (1994). A linguistic communication measure for aphasic narratives. *Aphasiology*, *8*, 343–359.

Miller, J.F., & Chapman, R.S. (1986). *Systematic analysis of language transcripts (SALT)*. University of Wisconsin-Madison, USA.

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.

Nicholas, L.E., & Brookshire, R.H. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech and Hearing Research*, *38*, 145–156.

Oelschlaeger, M.L., & Thorne, J.C. (1999). Application of the correct information unit analysis to the naturally occurring conversation of a person with aphasia. *Journal of Speech Language and Hearing Research*, *42*, 636–648.

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

Raven, J.C. (1976). *Coloured Progressive Matrices*. Oxford: Oxford Psychologists Press Ltd.

Shadden, B.B., Burnette, R.B., Eikenberry, B.R., & DiBrezzo, R. (1991). All discourse tasks are not created equal. *Clinical Aphasiology*, *20*, 327–342.

Stout, C.E., Yorkston, K.M., & Pimentel, J.I. (2000). Discourse production following mild, moderate, and severe traumatic brain injury: A comparison of two tasks. *Journal of Medical Speech-Language Pathology*, *8*, 15–25.

Ulatowska, H.K., Chapman, S.B., Highley, A.P., & Prince, J. (1998). Discourse in healthy old-elderly adults: A longitudinal study. *Aphasiology*, *12*, 619–633.

Ulatowska, H.K., Macaluso-Haynes, S., & North, A.J. (1980). Production of narrative and procedural discourse in aphasia. *Clinical Aphasiology*, *9*, 17–27.

Wambaugh, J.L., Thompson, C.K., Doyle, P.J., & Camarata, S. (1991). Conversational discourse of aphasic and normal adults: An analysis of communicative functions. *Clinical Aphasiology*, *20*, 343–353.

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

APPENDIX A

Information Units definition with inclusionary and exclusionary criteria.

An Information Unit (IU) is an identified word or phrase from the story stimulus that is intelligible and informative and that conveys accurate and relevant information about the story. All IUs must have been stated in the story. Liberal use of synonyms is used for the target IU. Correct grammatical word form is not necessary for IU identification.

Inclusions: IUs include nouns, pronouns (including ''what''), proper nouns, titles (e.g. Mr), verb phrases, adjectives, adverbs (including ''wh'' words, ''yes'' and ''no''), and prepositions. Hyphenated words, idioms, and colloquialisms are counted as single IUs. Contractions that include a copula are counted as two IUs (e.g. ''that's mine'').

Exclusions: Articles, conjunctions and auxiliary verbs are not counted (including contracted forms) as IUs.

APPENDIX B

CIU and IU comparison example

Note: All CIUs are in **bold print** and IUs are underlined.

Example 1.

Uh . . . **George took his** regular **place along the third** . . . bu . . . uh . . . bla . . . **baseline**. And . . . uh . . . **he donned an old cap**. And . . . uh . . . uh . . . uh . . . when . . . uh . . . **when** . . . **he flagged balls . . . then . . . the . . . batter up** . . . **sent a line drive to George**. And **he** . . . **fell over** . . . o . . . **the side onto the playing field**. And . . . **there's a** . . . **a fellow in a top hat** . . . **said gee I'm with a circus**. . . . **Will you apply for the job**.

Total Number of Words in Retell: 64
Total Number of **CIUs:** 57
% CIUs = #CIUs/Total # of Words in Retell
% CIUs = 89% = 57/64

Total Possible IUs in Story 1:111
Total Number of IUs: 24
%IUs = #IUs/Total Possible IUs
%IUs = 22% = 24/111

Example 2.

Jim Jim and **his wife they** were **sitting there** just sitting there **late at night with the snow**. And . . . um . . . they just thought they would sit there. And . . . h . . . **she was going to** feel . . . um . . . um . . . **sew her** pants. And . . . um . . . **pretty soon** . . . aft . . . **he decided he wanted to eat a sandwich**. Se . . . and . . . um . . . while he was . . . he wanted to eat **he also said he wanted some cookies . . . So then finally he ate a little bit** he ate the . . . ha . . . th . . . the salad I man **the sandwich**. And **then he was going to** . . . um . . . e . . . he was going to **get some** some **cookies too**. And **then later he decided he wanted some water too**. **Well** here **his** . . . wa . . . **wife** said . . . tha . . . she said she'll she'll get it for for . . . hims . . . for herself. Buy . . . yi . . . she **she told him that you better start getting it yourself**. And . . . um . . . that was about it.

Total Number of Words in Retell: 137
Total Number of **CIUs:** 76
% CIUs = #CIUs/Total # of Words in Retell
% CIUs = 55% = 76/137

Total Possible IUs in Story 5: 160
Total Number of IUs: 20
%IUs = #IUs/Total Possible IUs
%IUs = 13% = 20/160