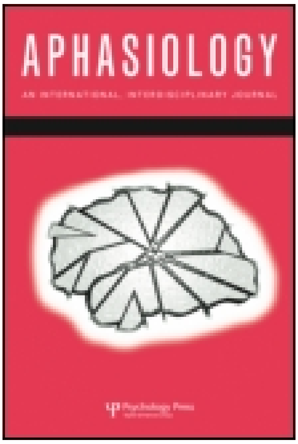

Aphasiology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/paph20

Towards a quantitative measure of verbal effectiveness and efficiency in the Amsterdam-Nijmegen Everyday Language Test (ANELT)

Marina B. Ruiter [a], Herman H. J. Kolk [b], Toni C. M. Rietveld [c], Nienke Dijkstra [c] & Erica Lotgering [c]

[a] Sint Maartenskliniek Research Development & Education, Nijmegen, the Netherlands

[b] Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University Nijmegen, Nijmegen, the Netherlands

[c] Department of Linguistics, Radboud University Nijmegen, Nijmegen, the Netherlands
Version of record first published: 19 May 2011.

APHASIOLOGY, 2011, 25 (8), 961–975

Psychology Press
Taylor & Francis Group

Towards a quantitative measure of verbal effectiveness and efficiency in the Amsterdam-Nijmegen Everyday Language Test (ANELT)

Marina B. Ruiter[1], Herman H. J. Kolk[2], Toni C. M. Rietveld[3], Nienke Dijkstra[3], and Erica Lotgering[3]

[1]Sint Maartenskliniek Research Development & Education, Nijmegen, the Netherlands
[2]Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University Nijmegen, Nijmegen, the Netherlands
[3]Department of Linguistics, Radboud University Nijmegen, Nijmegen, the Netherlands

Background: A well-known test for measuring verbal adequacy (i.e., verbal effectiveness) in mildly impaired aphasic speakers is the Amsterdam-Nijmegen Everyday Language Test (ANELT; Blomert, Koster, & Kean, 1995). Aphasia therapy practitioners score verbal adequacy *qualitatively* when they administer the ANELT to their aphasic clients in clinical practice.
Aims: The current study investigated whether the construct validity of the ANELT could be further improved by substituting the qualitative score by a *quantitative* one, which takes the number of essential information units into account. The new quantitative measure could have the following advantages: the ability to derive a quantitative score of verbal efficiency, as well as improved sensitivity to detect changes in functional communication over time.
Methods & Procedures: The current study systematically compared a new quantitative measure of verbal effectiveness with the current ANELT Comprehensibility scale, which is based on qualitative judgements. A total of 30 speakers of Dutch participated: 20 non-aphasic speakers and 10 aphasic patients with predominantly expressive disturbances.
Outcomes & Results: Although our findings need to be replicated in a larger group of aphasic speakers, the main results suggest that the new quantitative measure of verbal effectiveness is more sensitive to detect change in verbal effectiveness over time. What is more, it can be used to derive a measure of verbal efficiency.
Conclusions: The fact that both verbal effectiveness and verbal efficiency can be reliably as well as validly measured in the ANELT is of relevance to clinicians. It allows

Address correspondence to: Marina Ruiter, Sint Maartenskliniek Research, Development & Education, P.O. Box 9011, 6500 GM Nijmegen, The Netherlands. E-mail: m.ruiter@maartenskliniek.nl

The study described in this article was supported by the rehabilitation centre of the Sint Maartenskliniek in Nijmegen and conducted in collaboration with the Donders Centre for Cognition as well as the Department of Linguistics of the Radboud University Nijmegen. We would like to thank Marjolein Bruijstens, Lea Maessen, and Berber Spliethoff for their help in collecting the data. We would also like to thank the two anonymous reviewers for their valuable and constructive comments on an earlier version of this paper.

http://www.psypress.com/aphasiology DOI: 10.1080/02687038.2011.569892

them to obtain a more complete picture of aphasic speakers' functional communication skills.

Keywords: Functional communication; Quantitative measure; Verbal effectiveness; Verbal efficiency; ANELT.

A substantial body of research claims that the ultimate goal of aphasia therapy is to improve the functional communication skills of people with aphasia (cf. Blomert, 1990; Ramsberger, 2005; Worrall, 1995). Although functional communication can be defined in various manners, it is typically referred to as "the ability to receive a message or to convey a message, regardless of the mode, to communicate effectively and independently in a given environment" (American Speech-Language-Hearing Association, 1990, cited in Frattali, 1992, p. 64). In this article the focus will be on the category of aphasic patients whose language comprehension is relatively preserved, but whose spoken language production is mildly to moderately impaired (e.g., Broca's and anomic aphasia). In line with the classic classification of Weisenburg and McBride (1935), we will use the label *aphasic patients with predominantly expressive disturbances* to denote this subgroup of persons with aphasia. Focusing on this subgroup, the definition of functional communication can be more narrowly defined as the skill to get the message across effectively and independently through *spoken language*. Speakers with predominantly expressive disturbances not only differ in the extent to which they are able to convey their message through speech, but also in the number of words or time needed to do so. In order to obtain a more complete picture of their functional communication skills, *verbal efficiency* (i.e., verbal production per time unit) should be investigated as well.

THE MOST IMPORTANT OUTCOME MEASURE IN APHASIA THERAPY: FUNCTIONAL COMMUNICATION

In the rehabilitation of aphasic patients with predominantly expressive disturbances, there are two ways to reach the goal of improved functional communication: restoration and compensation. In order to demonstrate an effect of restoration therapy, linguistic aspects of spoken language production as well as functional communication skills need to be examined, as both aspects of recovery do not always seem to correlate (e.g., Irwin, Wertz, & Avent, 2002; Marini, Caltagirone, & Carlomagno, 2007; Ross & Wertz, 1999).

The assessment of functional communication becomes even more important if restoration therapy is not profitable. In that case, aphasic speakers can only improve functional communication skills by adapting their speech behaviour. Compensation, which results from adaptation, allows them to circumvent their permanent linguistic impairment, at least for the greater part. For example, chronically anomic patients may handle their word-finding difficulties independently by producing a circumlocution each time word retrieval fails (e.g., Tompkins, Scharp, & Marshall, 2006). The effectiveness of therapy aiming for independently accessed and initiated compensatory speech behaviour cannot be evaluated with tests that measure predominantly language function, such as the Dutch version of the Aachen Aphasia Test (AAT; Graetz, De Bleser, & Willmes, 1992) or the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA: Kay, Lesser & Coltheart, 1995), because improvement in language function is no longer expected. Instead, verbal effectiveness and efficiency should be investigated, independent of the linguistic form of the utterances used.

AMSTERDAM-NIJMEGEN EVERYDAY LANGUAGE TEST

A well-known test for measuring verbal adequacy (i.e., verbal effectiveness) in mildly impaired aphasic speakers is the Amsterdam-Nijmegen Everyday Language Test (ANELT; Blomert, Koster, & Kean, 1995). The ANELT does not include measures of communicative efficiency; focus is on communicative effectiveness only. In this test aphasic speakers are presented with scenarios of everyday life situations. Each scenario has a script-like character, in which the experimenter is an involved listener who does not engage in conversation. As a consequence, the communicative setting of each scenario is realistic, but consists of a monologue basically. An example of a scenario is: "Suppose you have an appointment with your family doctor; however, something else has come up. Therefore you need to reschedule your appointment. When you call the doctor, what do you say to him?" The verbal responses of the aphasic speaker to the ANELT scenarios are audio-taped and scored afterwards. The test has a standardised procedure, norm samples, and there are two parallel versions available. Each parallel version consists of 10 scenarios.

In the ANELT verbal adequacy is defined as a function of the comprehensibility of the spoken message and the intelligibility of the utterances used to convey it. The former relates to the content of the message, independent of the well-formedness of the utterances used. The latter relates to the degree to which the utterance can be perceived clearly, independent of its content or meaning. As Blomert, Kean, Koster, and Schokker (1994) point out, not only paralinguistic features (e.g., tempo and loudness) and voice quality have an effect on a speaker's intelligibility, but also disturbances that occur at the central levels of language production, such as phonemic paraphasias and neologisms. Thus two scales are used to score the aphasic speaker's verbal responses: the Comprehensibility A-Scale and the Intelligibility B-Scale. Each scale has five points ranging from *very bad* to *very good.* In effectiveness and efficacy studies the A-scale is often used exclusively as an outcome measure of verbal adequacy (e.g., Bastiaanse, Hurkmans, & Links, 2006; Doesborgh et al., 2003; Laska, Hellblom, Murray, Kahan, & Von Arbin, 2001). For that reason, we discuss only the ANELT A-scale here.

In order to establish inter-rater reliability of the ANELT A-scale, Blomert et al. (1994) requested six relatively untrained raters to score the responses that 14 aphasic speakers had given to both versions I and II. The judges were familiar with the ANELT scoring procedure, but they were not expert raters. Inter-rater reliability was satisfactory: Krippendorff's alpha, a reliability statistic used for content analysis (Krippendorff, 1970), was .92 and .94 for version I and version II respectively. Since internal consistency and stability over time were also found to be satisfactory, the ANELT is claimed to provide a reliable and valid assessment of aphasic speakers' verbal adequacy (Blomert, 1994; Blomert et al., 1994).

Blomert (1990) raised the question of how much and which information an aphasic speaker needs to express in each scenario in order to be judged *communicatively adequate* in the ANELT. In order to answer this question, Blomert first of all established which elements of meaning comprised the verbal responses to the ANELT of 30 non-linguistically impaired speakers of Dutch. Blomert does not specify the elements of meaning for each scenario, neither does he provide examples, but the elements of meaning probably consisted of single content words and phrases. For the scenario on the family doctor introduced above, the elements of meaning might have included the following ones: *appointment*, *unable to be there*, *new appointment.* Although the

ANELT consists of two parallel versions, Blomert only analysed the 10 scenarios of version I. Two independent raters were then asked to indicate whether 30 Dutch-speaking aphasic speakers produced these elements as well. The aphasic participants were between 20 and 71 years of age (*Mdn* = 52), 47% female. Median time post-onset was 14 months (Range: 2–70). Based on their AAT scores, the participants were classified as Broca (7), Wernicke (5), Global (6), and Anomic (6). Out of the 30 participants, 6 could not be classified. In line with the expectations the degree to which aphasic speakers were comprehensible—as rated with the A-scale—was directly related to the number of essential elements that they had produced in ANELT version I, $r = .77$. This finding not only lends support for the construct validity of the ANELT, but also illustrates that "adequacy of verbal communication can be reliably expressed in quantitative measures" (Blomert, 1990, p. 317).

Although Blomert derived a quantitative score of aphasic speakers' comprehensibility in his 1990 study, in current practice aphasia therapy practitioners score verbal adequacy *qualitatively* when they administer the ANELT to their aphasic patients. In this article we use the notion of "qualitative judgement" (and similar notions) to indicate that the current scoring procedure is concerned with subjective judgement of verbal effectiveness rather than with measurement of this parameter based on counts of events which can be objectively identified, because no external criteria (i.e., the amount of essential information to be conveyed) are provided for assigning scores on the (ordinal) 5-point Comprehensibility scale. That is, clinicians do not transcribe their patients' verbal responses orthographically when assigning scores to the A-scale; they only score the audio fragments, which they should not play back more than once according to the ANELT protocol. Clinicians should keep the following questions in mind when rating the aphasic speaker's verbal adequacy on the 5-point A-scale: (a) Which information was expressed? (b) To what extent does this information cover the message to be conveyed? So, even though reference is made to the amount and relevance of information conveyed, it is not indicated in the test manual *which* elements of meaning are considered essential in order to achieve the communicative goal. Also, it is not clear *how many* of these elements need to be expressed in order to be given a particular score. Aphasia therapy practitioners are only provided with some examples in the Appendix, such as the following response to Scenario 5 of version I (family doctor, introduced above), which should be assigned 2 points on the ANELT A-scale:

> *dan zou je ze zeggen niet niet te an gaat niet 't gaat niet dan het ik efgesreken gesproken ja . . . ja sa sorry eh kan niet .. ken nie weg eh ..*
> (Then you they should say not not to al not possible it is not possible then I has egree agreed well .. well sa sorry er cannot . . . cannot go away er . . .)

As stated above, the data obtained by Blomert (1990) indicate that adequacy of spoken language can be reliably expressed in quantitative measures, and his data thus lend support to the construct validity of the ANELT, based on qualitative judgements. This raises the question of whether the construct validity could not be further improved by substituting the qualitative score by a quantitative one, which takes the number of essential information units into account. One obvious improvement would be that now, in addition to a measure of verbal effectiveness, we are able to derive a verbal efficiency score, such as the average number of essential information units produced per minute. But the new quantitative measure could have other advantages as well, such as improved sensitivity to detect changes in functional communication over time. Although Grande et al. (2008) did not include any measures of verbal effectiveness in their analyses, they found that quantitative analyses of spontaneous speech of persons with aphasia were more sensitive to change in linguistic performance than speech rating scales. It therefore seems worthwhile to systematically compare the current qualitative scoring procedure with a quantitative measure of verbal adequacy, which is based on the number of essential information units produced in the ANELT.

RESEARCH AIMS

The present study systematically compares a quantitative measure of verbal effectiveness, based on the essential elements of meaning to be expressed in the ANELT, with the current ANELT A-scale, which is based on qualitative judgements. The reason for comparing both measures is that it allows us to investigate whether:

1. quantification of the number of essential information units produced in the ANELT yields at least the same inter-rater agreement as the current Comprehensibility A-scale;
2. a quantitative score of verbal effectiveness allows experimenters to reliably tell aphasic patients apart from linguistically unimpaired speakers, just as the traditional ANELT does;
3. quantifying the number of essential information units produced in the ANELT improves responsiveness of the traditional ANELT. A test is responsive if it is sensitive to detect changes in verbal effectiveness when guided recovery takes place. There are two conditions that have to be met in order to investigate responsiveness: the aphasic speakers should improve verbal effectiveness over time, and non-linguistically impaired speakers should not.
4. In addition to the research questions (1) to (3), we also investigated whether a measure of communicative efficiency can be derived from the quantitative score of verbal effectiveness in the ANELT. This is of relevance because, as mentioned above, verbal efficiency is not measured in the traditional ANELT.

To conclude, the current study both replicates and appends to the one conducted by Blomert (1990). In the latter study only the scenarios of version I were analysed and the quantitative score was not used for establishing a scoring procedure. As indicated above, the current study seeks to provide a quantitative scoring procedure for all scenarios, which can be beneficial to those involved in the rehabilitation of persons with aphasia.

METHOD

A quantitative scoring procedure for the ANELT

Before we could address the research questions mentioned in the preceding section, we had to develop a quantitative scoring procedure for the ANELT. In doing so we used an adapted version of the Content Unit analysis (Yorkston & Beukelman, 1980). Yorkston and Beukelman have defined a Content Unit (CU) as "a grouping of information that was always expressed as a unit by normal speakers" (p. 30) when presented with the *Cookie Theft* picture of the Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1972). A CU either consists of a single content word (e.g., *boy*) or a phrase (e.g., *on the table*).[1] Synonyms such as *lady* and *woman* are analysed as one and the same CU. We used the CU analysis to verify the extent to which *propositions* were conveyed by non-aphasic speakers. Propositions reflect preverbal messages as they stand for the structure of the lexical concepts. Although propositions correspond to words, one and the same proposition can be put into words in various manners. For example, the propositions *cause(you(go(shirt, from you to me)))* and *new(shirt)* can be verbalised as "I expect you to give me a new shirt" or "You should provide me with a fresh one".

More specifically, we orthographically transcribed the verbal responses that 24 non-linguistically impaired speakers of Dutch had given to the 20 ANELT scenarios. These speakers varied in age between 50 and 58 years ($M = 54$), 50% female. Their level of education (Verhage, 1964) ranged from 4 to 7 ($Mdn = 4$). In Verhage's classification, levels 1 and 2 indicate a low educational level (e.g., elementary school not fully completed); levels 3 through 5 an intermediate level; and levels 6 and 7 a high level of education (e.g., university). It is important to point out that none of these non-aphasic speakers participated in the subsequent phase of the current study. In line with the Content Unit analysis, we labelled (groupings of) information underlying propositions in their responses as CUs. We used a 30% criterion in composing a list of relevant CUs for each ANELT scenario, henceforward referred to as ANELT-CU. Thus each CU that was produced by at least 7 out of the 24 healthy speakers (i.e., 30%) was considered relevant to the communicative goal to be achieved in that particular scenario. One could wonder why a 100% criterion was not chosen here, which would have been in line with Yorkston and Beukelman (1980). The reason is that the non-linguistically impaired speakers who participated in the study of Blomert (1990) showed substantial inter-participant variability in the quantity of content units conveyed. None of the participants produced all elements of meaning. According to Blomert, this was due to the fact that some scenarios allowed alternative routes to achieve the communicative goal. What is more, the healthy control speakers varied in the degree to which they elaborated their responses, which was probably related to individual differences in conversational style. Because of these findings, we opted for a 30% criterion in the current study.

To accommodate the existence of alternative routes in each scenario, we distinguished several relevant preambles (i.e., the reason for starting communication) and requests (i.e., the communicative goal to be achieved). An example may illustrate this. In scenario 1 of the first ANELT version participants are presented with the following instruction: "Suppose you are at the dry cleaner's. When you come to collect this [experimenter handles participant a shirt], you get it back like this [experimenter points out the burn hole at the front of the shirt]. What would you say?" At least 30% of the non-aphasic speakers started conversation by pointing out that the shirt is spoiled. Other non-aphasic speakers, who also represented (at least) 30%, indicated that the shirt's current state is not acceptable. Accordingly, two different preambles were included in the ANELT-CU list (see Table 1). In addition to these preambles, several requests were put forward by (at least) 30% of the non-linguistically impaired speakers: they asked for compensation, a new shirt, or an explanation. Since three

[1] The definition of Content Units implies that words that were produced together were considered to form a unit; however, some inconsistency remains in the definition of CUs. This for example holds for adjectives. Regarding the Cookie Theft Picture of the BDAE, Yorkston and Beukelman (1980) sometimes analysed adjectives as an individual CU (e.g., *little*) and sometimes as part of another CU (e.g., *on the high shelf*).

different requests were communicated, all these requests were considered to be of relevance. In Table 1, P1 and P2 indicate alternative preambles. Similarly, R1 to R3 point to alternative requests.

Whereas a 30% criterion was used in deriving de CU-list, a 100% criterion was used in quantitatively scoring aphasic speakers' responses to the ANELT. That is, in order to be rated *communicatively adequate*, aphasic speakers should produce *all* CUs underlying one of the alternative preambles and requests in their responses. More specifically, although alternative preambles and requests can be produced in order to achieve the communicative goal in each scenario, an aphasic speaker should either produce both CUs underlying P1 or both CUs underlying P2. What is more, the speaker should produce all three CUs underlying one of the three possible requests (i.e., R1, R2, or R3) as well. Only then will the aphasic speaker be assigned the maximum score (see example below). The number of essential CUs per scenario—and thus the maximum score per scenario—ranges from two to nine ($Mdn = 5$).

A protocol was made for the ANELT-CU scoring procedure, which for example includes instructions on handling circumlocutions as well as semantic paraphasias (whenever comprehensible to the context, the former are considered to adequately express a CU, the latter are not). The protocol also provided some examples on assigning scores, among which the following example (Scenario 1 of ANELT version I):

> *Uh d'r is een . . . d'r is een brandvlek in de ... d'r is een brandvlek gemaakt. Dus ik kan deze zo niet meemaken, zo niet mee uh nemen en ik wil hem graag gemaakt hebben.*
> (Er there is a . . . there is a burn hole in the . . . . there is a burn hole $_{\mathbf{P1.1}}$ now. Thus, I cannot experience it like this, cannot take it er back and I $_{\mathbf{R1.2}}$ would like to have $_{\mathbf{R1.1}}$ it fixed $_{\mathbf{R1.3}}$)

In the English translation we marked the essential CUs that were expressed. This speaker opted for Preamble 1, but produced only one of the underlying CUs. In addition he produced Request 1, of which he produced all three underlying CUs. In total, this aphasic speaker was assigned 4 out of the 5 points for his verbal responses to this scenario.

TABLE 1
ANELT-CU scenario 1: Dry cleaner

Instruction:	*"Suppose you are at the dry cleaner. When you come to collect this, you get it back like this [experimenter handles participant a shirt that has a burn hole at the front]. What would you say?"*		
Preamble (P) / Request (R)		Sub-score	Sum score: 5
P1	*spoiled* $_{P1.1}$ *(shirt* $_{P1.2}$*)*	2	2
P2	OR *unacceptable* $_{P2.1}$ *(spoil* $_{P2.2}$*)*	(2)	
R1	**AND** *get* $_{R1.1}$ *(I* $_{R1.2}$*, new one (for free) / it fixed* $_{R1.3}$*)*	3	
R2	OR *get from you* $_{R2.1}$ *(I* $_{R2.2}$*, compensation* $_{R2.3}$*)*	(3)	3
R3	OR *get from you* $_{R3.1}$ *(I* $_{R3.2}$*, explanation* $_{R3.3}$*)*	(3)	

Sub-scores for alternative causes or requests (indicated with "OR") are given in parentheses. In order to be assigned the maximum score speakers should produce all CUs of preamble P1 or P2 as well as all CUs of one request (i. e., R1, R2, or R3). The fact that both a preamble and request should be produced is indicated with "AND" (printed in bold).

Participants

A total of 30 speakers of Dutch participated in the present study.

Healthy control speakers (n = 20). There were 8 male and 12 female healthy control speakers, varying in age between 25 and 63 years ($M = 51$). Their level of education (Verhage, 1964) ranged from 4 to 7 ($Mdn = 5$). The participants had neither neurological damage nor severe visual or hearing disorders. None of the healthy control participants' responses had been used to derive the relevant CUs for the ANELT-CU scoring procedure, which was described in the preceding section.

Aphasic speakers with predominantly expressive disturbances (n = 10). A total of 10 aphasic speakers with predominantly expressive disturbances were recruited among patients of two rehabilitation centres in the Netherlands: the Sint Maartenskliniek in Nijmegen ($n = 7$) and VieCuri in Venlo ($n = 3$). All participants were in the subacute phase. Although all participants were hampered in spoken language production, some of them also demonstrated mild apraxia of speech (IH, FM, FP, AS, and MP). These five participants nevertheless produced at least 95% intelligible speech output (for participants' characteristics, see Table 2). In using the nonparametric discriminant analysis programme (ALLOC) of the Dutch AAT (Graetz et al., 1992), participants IH, AS, and TV were classified as Wernicke's aphasia. However, the clinical presentation of their language disorders—as judged by their trained speech and language therapist—was indicative of *predominantly expressive disturbances* (Weisenburg & McBride, 1935). Since it is acknowledged that ALLOC classification and clinical presentation do not always coincide (e.g., Günther, Hofman, & Promes, 2009), participants IH, AS, and TV were nevertheless included in the present study.

Design and assessments

Our experiment was for each outcome measure a split-plot design with two fixed factors: the between-participant factor Group, which contains two levels: aphasic and healthy control speakers, and the within-participant factor Time: T1 and T2. The time interval between T1 and T2 was 8 weeks, in which the control speakers did not receive any intervention. The aphasic speakers received therapy aiming for restoration of their spoken language production difficulties. Both goal and intensity of treatment were somewhat different for each aphasic speaker. However, the current study does not seek to evaluate therapy outcome; focus is on the systematic exploration of a quantitative measure of functional communication. Therefore the fact that guided recovery (i.e., recovery that is increased as a result of training, see Robertson & Murre, 1999) was occurring was more important than establishing what induced it.

The three outcome measures were: the rating of verbal adequacy on the 5-point ANELT A-scale (henceforward referred to as *ANELT-traditional*), the percentage of relevant Content Units produced in the ANELT (i.e., ANELT-CU), and the percentage of essential information units expressed in a Picture Description Task (PDT; Ruiter, 2008). In the PDT, participants are instructed to describe 40 pictures of every day life activities (e.g., a man who squeezes oranges in the kitchen), which are displayed on a computer screen. The task is self-paced and participants do not receive any feedback on their spoken language production. The PDT was used in the current study as a reference test, because it also measures verbal effectiveness quantitatively.

TABLE 2
Characteristics of aphasic speakers in sub-acute phase after stroke ($n = 10$)

Participant		*Age (yrs)*[a]	*TPO (weeks)*[b]	*Syndrome AAT*[c]	*Severity aphasia*[d]	*Handedness*	*Aetiology / localisation*[e]	*Education*[f]
IH	f	49	8	Wernicke	47.43	R	SAH, LH, parietal	4
HS	f	52	48	*md*	*md*	R	*md*	5
BM	f	49	11	Broca	36.81	R	ICVA, LH, arteria cerebri media	5
FM	m	57	7	Broca	43.67	L	ICVA, LH, exact location unknown	6
FP	m	61	8	Broca	40.97	R	ICVA, LH, exact location unknown	5
AS	f	77	5	Wernicke	58.20	R	ICVA, LH, exact location unknown	6
TV	m	67	8	Wernicke	51.63	R	ICVA, LH, exact location unknown	6
MP	f	57	10	Broca	44.38	R	HCVA, LH, exact location unknown	5
JH	m	51	6	*md*	*md*	L	*md*	5
FK	m	64	7	*md*	*md*	R	*md*	4
Mean		58	12					
Range		49–77	5–48					

Participants HS, JH, and FK have missing data (*md*) for the aphasia syndrome, severity of aphasia and aetiology.
[a, b]Age (in years) versus time post-onset (in weeks) at T1.
[c]AAT diagnosis based on the nonparametric discriminant analysis programme (ALLOC).
[d]Severity of aphasia is expressed as the average profile height ($M = 50$, $SD = 10$) of the Dutch Aachen Aphasia Test (Graetz et al., 1992)
[e]SAH = subarachnoid haemorrhage, ICVA = ischaemic cerebrovascular accident, HCVA = hemorrhagic cerebrovascular accident.
[f] Level of education according to Verhage (1964): levels 1 and 2 indicate a low educational level; levels 3 through 5 an intermediate level; and levels 6 and 7 a high level of education.

That is, a list of essential Content Units is available for the PDT. This list was established on the basis of the verbal responses that 10 linguistically unimpaired speakers of Dutch had given to the PDT when participating in the study of Ruiter (2008).[2] To the best of our knowledge there are no other standardised and norm-referenced tests available that measures verbal effectiveness quantitatively.

Procedure and statistical analyses

At T1 and T2 all 30 participants were presented with both versions of the Dutch ANELT as well as the PDT. The order in which the participants were presented with the tests was counterbalanced. Participants' responses to both tests were tape-recorded. With the audio fragments, a trained speech and language pathologist rated

[2]As the PDT had yielded homogeneous verbal responses of the 10 healthy control speakers in the study by Ruiter (2008), the criterion for relevance was set at 75% for the PDT, which is in line with Christiansen (1995) as well as Huber (1990). As discussed above, the ANELT had yielded much more heterogeneous responses from healthy control speakers (Blomert, 1990). That is the reason why the criterion for relevance was set at 30% for the ANELT-CU scoring procedure.

each speaker's verbal adequacy on the 5-point ANELT A-scale. Subsequently, the audio-fragments of the PDT and ANELT were orthographically transcribed. Using scoring protocols for the ANELT-CU and the PDT, the speech and language pathologist derived a quantitative score of verbal effectiveness for both the ANELT and PDT. The ANELT A-scale ratings were not available to the speech and language pathologist when doing the quantitative rating. On average, it took 1 hour to derive an ANELT-CU score for each aphasic speaker, the time needed for orthographic transcription included. A second speech and language pathologist scored the verbal responses of three randomly chosen aphasic speakers (FM, FP, and AS) and three control speakers (DJ, HM, and PB) as well. In order to be able to derive a measure of verbal efficiency the ANELT samples were timed. All time taken by the experimenter's speech was subtracted; however, pauses that occurred prior to or within the boundaries of the aphasic speaker's turn of talk were included in the time for the participant's speech, as suggested by Oelschlaeger and Thorne (1999). Subsequently the total number of words was counted. Verbal efficiency was operationally defined as the average number of Essential Content Units per minute (*ECU/min*). In calculating this parameter, 30 seconds was used as a cut-off to round to full minutes (i.e., n resp. $n+1$ minutes).

For all outcome measures obtained with the randomly chosen control speakers ($n = 3$) and aphasic speakers ($n = 3$), Intra-class Correlation Coefficients (ICCs, objects random, raters fixed, using the consistency definition) were calculated at T1, which allowed investigation of research question (1). The raters were not randomly chosen from a large population of raters, that is why we did not label them as a random factor. However, this does not preclude generalisation to other raters, as we are convinced that the raters could be replaced by other raters who, of course, should receive the same amount of training and show the same degree of interest in the subject.

To investigate questions (2) to (4) we planned to use two-way repeated-measures ANOVA (GLM repeated measures procedure of SPSS 16.0). However, homogeneity of variances—a basic assumption in analysis of variance—was intrinsically not warranted (as confirmed by Levene's tests, which were always significant, also after arcsin transformations of the data). As expected, the control group was much more homogeneous and their scores were in close proximity to ceiling level. That is why we resorted to non-parametric alternatives implemented in SPSS 16.0: Wilcoxon's matched pairs signed ranks test to assess the changes between the scores obtained at T1 and T2 for each of the groups and each of the dependent measures, and bootstrapped Welch's t tests for independent samples to assess differences between the groups at each moment in time and for each of the dependent variables (Boos, 2003; Ruxton, 2006). One-tailed p values are reported, as the expectations (H1s) were directional; for the Wilcoxon tests exact p values were calculated. One of the effects of using one-tailed tests is a welcome one: an increase in power. Since six comparisons were made between the groups in total as well as three comparisons per group across time, we used the Holm method (e.g., Holm, 1979) to control for inflation of the Type I error. Thus we report Holm's adjusted p levels for all outcome measures.

Sensitivity to change in verbal effectiveness over time was assessed with d_{paired}, which is one of the available effect size indices used to gauge the responsiveness of scales to clinical change. In the current study Pearson's correlation coefficient, r, was used not only as a measure of the strength and direction of linear dependency between two variables, but also as an indication of the strength of the experiment effect.

RESULTS AND DISCUSSSION

Inter-rater agreement

For each of the randomly chosen aphasic speakers, ICCs were calculated for the outcome measures. The ICCs for ANELT-traditional ranged from .656 to .926 (M = .787), for ANELT-CU from .712 to .953 (M = .821), and for the PDT from .914 to .979 (M = .948). ICCs for the control speakers ranged from .892 to 1.000 (M = .942) in ANELT-CU and from .918 to .971 (M = .942) in the PDT.[3] All ICCs were significantly different from zero ($p < .001$) and—using the criteria of Landis and Koch (1977, in Rietveld & Van Hout, 2005) for interpretation—the strength of the relation between all ratings was substantial (i.e., .61–.80) to almost perfect (i.e., >.81).

Distinguishing healthy and aphasic speakers

As Table 3 shows, the two groups differed significantly on all three dependent variables (i.e., ANELT-traditional, ANELT-CU, and PDT) at each point in time: all p values < .01. So, at each point in time, the healthy control speakers scored significantly higher than the aphasic speakers.

Stable performance over time for healthy control speakers

It was expected that the control speakers would show stable performance over time on all outcome variables. In line with our expectations, non-significant differences between the scores obtained at T1 and T2 were found for ANELT-CU as well as the PDT. For ANELT-traditional, however, the control speakers had significantly lower scores at T2 than at T1 (cf. Table 4). The latter finding may be due to the fact that the healthy control speakers' responses nearly reached ceiling level and were not spread out much (as indicated by the low *SD*) at T1. In a situation with a very small standard error, even a small decrease in the average score, as occurred at T2, may lead to significance.

TABLE 3
Differences in percentage scores between control (n = 20) and aphasic speakers (n = 10) at each point in time

Outcome measure		*Control speakers*		*Aphasic speakers*		*M difference*	*Bootstrapped Welch t test*
ANELT-traditional	T1	99.55	(*.69*)	64.10	(*13.45*)	35.45	$t(28) = -8.327, p < .001$
	T2	98.55	(*1.85*)	69.00	(*13.16*)	29.55	$t(28) = -7.067, p < .001$
ANELT-CU	T1	80.98	(*10.16*)	45.80	(*14.48*)	35.18	$t(28) = -6.881, p < .001$
	T2	80.69	(8.13)	52.46	(*15.11*)	28.23	$t(28) = -5.523, p < .001$
PDT	T1	90.49	(*6.52*)	51.02	(*23.09*)	39.47	$t(28) = -5.302, p < .001$
	T2	90.43	(*6.42*)	60.94	(*19.41*)	29.49	$t(28) = -4.678, p < .001$

[3]Because of a ceiling effect, the ICC is not given for the ANELT-traditional administered to the healthy control speakers.

TABLE 4
Average percentage scores (*SD*) at T1 and T2; changes in scores over time for both control ($n = 20$) and aphasic speakers ($n = 10$)

Outcome measure	*Group*	*T1*		*T2*		Δ_{T2-T1}	*Wilcoxon signed-rank test*	*Effect size*
ANELT-traditional	control	99.55	(*.69*)	98.55	(*1.85*)	−1.00	$T^+ = 3.5, p = .009$	$r = -.42$
	aphasic	64.10	(*13.45*)	69.00	(*13.16*)	4.90	$T^- = 19, p = .209$	$r = -.19$
ANELT-CU	control	80.98	(*10.16*)	80.69	(8.13)	.29	$T^+ = 93, p = .654$	$r = -.07$
	aphasic	45.80	(*14.48*)	52.46	(*15.11*)	6.66	$T^- = 1, p = .006$	$r = -.60$
PDT	control	90.49	(*6.52*)	90.43	(*6.42*)	−.06	$T^- = 77, p = .654$	$r = -.06$
	aphasic	51.02	(*23.09*)	60.94	(*19.41*)	9.92	$T^- = 2, p = .006$	$r = -.58$

Increased performance over time for aphasic speakers

We hypothesised that the aphasic speakers would perform better on all dependent measures as a result of guided recovery. In line with our expectations, significant differences between the scores obtained at T1 and T2 were found for all dependent variables, except for ANELT-traditional (cf. Table 4).

Responsiveness

As indicated above, the aphasic speakers on average exhibited higher performance on ANELT-CU and PDT over time. In line with these findings the lowest value of d_{paired} was obtained in the aphasic speakers with ANELT-traditional (.31). When compared to ANELT-traditional, ANELT-CU yielded a higher d_{paired} value (1.36). This value was even somewhat higher for the PDT (1.44). The fact that PDT and ANELT-CU correlated positively and significantly at T2 ($r_8 = .773$, $p = .009$), whereas PDT and ANELT-traditional did not ($r_8 = .308$, $p = .386$), further supports the hypothesis that ANELT-CU more sensitively measures verbal effectiveness in aphasic speakers than ANELT-traditional.

Verbal efficiency

At T1 the aphasic participants on average produced 4.50 ECU/min (*SD* 2.44) in ANELT parallel versions I and II. At T2, this was 6.83 ECU/min (*SD* 4.26). A Wilcoxon signed-ranks test revealed that the 10 aphasic speakers on average increased verbal efficiency over time: $T^- = 3$, $p = .005$, effect size $r = -.56$.

GENERAL DISCUSSION

The study reported here sought to develop a quantitative score of verbal effectiveness, based on the elements of meaning that are essential to achieve the communicative goal in each ANELT scenario. We investigated whether such a quantitative score increases responsiveness of the ANELT and also allows derivation of a measure of verbal efficiency. Although our findings need to be replicated in a larger group of aphasic speakers, the main results suggest that there is a positive answer to both questions.

The fact that valid and reliable measures of verbal effectiveness as well as efficiency can be calculated for the ANELT allows clinicians to obtain a more complete picture of aphasic speakers' functional communication skills. Further research is warranted in order to establish critical differences for the quantitative outcome measures for the ANELT, which are essential to decide on individual patient's performance difference. Critical differences are also needed for conducting studies into the efficacy and effectiveness of aphasia therapy. Only then the new outcome measures may become a new standard for measuring verbal effectiveness and efficiency in aphasic speakers.

In interpreting the aphasic speaker's functional communication skills it is important to point out that aphasia therapy may affect both aspects differently in the several phases of rehabilitation. Whereas both aspects of functional communication may increase as a result of restoration therapy, compensation therapy may affect verbal effectiveness and efficiency in a different way, depending on the strategy used. For example, some chronically anomic speakers produce a hypernym (e.g., *animal*) when they fail to retrieve a word (e.g., *giraffe*). Teaching such speakers to produce a circumlocution (e.g., *A tall animal that lives in Africa. It has a very long neck and legs*) may lead to an increase in verbal effectiveness. However, verbal efficiency is likely to decrease as the speaker produces more words.

Another example regarding the effect of compensation therapy on functional communication relates to the error-strewn production of sentences that may occur in chronically agrammatic speakers. The *Dutch and adapted version of Reduced Syntax Therapy* (REST; Ruiter, Kolk, & Rietveld, 2010, for the original REST programme see Schlenck, Schlenck, & Springer, 1995) aims at the continuous use of normal but pre-morbidly infrequently used language routine: the production of ellipses. Ellipses are grammatically well formed, but incomplete utterances, such as *problem solved* or *Tim drinking coffee* (Progovac, 2006). When such speakers learn to simplify their grammatical output by producing ellipses continuously, fluency is typically increased because the complexity of the utterance is now optimally tuned to the speaker's linguistic capacity (Kolk, 2006). The aphasic speakers who participated in the study of Ruiter, Kolk, and Rietveld typically compensated the reduction in utterance length by increasing the number of utterances produced. As a consequence the amount of information conveyed on average remained the same. Thus verbal efficiency increased in these chronically agrammatic speakers of Dutch, whereas hardly any improvement in verbal effectiveness was observed (Ruiter et al., 2010). These examples illustrate that further research is warranted in order to investigate the effect that compensation therapy has on verbal effectiveness and efficiency.

Clinical implications

A clinically relevant matter that has to be addressed is the time involved to derive a quantitative score of verbal effectiveness in the ANELT. Compared to ANELT-traditional, clinicians need to spend more time to derive the ANELT-CU score. Although the amount of time required for the ANELT-traditional procedure varies between speakers, depending on their functional communication skills, time needed for scoring is almost one-to-one related to the participant's time for speech. That is, clinicians assign their scores to the A-scale while playing back the audio-tape. The ANELT-CU scoring procedure, on the other hand, is much more time-consuming, because it requires transcription of the speech sample and marking of the CUs in the transcription. It is commonly acknowledged that quantitative scoring procedures are very time-consuming (e.g., Prins & Bastiaanse, 2004). Our experience is that the time needed to derive a quantitative score of verbal effectiveness is at least doubled, when compared to ANELT-traditional. We nevertheless are inclined to believe that the ANELT-CU scoring procedure is valuable in both academic and clinical practice. Compared to the traditional test, ANELT-CU seems more sensitive to measure improvement in verbal effectiveness and also allows measurement of verbal efficiency. This is of relevance because, in evaluating the efficacy and effectiveness of aphasia therapy, both aspects of functional communication need to be investigated. Regarding the increased time effort, human language technology may be used to provide reliable and valid quantitative measures of functional communication without human intervention. That is why we plan to have language and speech technology developed for the automatic calculation of degree of functional communication on the basis of speech recordings.

Manuscript received 14 November 2010
Manuscript accepted 5 March 2011
First published online 27 May 2011

REFERENCES

Bastiaanse, R., Hurkmans, J., & Links, P. (2006). The training of verb production in Broca's aphasia: A multiple-baseline across-behaviours study. *Aphasiology*, *20*, 298–311.

Blomert, L. (1990). What functional assessment can contribute to setting goals for aphasia therapy. *Aphasiology*, *4*, 307–320.

Blomert, L. (1994). *Assessment and recovery of verbal communication in aphasia*. Doctoral dissertation, Radboud University, Nijmegen.

Blomert, L., Kean, M. L., Koster, Ch., & Schokker, J. (1994). Amsterdam-Nijmegen Everyday Language Test: Construction, reliability and validity. *Aphasiology*, *8*, 381–407.

Blomert, L., Koster Ch., & Kean, M-L. (1995). *Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheid: Handleiding* [Amsterdam-Nijmegen Everyday Language Test: Manual]. Lisse: Swets & Zeitlinger.

Boos, D. D. (2003). Introduction to the Bootstrap World. *Statistical Science, 18*, 168–174.

Christiansen, J. A. (1995). Coherence violations and propositional usage in narratives of fluent aphasics. *Brain and Language*, *51*, 291–317.

Doesborgh, S. J. C., Van de Sandt-Koenderman, M., Dippel, D. W. J., Van Harskamp, F., Koudstaal, P. J., & Visch-Brink, E. (2004). Effects of semantic treatment on verbal communication and linguistic processing in aphasia after stroke: A randomized control trial. *Stroke*, *35*, 141–146.

Frattali, C. M. (1992). Functional assessment of communication: Merging public policy with clinical views. *Aphasiology*, *6*, 63–83.

Goodglass, H., & Kaplan, E. (1972). *Boston Diagnostic Aphasia Examination*. Philadelphia: Lea & Febiger.

Graetz, P., de Bleser, R., & Willmes, K. (1992). *De Akense Afasietest* [The Aachen Aphasia Test]. Lisse: Swets & Zeitlinger.

Grande, M., Hussmann, K., Bay, E., Christoph, S., Piefke, M., Willmes, K., et al. (2008). Basic parameters of spontaneous speech as a sensitive method for measuring change during the course of aphasia. *International Journal of Language and Communication Disorders*, *43*, 408–426.

Günther, T., Hofman, M., & Promes, M. (2009). Afasiesyndromen: Twijfels over de klassieke taxonomie [Aphasia syndromes: Questioning the classic taxonomy]. *Logopedie en Foniatrie*, *5*, 148–152.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Huber, W. (1990). Text comprehension and production in aphasia: Analysis in terms of micro- and macroprocessing. In Y. Joanette & H. H. Brownell (Eds.). *Discourse ability and brain damage* (pp. 154–179). New York: Springer-Verlag.

Irwin, W. H., Wertz, R. T., & Avent, J. R. (2002). Relationships among language impairment, functional communication, and pragmatic performance in aphasia. *Aphasiology*, *16*, 823–835.

Kay, J., Lesser, R., & Coltheart, M. (1995). *Psycholinguistic Assessments of Language Processing in Aphasia*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Kolk, H. (2006). How language adapts to the brain: An analysis of agrammatic aphasia. In L. Progovac, K. Paesani, E. Casielles, & E. Barton (Eds.), *The syntax of nonsententials: Multidisciplinary aspects* (pp. 229–258). Amsterdam: John Benjamins Publishing Company.

Krippendorff, K. (1970). Estimating the reliability, systematic error and random error of interval data. *Educational and Psychological Measurement*, *30*, 61–70.

Laska, A.C., Hellblom, A., Murray, V., Kahan, T., & Von Arbin, M. (2001). Aphasia in acute stroke and relation to outcome. *Journal of Internal Medicine*, *249*, 413–422.

Marini, A., Caltagirone, C., Pasqualetti, P., & Carlomagno, S. (2007). Patterns of language improvement in adults with non-chronic non-fluent aphasia after specific therapies. *Aphasiology*, *21*, 164–186.

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.

Prins, R., & Bastiaanse, R. (2004). Review: Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, *18*, 1075–1091.

Progovac, L. (2006). The syntax of nonsententials: Small clauses and phrases at the root. In L. Progovac, K. Paesani, E. Casielles, & E. Barton (Eds.), *The syntax of nonsententials: Multidisciplinary aspects* (pp. 33–71). Amsterdam: John Benjamins Publishing Company.

Ramsberger, G. (2005). Achieving conversational success in aphasia by focusing on non-linguistic cognitive skills: A potentially promising new approach. *Aphasiology*, *19*, 1066–1073.

Rietveld, T., & van Hout, R. (2005). *Statistical techniques for the study of language and language behaviour*. Berlin: Mouton de Gruyter.

Robertson, I. H., & Murre, J. M. J. (1999). Rehabilitation of brain damage: Brain plasticity and principles of guided recovery. *Psychological Bulletin*, *125*, 544–575.

Ross, K. B., & Wertz, R. T. (1999). Comparison of impairment and disability measures for assessing severity of, and improvement in, aphasia. *Aphasiology*, *13*, 113–124.

Ruiter, M. B. (2008). *Speaking in ellipses: The effect of a compensatory style of speech on functional communication in chronic agrammatism*. PhD thesis, Radboud University, Nijmegen. Retrieved 23 February 2009 from http://webdoc.ubn.ru.nl/mono/r/ruiter_m/speainel.pdf

Ruiter, M. B., Kolk, H. H. J., & Rietveld, A. C. M. (2010). Speaking in ellipses: The effect of a compensatory style of speech on functional communication in chronic agrammatism. *Neuropsychological Rehabilitation*, *20*, 423–458.

Ruxton, G. D. (2006). The unequal variance *t*-test is an underused alternative to Student's *t*-test and the Mann-Whitney *U* test. *Behavioral Ecology*, *17*, 688–690.

Schlenck, C., Schlenck, K. J., & Springer, L. (1995). *Die Behandlung des schweren Agrammatismus – Reduzierte-Syntax-Therapie (REST)* [Treatment of severe agrammatism – Reduced Syntax Therapy (REST)]. Stuttgart: Thieme.

Tompkins, C. A., Scharp, V. L., & Marshall, R. C. (2006). Communicative value of self cues in aphasia: A re-evaluation. *Aphasiology*, *20*, 684–704.

Verhage, F. (1964). *Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar* [Intelligence and age: Study with Dutch people from age 12 to 77]. Assen: Van Gorcum.

Weisenburg, T. H., & McBride, K. E. (1935). *Aphasia: A clinical and psychological study*. New York: Commonwealth Fund.

Worrall, L. (1995). The functional communication perspective. In D. Muller & C. Code (Eds.), *Treatment of aphasia*. London: Whurr Publishers.

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.