Journal of Psycholinguistic Research, Vol. 31, No. 5, September 2002 (© 2002)

On the Categorization of Aphasic Typologies: The SOAP (A Test of Syntactic Complexity)

Tracy Love[1,3] and Elizabeth Oster[1,2]

*This paper presents a new measure of syntactic comprehension abilities in brain-damaged populations known as the SOAP (**S**ubject-relative, **O**bject-relative, **A**ctive, and **P**assive), along with data supporting its sensitivity and specificity. This assessment tool examines comprehension of sentences (matched for length) of four syntactic construction types: active, passive, subject-relative, and object-relative. Data are presented that indicate that the SOAP provides a sensitive and reliable differentiation of aphasia subgroups. The SOAP's sensitivity in differentiating broad behavioral (anterior/posterior-lesioned) groups is compared to the auditory comprehension component of the Boston Diagnostic Aphasia Examination (BDAE), supporting its sensitivity differentiating between anterior- and posterior-lesioned groups. It is argued that this tool can be an important accompaniment to standard aphasia assessment batteries in more sensitively defining syntactic comprehension deficits.*

KEY WORDS: aphasia; syntax; assessment; sentence comprehension.

INTRODUCTION

A number of behavioral tests in the fields of neurolinguistcs and aphasiology have sought to address issues of how varying types of brain damage can result in language impairment. Most diagnostic tests, such as the Porch Index of Communicative Ability (PICA; Porch, 1967), the Minnesota Test for Differential Diagnosis of Aphasia (MTDDA; Schuell, 1965), the

The authors wish to acknowledge the support of NIH grant R01 DC02984 and to thank Tamar Gollan and Vikki Bouck for helpful comments during earlier drafts of this paper and to Glenice Tobiano and Yosef Grodzinsky for contributions to the work presented in this paper.

[1] University of California, San Diego, California.

[2] San Diego State University, San Diego, California.

[3] To whom all correspondence should be addressed at University of California, San Diego, Department of Psychology-0109, La Jolla, CA 92093-0109. email: tlove@ucsd.edu

0090-6905/02/0900-0503/0

Western Aphasia Battery (WAB; Kertesz, 1982), or the Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1972), assess varying levels of comprehension and production (e.g., single word comprehension for presented nouns and verbs or open and closed class items; the comprehension and production of simple sentences; repetition). These measures have afforded researchers and clinicians the opportunity to begin characterizing specific language deficits in service of disentangling the complex systems involved in language comprehension and production and to help direct treatment and therapy.

The original goals of these classic tests were twofold. First, they were intended as comprehensive assessments that could provide a basis for the diagnosis and typology of aphasias. Second, they were guides for inferring the cerebral localization of neural damage; when a particular pattern indicative of a syndrome was present, the probable site of brain lesion could be *inferred.* The advent of modern neuroimaging techniques has supplanted the need for the latter of those goals. We argue below that neuroradiological findings should be used *a priori* to classify patients before behavioral data collection and analysis.

However, these classic behavioral tests commonly focus on the general functioning of broad language subsystems, such as comprehension and production as a whole. There are only minimal attempts in these tests to analyze specific language subsystems. Because of the range of issues targeted in these broad diagnostic measures, one cannot expect them to necessarily demonstrate sensitivity to individual subsystems of language processing. For this reason, a number of more focused subtests have begun to be developed by researchers to target specific subsystems of interest (see, for example, BDAE 3rd edition, Goodglass, Kaplan, & Barresi, 2001). The test presented in this paper was developed to provide a means of differentiating the abilities of brain-damaged populations to understand various syntactic constructions presented in the auditory modality.

The goal of this paper is twofold. One is to review the efficacy of this new supplementary assessment tool (termed the SOAP; **S**ubject-relative, **O**bject-relative, **A**ctive, **P**assive test) as a valid and sensitive measure of certain comprehension abilities in brain-damaged populations. The second is to present evidence speaking to the need for more reliance on lesion-based classification of aphasic patient populations (in addition to behavioral assessments). It is argued that there exists a critical interdependence between the behavior and lesion basis of classification that has, in the past, been disregarded in lieu of concentrating on one or the other generalization. With regard to the first issue, prior observations have held that patients with diagnoses of agrammatism have specific difficulties with certain types of sentence constructions, in particular those sentences with

moved constituents[4] (e.g., Drai, Grodzinsky, & Zurif, 2001; Hickok & Avrutin, 1995a, b; Zurif, Gardner, & Brownell, 1989). Indeed, a number of research laboratories cite anecdotal and informal evidence supporting the further investigation and evaluation of this phenomenon via nonstandardized, homegrown tools. Unfortunately, the development of such informal assessments makes comparisons across studies difficult, if not impossible. In this paper, we will demonstrate that the SOAP can reliably distinguish groups of patients who have difficulty understanding such complex constructions yet display seemingly intact auditory comprehension on the BDAE. In particular, patients who have sustained damage to the anterior regions of the left hemisphere all demonstrate these specific patterns of performances (including, but not exclusively, agrammatic aphasics).

METHODS

Subjects

Seventy-six subjects participated in the study. Testing took place at The Laboratory for Research on Aphasia and Stroke (LRAS) at The University of California, San Diego (UCSD) or at The Aphasia Research Center at the Boston Veterans' Administration Medical Center. Subjects participated in one of three testing groups: (1) those with neural involvement, (2) younger unimpaired controls, and (3) older (age-matched) unimpaired controls.

Group 1: Neural Involvement

Twenty-six neurologically involved participants are included in this report. The participants ranged in age from 39 to 72 years old at the time of testing (mean age = 61.73). All subjects were native English speakers and were right-handed before their stroke (except for one participant who was, based on self-report, left-handed premorbidly [TY]). All met the following criteria: a *single*, relatively localized lesion site; no previous history of infarct; neurologically and physically stable (i.e., at least 6 months post onset); no history of active or significant alcohol and/or drug abuse; no history of active psychiatric illness; no history of other significant brain disorder or dysfunction (e.g., Alzheimer's/dementia, senility, Parkinson's, Huntington's, Korsa-

[4] In English, and many other languages, the canonical ordering of words is generally taken to be Subject-Verb-Object (e.g., The boy kicks the ball.). But it is legal to "move" the object in front of the verb (e.g., The ball is kicked by the boy.). In such noncanonical constructions, it is also commonly argued that the listener must reconstruct the underlying Subject-Verb-Object ordering so as to ultimately comprehend the sentence.

koff's, mental retardation); and corrected-to-"normal" auditory and visual acuity for age. Participants were paid $15 for their participation in this study.

Left Hemisphere-Damaged

For the left hemisphere–involved participants ($N = 22$), diagnosis of aphasia was based on the convergence of clinical consensus and the results of a standardized aphasia examination—the BDAE (Goodglass & Kaplan, 1972; see Appendix 1 for summary of scores for auditory comprehension subtests). At the time of testing, all participants had retained the defining features of their original diagnosis. On this basis, of the 22 left hemisphere–involved participants, 15 anterior-lesioned patients were diagnosed and classified as expressive aphasics with severity rating ranging from 1 to 4 (on a low-to-high oral communication capacity scale of 0 to 5, see Appendix 2 for breakdown of BDAE severity rating scale). Those expressive aphasics who received a BDAE severity rating below 3 (hereafter noted as severe expressive aphasics) presented with nonfluent, telegraphic, verbal output and with relative sparing of auditory comprehension at the conversational level. Mild expressive aphasics were those who received a BDAE score of 3 or greater displayed fairly fluent, grammatically well-formed speech with word-finding difficulties marked with circumlocutory strategies and with relatively intact auditory comprehension. Seven posterior-lesioned patients were diagnosed and classified as receptive aphasics, with all but one receiving severity ratings of 1 or 2 (with the remaining participant receiving a score of 3.5). The receptive aphasics maintained fluent, relatively empty speech and tended to show more overall comprehension problems than did the severe expressive aphasics.

Neurally Involved Controls

In addition to these core groups, one left hemisphere–damaged (LHD) nonaphasic and three right hemisphere–damaged (RHD) participants served as brain-injured controls, with severity ratings of oral communication skills ranging from 4.5 to 5.

A correlation analysis of the assigned BDAE severity rating scores and the achieved BDAE overall auditory comprehension score for the neurally involved participants demonstrated strong intratest reliability [$r = 0.631$; $F(1,13) = 8.611$, $p = .01$]. As a result, it was determined that we could confidently group the participants based on severity ratings (knowing that this measure accurately reflects the BDAE objective scores).

Group 2: Younger Unimpaired Controls (No Neural Involvement)

Forty-six undergraduates from UCSD participated in this experiment for 1 hr of course credit. All subjects were right-handed, native English speakers

with no history of learning disabilities or neural injury, and "normal"-to corrected visual acuity and hearing. Subjects ranged in age from 18 to 22 years of age (mean age 19.7).

Group 3: Older Mean Age-Matched Unimpaired Controls (No Neural Involvement)

As an additional control for extraneous (i.e., age-related) variability, four unimpaired subjects were specifically age- and socioeconomic-matched with four of the LHD participants. This group was examined to provide a closer mean age-matched population; the mean age was 62 (ages ranged from 47 to 74). All subjects were native English speakers, right-handed, had corrected-to-"normal" visual and auditory acuity for age, and had no history of (a) active or significant alcohol and/or drug abuse, (b) active psychiatric illness, or (c) other significant brain disorder or dysfunction. Participants were paid $15 for their participation in this study. As will be seen in detail below, performance between groups 2 and 3 did not differ with respect to SOAP performance.

Table I provides the available demographic information for each neurologically involved participant and the accompanying neuroradiological (CT/MRI) findings, where appropriate. In addition, demographic information for the older age-matched controls (group 3) is also presented.

MATERIALS

The SOAP syntactic battery is composed of the following materials: 5 practice sentences (involving active and subject-relative constructions only) and 40 experimental sentences targeting reversible actions with active, passive, subject-relative, and object-relative constructions (10 each). Appendix 3 provides a list of the entire test.

Experimental sentences were controlled across syntactic category for length per sentence (within 2 words), and were pseudo-randomized within the test so that no more than two items of a particular syntactic structure were presented consecutively. Practice and experimental sentences were matched with simple line drawings involving two characters. Correct experimental pictures ("match" condition) involved the agent and patient described in the sentence accurately depicting the action. Incorrect experimental pictures ("mismatch" condition) involved the same two characters but with a reversal of thematic roles (so that the patient was depicted as the agent). Defining modifiers (such as color of hair, shirt, pants, etc.) were replicated for each character per item to avoid possible nonsyntactic cues. The third (foil) picture consisted of items/characters unrelated to the characters in the sentence, as

Table I. Demographic and Lesion Information

Expressive Aphasia (N = 15)							
Subject	BDAE Aphasia Severity Level	Gender	Age at testing	Years post onset	Hemiparesis?	Education	Lesion
TA	1	M	52	11	R/wheelchair	M.D.	L frontal lobe extending into parietal and temporal pole regions, sparing the STS
YY	1	F	62	1½	R/wheelchair	M.A.	Impression: Basal ganglia ischemic insult; left parasylvian and peri-insular ischemic insult—particularly the insular cortex and superior opercular regions bordering the Rolandic fissure
LK	2	M	60	8	R weakness	MA	L basal ganglia, deep white matter, frontoparietal cortex
LN	1	M	72	11	R weakness	10th grade	Left frontoparietal infarct
AH	1	F	39	4	R weakness	HS	L basal ganglionic structures, periventricular white matter, posterofrontal lobe
HPR	1	M	72	17	R weakness	BA	Massive cerebral infarct, L middle cerebral artery with anterior extension of infarct
TR	1	F	46	1½	R weakness	HS	L MCA extending into anterior frontal regions.
HN	2	M	68	3	R weakness	Jr. College	L frontoparietal infarct involving perislyvian region
NN	1	M	55	7	R weakness	HS	Left periventricular lesions adjacent to the lateral ventricle, involvement of temporal operculum and insular region extending into basal ganglia
BJ	2	F	50	2	R weakness	Nursing	L frontoparietal lesion involving all inferior frontal gyrus
DR	1.5	M	82	21	R weakness	2 years college	L frontal lesion(s) involving Broca's area with deep extension across to L frontal horn and into anterior temporal pole sparing Wernicke's area.
YRY	3	M	52	2½	R weakness	BA	L temporoparietal lobe
TH	4	M	61	5	NO	8th grade	L frontal extending basal ganglia, internal capsule, lenticular nucleus
EN	4	F	52	14	R weakness	2 yrs college	L basal ganglia, internal capsule, lenticular nucleus
TY	4	M	52	4	R weakness	HS	L frontal cortical region and deeper structures in the basal ganglia

Receptive Aphasia (N = 7)							
RL	2	M	58	2	NO	MBA	Extends from L temporal lobe into parietal and occipital lobes without significant mass effect
HTN	2	F	78	1	NO	MA	L temporal extending into parietal and occipital lobes
RH	2	F	73	2	NO	HSS	L MCA involving temporal lobe
CC	3.5	M	66	8	NO	HS	LH temporal lesion extending into SMG and angular gyrus (surface and deep) extending to occipital lobe
BuJ	1	M	56	1	NO	HS	L temporoparietal involving Wernicke's area extending into MTG and posterior SMG
HR	1	M	67	3	NO	HS	L MCA involving temporal lobe
TAN	1	M	68	3	NO	Ph.D.	Left middle cerebral artery stroke involving the central parietal and temporal areas.
RHD (N = 3)							
NE	4.5	M	63	5	L/weakness	MA	R temporoparietal lobe
EIN	5	M	67	3½	L weakness	MA	R medial temporal lobe
NR	5	M	62	7 mos.	L/wheelchair	Jr. college	R temporoparietal lobe
LHD (N = 1)							
KY	5	M	72	1½	R weakness	BA	Left pontine infarction
Unimpaired Control (N = 4)							
BJ	N/A	F	57	N/A	N/A	14 years	N/A
CR	N/A	F	47	N/A	N/A	H.S.	N/A
QC	N/A	F	71	N/A	N/A	B.A.	N/A
SM	N/A	F	74	N/A	N/A	H.S.	N/A

well as an unrelated action. Picture order for each sentence was pseudo-randomized so as to discourage perseverative responses.

PROCEDURE

Subjects were read instructions informing them of the task. As demonstrated in Fig. 1, they were shown three pictures for each test item. Before reading each sentence, the experimenter identified the characters in each of the three pictures in turn, using the defining features critical to the sentence. So, for the example in Fig. 1, the experimenter said, "This is the woman," "This is the child," "This is the old man," and "This is the little boy."

Once the subject was familiar with the characters, the relevant sentence was read aloud twice at a normal rate of fluent speech. Subjects were then asked to choose which picture correctly depicted the sentence they had just heard and to indicate their choice by pointing to the appropriate picture. The experimenter recorded the numerical code for their selection on the response form.

As noted, five practice items were presented first, during which time subjects were given feedback regarding accuracy, and errors were reviewed in order to ensure comprehension of the instructions and the task. Each of the practice items was repeated as often as necessary until the participant understood.

RESULTS

Data were tallied as to the number of correct items per subject for each of the four syntactic constructions (10 exemplars each), and the percent correct was calculated for each subject (Table II).

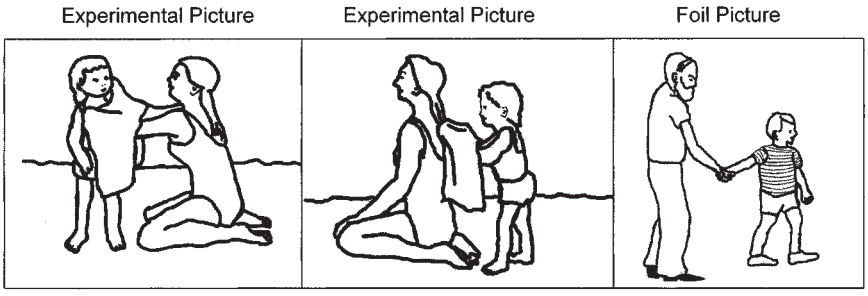

Fig. 1. Sample of experimental pictures.

Table II. Percent Correct for the SOAP

Subject	Aphasia Severity Rating	SOAP Comprehension			
		Act (%)	Pass (%)	SR (%)	OR (%)
Expressives					
TA	1	90	80	70	70
YY	1	90	20	40	50
LK	2	90	50	50	40
HR	1	70	50	60	40
TR	1	70	60	70	50
HN	2	100	70	50	70
LN	1	80	70	80	60
AH	1	70	50	70	70
NN	1	100	40	70	40
BJ	2	100	100	80	60
DR	1.5	100	40	100	30
EN	4	100	100	100	90
TH	4	90	40	80	30
YRY	3	100	90	100	100
TY	4	70	80	80	70
Receptive					
RL	2	60	40	50	50
HTN	2	60	30	70	40
HR	1	90	40	70	40
PH	2	60	20	30	50
TAN	1	80	70	70	20
CC	3.5	100	90	90	80
BuJ	1	70	70	60	40
LHD					
KY	5	100	100	100	80
RHD					
NE	4.5	100	100	90	90
EIN	5	100	100	100	90
NR	5	70	80	90	90
Unimpaired control					
BJ	N/A	100	100	100	100
CR	N/A	100	100	100	90
QC	N/A	100	100	100	100
SM	N/A	100	90	100	90

To better understand group patterns, the data were then averaged to get the percent correct for each of the four syntactic constructions for each patient group. As seen in Table III, the neurologically involved control participants (one LHD and three RHDs) performed quite well on the task, with scores ranging from 80% (on the hardest construction) to 100% (the easi-

Table III. SOAP Data Collapsed Across Subjects Within Each *(a Priori)* Behaviorally Classified Group for all Four Conditions

N		Act (%)	Pass (%)	SR (%)	OR (%)
15	Expressive	88	63	73	58
7	Receptive	74	51	63	46
1	LHD	100	100	100	80
3	RHD	90	93	93	90

est). The receptive aphasics demonstrated an overall weak performance across all constructions. The expressive aphasics demonstrated relatively preserved performance on the actives and subject-relative constructions, with weaker performance on the passives and object-relative constructions (these generalizations will be explored in depth below).

As noted above, expressive aphasics have been claimed to have particular difficulty when they encounter any type of noncanonical format, which would include both the passive and object-relative constructions. To better understand this canonical/noncanonical distinction across all the participant groups involved in this study, we collapsed the four data sets into two categories: canonical (active and subject-relative constructions, which retain the order of subject-verb-object) and noncanonical (passives and object relatives, which involve objects in nonstandard syntactic positions).

Canonical Versus Noncanonical

*Neurally Involved Controls (*N = *4) versus Older Age-Matched Unimpaired Controls (*N = *4)*

As shown in the data in Table IV below, the LHD and RHD patients combined showed no distinguishable patterns compared to the age-matched unimpaired control subjects (The following are two-tailed unpaired *t* test for each of the four conditions: actives, $t_6 = -1.0$, $p = .36$; passives, $t_6 = -0.45$,

Table IV. Collapsed Data Groups: Canonical (Actives and Subject Relatives) and Noncanonical (Passives and Object Relatives)

N		Canonical (%)	Noncanonical (%)
15	Expressive	81	60
7	Receptive	69	49
1	LHD	100	90
3	RHD	92	92
46	Young UNC	99	96
4	Age-matched UNC	100	96

$p = .67$; subject relatives, $t_6 = -1.7320$, $p = .13$; object relatives, $t_6 = -1.964$, $p = .1$)

Controls versus Left Hemisphere-Damaged

In contrast, both control groups (impaired and unimpaired) demonstrated a very different pattern from the aphasic populations. The expressive aphasics showed relatively preserved performance on the canonical exemplars (81% correct), but poor performance on the noncanonical items (60% correct). The receptive aphasic group showed an unreliable pattern, characteristic of the group's behavioral syndrome. This dissociation in performance is explored in depth below.

This overall pattern of enhanced performance on canonical versus noncanonical constructions held for all but one expressive aphasic, suggesting an inherent reliability in this behavioral effect for most subjects, regardless of general aphasia subtype classification.

Of interest was whether or not the patients' performance was better or worse than chance level for each of these groups. Because there were three pictures to choose from (semantically and syntactically correct, semantically correct but syntactically incorrect, and a semantically and syntactically unrelated foil), 33% would be considered chance performance. Performing statistical analysis on the data in this fashion would not tell the whole story, however, because 99.36% of the time the semantically unrelated foil did not lure participants. Among the control subjects (both neurologically involved and unimpaired), the semantically unrelated foil was never chosen as a response. Of the 15 expressive aphasics, the semantically unrelated foil was chosen 0.6% of the time. For the receptive aphasic group, the semantically unrelated foil was chosen 0.7% of the time. The low percentage of semantically irrelevant interpretations indicates that the aphasic participants clearly understood the semantic of the words (nouns) being spoken to them, despite varying aphasia severity levels. Of interest to the hypothesis under investigation is whether or not participants understood the *syntactic* constructions being presented. Thus a subanalysis of the data was performed, setting chance level at 50% (rather than at 33%). All the analyses presented for this section refer to Table V.

Results: SOAP Performance Versus Chance

Controls

*Unimpaired (Younger and Older) Controls (*N = *50).* Neurologically unimpaired younger college-age control participants demonstrated better than chance performance on both the canonical (99% correct) and noncanonical

Table V. Expressive Aphasics Separated into Two Categories, Severe (BDAE Scores <3) and Mild (BDAE Ratings ≥3) Along with Other Neurologically Involved Participants and Their Corresponding Group Averages on the SOAP

			Percent Correct	
N		BDAE auditory comprehension score (%)	Canonical (%)	Noncanonical (%)
11	Severe expressive	69	77*	55x
4	Mild expressive	87	90*	75*
7	Receptive	62	69*	49x
1	LHD	96	100*	90*
3	RHD	97	92*	92*
4	Age-matched UNC	–	100*	96*
46	Young UNC	–	99*	96*

* Statistically different from chance.
x Statistically equal to chance.

(96% correct) constructions ($t_{45} = 100.569$, $p < .0001$, $t_{45} = 42.641$, $p < .0001$, respectively). In addition, older age-matched controls also demonstrated better than chance performance on both constructions (canonical, 100% correct[5], noncanonical, 96% correct, $t_3 = 19.323$, $p = .0003$).

*Neurologically Involved Controls (*N = *4).* As seen with the unimpaired participants, the four neurologically involved subjects showed better than chance performance for both canonical (96% correct) and noncanonical (91% correct) constructions ($t_3 = 9.245$, $p = .0027$, $t_3 = 17.234$, $p = .0004$, respectively), revealing intact comprehension for these syntactic forms.

Left-Hemisphere Damaged Participants

*Expressive Aphasics (*N = *15).* Statistically, severe, expressive patients demonstrated better than chance performance on the canonical items (77% correct, $t_{10} = 8.195$, $p < .0001$), but their performance on the noncanonical items did not differ from chance (55% correct, $t_{10} = 1.028$, $p = .3280$). This pattern was shown for the mild expressive aphasics as well (canonical, 90% correct: $t_3 = 6.532$, $p = .0073$; noncanonical, 75% correct: $t_3 = 1.768$, $p = .1753$).

*Receptive Aphasics (*N = *7).* For the receptive aphasics, when looking at the group as a whole, we see a similar pattern of effect (canonical, 69% correct: $t_6 = 2.976$, $p < .03$; noncanonical, 49% correct: $t_6 = -0.216$, $p = .8361$).

[5] All four subjects on the canonical condition scored 100%. Statistical tests are not required, but clearly demonstrate performance significantly better than chance.

Table VI. Receptive Aphasics Separated into Two Categories, as a Group ($N = 7$) and Only the Severe (BDAE Scores <3) ($N = 6$) with Their Corresponding Group Averages on the SOAP

N		BDAE auditory comprehension score (%)	Canonical (%)	Noncanonical (%)
7	Receptive	62	69*	49x
6	Severe receptive	56	64*	42*

* Statistically different from chance.
x Statistically equal to chance.

However, as noted above, one patient (CC) had a BDAE severity rating of 3.5. If we remove that patient from the analysis (as a mild receptive aphasic) and look at the six remaining severe receptive aphasics (Table VI), the pattern remains the same for the canonical items (64% correct, $t_5 = 2.708$, $p = .04$), *but* it changes for the noncanonical items (42% correct), with the severe receptive aphasics performing significantly *worse* than chance ($t_5 = -2.423$, $p = .059$).

SUMMARY OF PERFORMANCE ON SOAP (VERSUS CHANCE)

Overall, along with the neurologically involved and unimpaired individuals, both the expressive and receptive aphasics reliably understand canonical constructions—those sentence constructions that retained canonical subject-verb-object order. This, in a short test, replicates the pattern of effect that has been demonstrated numerous times in the literature (Drai *et al.*, 2001; Hickok & Avrutin, 1995a, b; Zurif *et al.*, 1989). However, the severe receptive aphasics demonstrated below-chance performance on noncanonical constructions and both mild and severe expressive aphasics performed at chance levels. This will be explored in more detail later. These results provide evidence that the SOAP distinguishes groups with anterior and posterior lesions (regardless of severity level) and serves as a sensitive measure of syntactic complexity. The manner by which the SOAP provides additional information regarding the comprehension abilities of patient populations will be explored below.

Comparison of SOAP Performance and BDAE Performance

One issue to consider in the evaluation of the SOAP as a new contribution to aphasiological typologies is to examine whether it provides important information. By definition, patients with mild severity ratings do not

present with obvious language difficulties (of interest in this paper, no obvious comprehension difficulties). The high scores (mild deficits) obtained by these participants on the BDAE comprehension subtests support these findings. In our investigations, we have noticed that these high BDAE scores do not predict performance on the SOAP. Critically, what appears to play a more integral role is site of lesion (broadly, anterior versus posterior). Interestingly, whereas severity rating appears to relate to the BDAE, it in fact has no affect on the overall pattern of performance on the SOAP (i.e., anterior-lesioned patients, both mild and severe expressive aphasics, demonstrate similar patterns of performance). In the following discussion, we present evidence demonstrating that the SOAP can detect comprehension deficits in those patients labeled mild expressive aphasics who have performed "flawlessly" on the BDAE subtests.

Table VII presents a comparison of the participant group's overall BDAE auditory comprehension score (again, see Appendix 1 for individual patient score breakdown) and their corresponding canonical and noncanonical SOAP scores. Correlation analyses were performed for each of the groups to compare performance on the SOAP in the canonical and noncanonical conditions and the auditory comprehension score from the BDAE.

Of interest to the hypothesis under investigation is whether or not the BDAE auditory comprehension score (which is based on four subtests) and/or the severity rating correlate with the canonical and noncanonical conditions of the SOAP. By comparing performance on the group of subtests used to diagnose auditory comprehension in the BDAE with performance on canonical and noncanonical syntactic conditions in the SOAP, we can highlight the sensitivity and specificity of the SOAP as a tool for assessing syntactic comprehension. Upon segregating the aphasics into "severe" and "mild," we can better explore the influence of severity level on overall auditory comprehension, as well as the SOAP's sensitivity to distinguishing groups differing in general site of lesion (anterior/posterior).

Table VII. BDAE Auditory Comprehension Score and their Corresponding Canonical and Noncanonical SOAP Scores

N		BDAE auditory comprehension score (%)	Canonical (%)	Noncanonical (%)
11	Severe expressive	69	77	55
4	Mild expressive	87	90	75
7	Receptive	62	69	49
1	LHD	96	100	90
3	RHD	97	92	92

SEVERE EXPRESSIVE APHASICS (B.D.A.E. SEVERITY RATING <3)

As can be seen in Table VIII below, for the severe group of expressive aphasics, the SOAP's canonical and noncanonical conditions did not correlate [$r = 0.112$, $F(1,9) = 0.115$, $p = .742$]. In addition, although the BDAE auditory comprehension score correlated with their performance on the canonical condition of the SOAP [$r = 0.678$, $F(1,9) = 7.665$, $p = .02$; Fig. 2a], it did not correlate at all with the noncanonical condition [$r = 0.217$, $F(1,9) = 0.446$, $p = .52$; Fig. 2b].

MILD EXPRESSIVE APHASICS (SEVERITY RATING ≥3)

For the group of mild expressive aphasics (with severity level ratings higher than 3), performance on the BDAE auditory comprehension subtest correlated well with the SOAP's canonical condition ($r = 0.903$; Table IX). Because there were fewer participants in this group (low power for this particular analysis), the regression analysis, while clearly demonstrating clustering around the intercept (Fig. 3a), is not statistically significant [$F(1,3) = 8.858$, $p = .09$]. For the noncanonical condition, no reliable patterns were found [$r = 0.313$, $F(1,3) = 0.217$, $p = .387$; Fig. 3b]. This pattern of effect is strikingly similar to that found with the severe expressive aphasics, despite significant distinctions in oral communicative capacity and auditory comprehension scores (with average values for the BDAE subtests never falling below 75% accuracy).

ALL EXPRESSIVE APHASICS COMBINED (ANTERIOR-LESIONED PARTICIPANTS)

The basic pattern of the BDAE auditory subtests showed reliable predictions for performance on the SOAP canonical construction, but not for the noncanonical construction; therefore the two behavioral groups, severe

Table VIII. Correlation Matrix for 11 Severe Expressive Aphasics (BDAE Severity Rating <3) Comparing BDAE Auditory Comprehension Scores and SOAP Canonical and Noncanonical Scores

	SOAP canonical	SOAP Noncanonical
BDAE auditory comprehension	.678	.217
SOAP noncanonical	.112	–

Eleven observations were used in this computation.

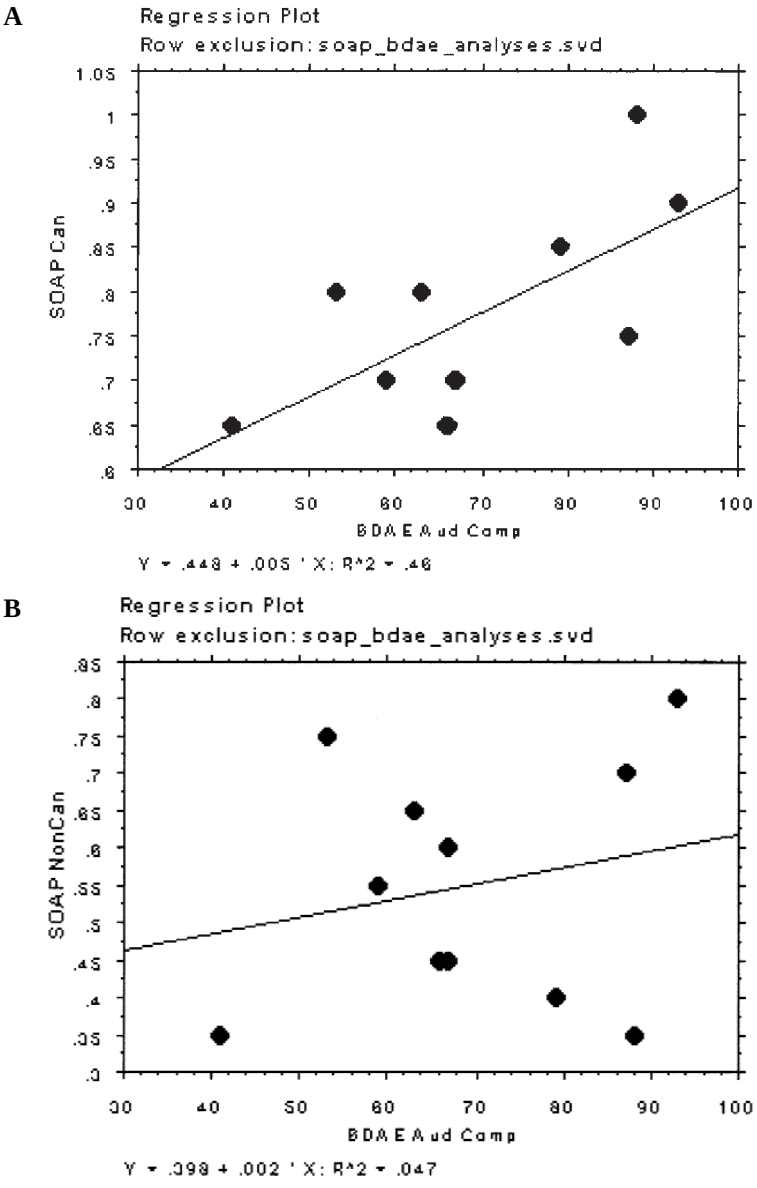

Fig. 2. (a) Simple regression plot for severe expressive aphasics comparing the BDAE's overall auditory comprehension score as percentage and the SOAP's canonical score (Note scaling on Y-axis: 1 = 100%) (b) Simple regression plot for severe expressive aphasics comparing the BDAE's overall auditory comprehension score as percentage and the SOAP's noncanonical score (Note scaling on Y-axis: 1 = 100%)

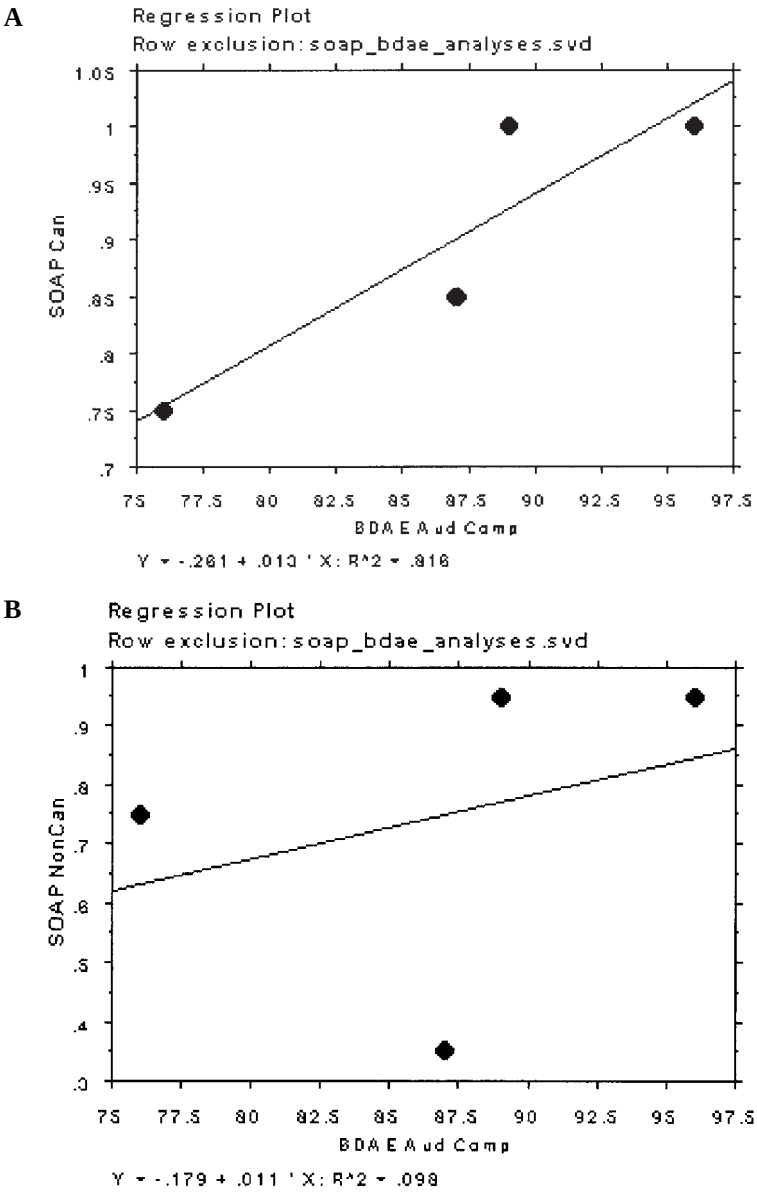

Fig. 3. (a) Simple regression plot for mild expressive aphasics (severity rating ≥3) comparing the BDAE's overall auditory comprehension score as percentage and the SOAP's canonical score (Note scaling on Y-axis: 1 = 100%) (b) Simple regression plot for mild expressive aphasics (severity rating ≥3) comparing the BDAE's overall auditory comprehension score as percentage and the SOAP's noncanonical score (Note scaling on Y-axis: 1 = 100%)

Table IX. Correlation Matrix for Four Mild Expressive Aphasics (BDAE Severity Rating ≥3) Comparing BDAE Auditory Comprehension Scores and SOAP Canonical and Noncanonical Scores

	SOAP canonical	SOAP noncanonical
BDAE auditory comprehension	.903	.313
SOAP noncanonical	.577	–

Four observations were used in this computation.

and mild expressive aphasics, were combined. This general group of expressive aphasics all had anterior lesions (with no involvement of the superior temporal sulcus and surrounding regions). When the two groups of expressive aphasics (mild and severe) were combined, the same pattern of effects emerged (Table X). Performance on the BDAE auditory comprehension subtest correlated well with the SOAP canonical condition [$r = 0.759$, $F(1,13) = 17.679$, $p = .0010$; Fig. 4a] but not with the noncanonical condition [$r = 0.381$, $F(1,13) = 2.210$, $p = .1609$; Fig. 4b]. These findings will be discussed further below.

RECEPTIVE APHASICS

All factors correlated relatively well for analyses involving performance of the seven receptive aphasics so that performance on both canonical and noncanonical conditions of the SOAP correlated [$r = 0.713$, $F(1,5) = 35.780$, $p = .0002$; Table XI]. In addition, performance on the BDAE auditory comprehension test also correlated reliably with the canonical condition [$r = 0.941$, $F(1,5) = 38.727$, $p = .0016$; Fig. 5a], but for the noncanonical condition the correlation was not statistically reliable [$r = 0.614$, $F(1,5) = 3.028$, $p = .1423$; Fig. 5b]. This nonreliable effect was probably due to the outlier from the one receptive aphasic who received a "mild" severity rating of 3.5 on the BDAE scale.

DISCUSSION

As a reminder, the first purpose outlined in this paper was to review the efficacy of the SOAP as a valid and sensitive measure of comprehension abilities in aphasic populations. The data presented above clearly demonstrate that the SOAP is, in fact, a valid measure of syntactic complexity for both canonical and noncanonical constructions.

Table X. Correlation Matrix for 15 Expressive Aphasics Comparing BDAE Auditory Comprehension Scores and SOAP Canonical and Noncanonical Scores

	SOAP canonical	SOAP noncanonical
BDAE auditory comprehension	.759	.381
SOAP noncanonical	.425	–

Fifteen observations were used in this computation.

It was first important to show that nonaphasic individuals (both brain-injured and non–brain-injured) all performed statistically better than chance on both canonical and noncanonical formats. Once that pattern of effect (and, thereby, the SOAP's specificity) was established, the data from the aphasic population could be examined more critically.

Of all the neurally involved participants tested, the control subjects (1 LHD and 3 RHD participants) all performed similarly to the age-matched unimpaired controls and college-age controls (we note that we had no *a priori* reason to expect deficits with these groups). Their SOAP scores on the canonical and noncanonical conditions were at or above 90% correct; their BDAE auditory comprehension scores were above 95%; and their BDAE auditory comprehension subtests were all above 80%.[6]

THE SENSITIVITY OF THE SOAP IN DISTINGUISHING POPULATIONS

The interesting comparisons begin to arise when we explore the sensitivity of the test to distinguishing aphasic individuals (and subgroups therein). In examining auditory comprehension, one issue is the efficacy of generalized assessment batteries in providing researchers a means of distinguishing one behavioral population from another. In this section, we address this issue by demonstrating the sensitivity of the SOAP in distinguishing among aphasia populations that would otherwise be grouped together. We presented data looking at the correlation between the BDAE's overall auditory comprehension scores and performance on the SOAP.

As stated above, neurologically involved control subjects demonstrated spared comprehension as shown by their performance on both the BDAE's auditory comprehension tests and both the SOAP's canonical and noncanonical constructions.

[6] One RHD patient, NE, was only run on two of the four BDAE subtests (word discrimination and complex ideational). It is noted, however, that the one of the subtests administered is one that is considerably difficult. This condition generally generates the lowest score achieved by most, if not all, aphasic patients.

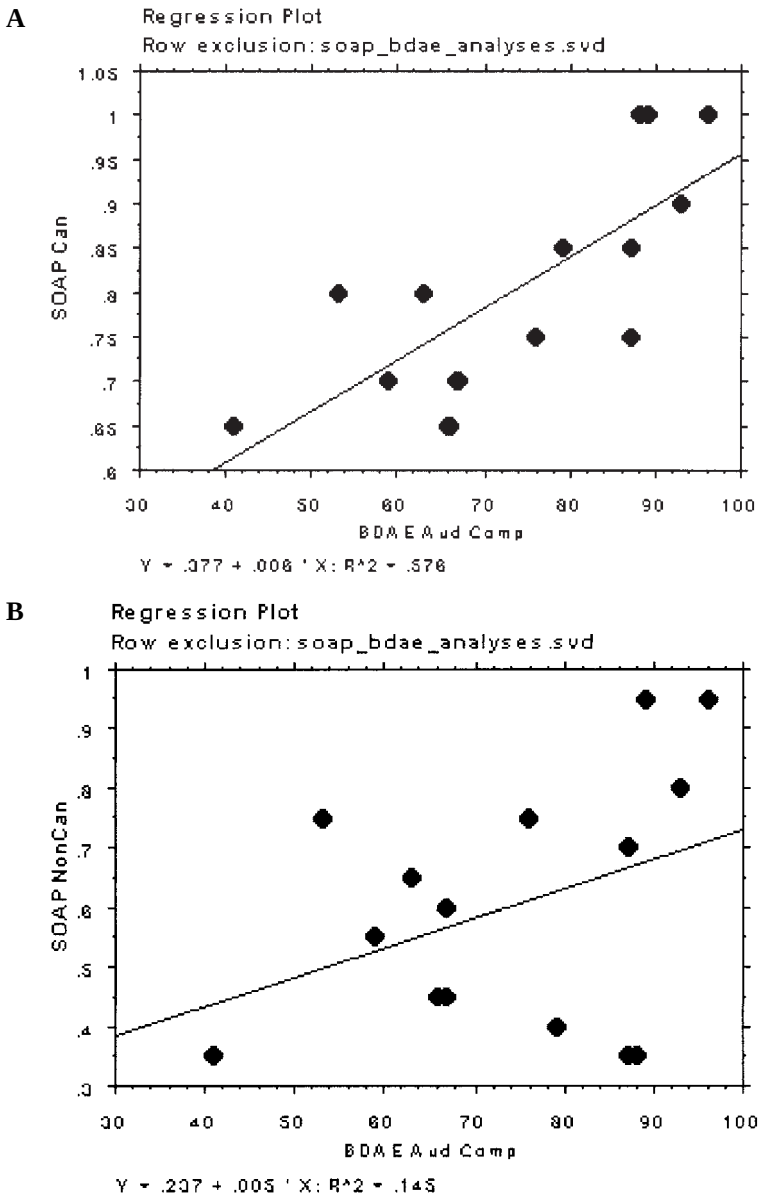

Fig. 4. (a) Simple regression plot for all expressive aphasics comparing the BDAE's overall auditory comprehension score as percentage and the SOAP's canonical score (Note scaling on Y-axis: 1 = 100%) (b) Simple regression plot for all expressive aphasics comparing the BDAE's overall auditory comprehension score as percentage and the SOAP's noncanonical score (Note scaling on Y-axis: 1 = 100%)

Table XI. Correlation Matrix for Seven Receptive Aphasics Comparing BDAE Auditory Comprehension Scores and SOAP Canonical and Noncanonical Scores

	SOAP canonical	SOAP Noncanonical
BDAE auditory comprehension	.941	.614
SOAP noncanonical	.713	–

Seven observations were used in this computation.

Receptive Aphasics

As one would expect with this population, this group demonstrated poor BDAE auditory comprehension performance. This pattern of performance was not indicative of the group's performance on the SOAP, however. Although the scores were by no means as good as those of their neurologically involved controls, the receptive aphasics showed some sparing of canonical constructions. The group's performance on the noncanonical exemplars was not as clear-cut, however. As a whole, the group showed chance performance, but when we apply the same criterion of severity level grouping, we end up excluding one higher-functioning participant. The noncanonical condition then drops to significantly worse than chance performance. This susceptibility reflects the stereotypical aberrant comprehension patterns that are central to the behavioral deficit. It is important to note that all patients in this behavioral classification have lesions involving aspects of the temporal lobe, with sparing of critical frontal operculum regions, those typically found to be involved with Broca's aphasia.

Expressive Aphasics

All patients in this category have damage to the frontoparietal region with sparing of the superior temporal sulcus and surrounding region. We subdivided this group to include those individuals who demonstrated severe behavioral deficits, those warranting a score on the BDAE's severity rating scale below 3 (severe expressive aphasics), and those who had more spared functioning, those who received a 3 or higher (mild expressive aphasics).

The severe expressive aphasics demonstrated intact comprehension for canonical constructions but not for the noncanonical exemplars. This pattern of effect has been demonstrated in the literature, and we have now shown that the SOAP reliably distinguishes this group's ability to understand syntactic forms that do not require more complex syntactic analyses (i.e., active and subject-relative constructions, both part of the canonical group). This effect correlates well with the BDAE's overall auditory comprehension score.

When this group is presented with sentence forms that have fronted or "moved" objects, their performance drops to chance level. This effect does

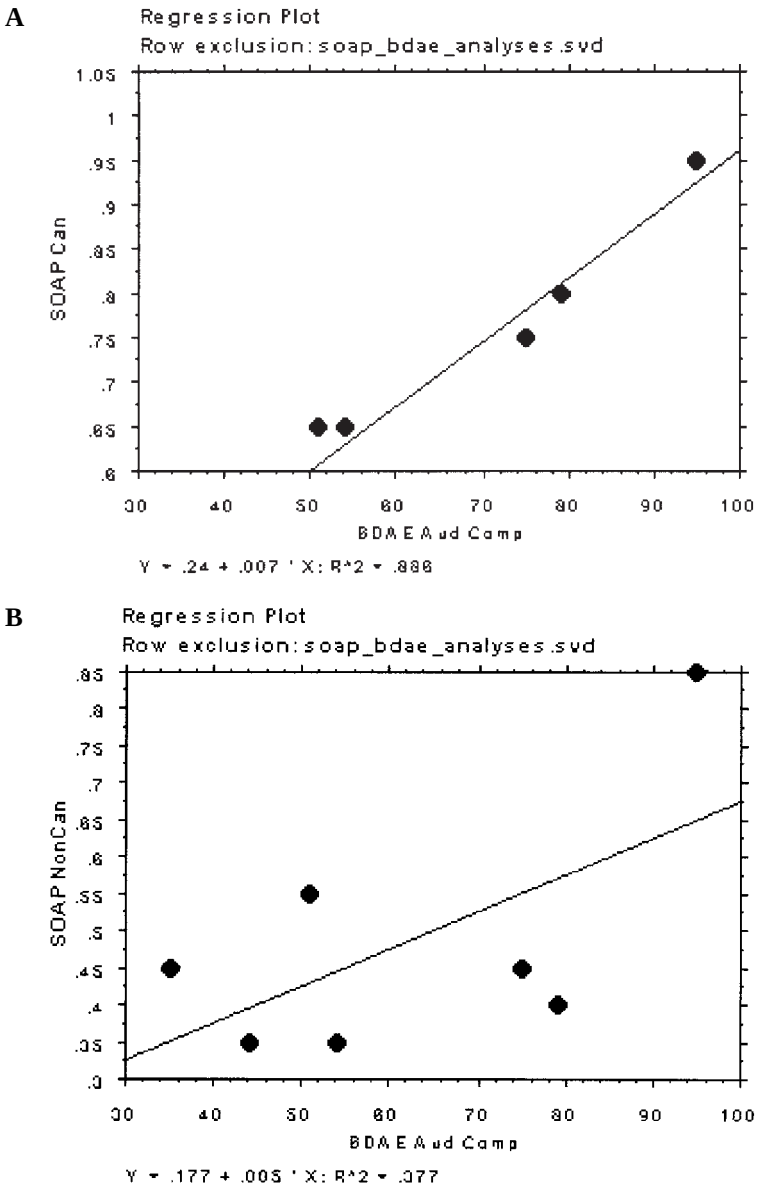

Fig. 5. (a) Simple regression plot for receptive aphasics comparing the BDAE's overall auditory comprehension score as percentage and the SOAP's canonical score (Note scaling on Y-axis: 1 = 100%) (b) Simple regression plot for severe expressive aphasics comparing the BDAE's overall auditory comprehension score as percentage and the SOAP's non-canonical score (Note scaling on Y-axis: .5 = 50%)

not correlate with the achieved scores from the BDAE's overall auditory comprehension test. In other words, the BDAE measure would not predict this groups poor performance on this subtest and therefore fails to identify an essential characteristic of the language deficit of this population.

For the mild expressive aphasics, those receiving a BDAE severity rating at or above 3, the same exact pattern of effect holds, something the BDAE auditory comprehension scores alone would not predict. Behaviorally, this group demonstrates relatively well-preserved language skills. Overall, they are fluent and have fairly intact comprehension skills that are clearly distinct from their more severe counterparts. With respect to lesion sites, this group also sustained damage to anterior regions of the left hemisphere with sparing of the temporal lobe. The SOAP is sensitive to this group's more subtle comprehension deficits and clearly provides additional information for diagnostic categorization that the BDAE auditory comprehension subtest did not tap into (as evident by showing little, if any, deficit).

FINAL THOUGHTS

It is clear that broad behavioral assessment batteries should not be the sole basis for grouping subject populations. The data presented above clearly show that a group that exhibits only mild behavioral manifestations of language impairment (the mild expressive group), as diagnosed by standardized assessment protocols, do, in fact, have difficulties in understanding complex syntactic constructions. The clinical diagnostic measures employed in this study alone would cause researchers to overlook such deficits. An important note is that in looking at the overall construction of the matched-length sentences, one cannot argue that the effects stem from complexity as defined by length, but, in fact, it is the syntactic complexity that drives this effect.

In speaking to the issue of lesion-based classification of patient populations, it has been demonstrated that the SOAP can aide researchers in teasing apart subtle auditory language deficits that more generalized tests do not detect. The SOAP is clearly sensitive to sites of lesion; both the mild and severe expressive aphasics (two behaviorally distinct groups with anterior lesions) have similar patterns of performance with respect to canonical and noncanonical constructions. We agree with the arguments made in recent publications suggesting that *a priori* categorization based on behavioral classification alone is inappropriate (Caplan, 2001). Because aphasiologists now have the benefit of neuroradiological findings early on in the diagnostic process, this information should be considered in conjunction with behavioral profiles when grouping subjects for research endeavors. As

the findings reported here suggest, the SOAP is sensitive to anterior/posterior neural distinctions and, moreover, has proven itself to be a valuable assessment tool that, when used in conjunction with broader assessment protocols, gives a more complete picture as to the processing abilities participants have for varying types of complex syntactic forms.

APPENDIX 1: BDAE SCORES ON AUDITORY COMPREHENSION SUBTESTS FOR BRAIN-INJURED PARTICIPANTS

BDAE Auditory Comprehension: Converted to Percent

Subject	Aphasia severity rating	Word discrimination	Body part identification	Commands	Complex ideational
Expressive					
TA	1	72	45	60	33
YY	1	38	30	47	50
LK	2	96	100	40	33
HR	1	83	65	47	67
TR	1	81	60	60	33
HN	2	87	90	80	92
LN	1	61	85	40	67
AH	1	90	68	67	42
NN	1	96	75	93	50
BJ	2	97	90	100	83
DR	1.5	97	80	100	75
EN	4	100	100	100	83
TH	4	100	95	93	58
YRY	3	96	98	93	67
TY	4	72	85	100	50
Receptive					
RL	2	76	50	7	8
HTN	2	82	60	40	33
HR	1	97	90	87	42
RH	2	79	63	27	8
TAN	1	85	90	67	58
CC	3.5	100	98	80	100
BuJ	1	70	50	33	50
LHD					
KY	5	100	100	100	83
RHD					
NE	4.5	93	N/A	N/A	100
EIN	5	97	90	100	100
NR	5	86	100	100	100

APPENDIX 2: BDAE SEVERITY RATING SCALE

Aphasia Severity Rating Scale

0. No usable speech or auditory comprehension.
1. All communication is through fragmentary expression; great need for inference, questioning, and guessing by the listener. The range of information that can be exchanged is limited, and the listener carries the burden of communication.
2. Conversation about familiar subjects is possible with help from the listener. There are frequent failures to convey the idea, but patient shares the burden of communication with the examiner.
3. The patient can discuss almost all everyday problems with little or no assistance. Reduction of speech and/or comprehension, however, makes conversation about certain material difficult or impossible.
4. Some obvious loss of fluency in speech or facility of comprehension, without significant limitation on ideas expressed or form of expression.
5. Minimal discemible speech handicaps; patient may have subjective difficulties that are not apparent to listener.

APPENDIX 3: SOAP STIMULI

Practice

1. The old man with the beard leads the little boy.
2. The woman in the bathing suit dries the child.
3. The teacher that scolds the student has black hair.
4. The nurse on the 5th floor examines the doctor.
5. The policeman that threatens the man has black hair.

Test

1. (P) The doctor with blonde hair is questioned by the soldier.
2. (SR) The man that pushes the boy is wearing a red shirt.
3. (A) The thin cowboy with the blue pants captures the Indian.
4. (A) The young boy with the brown hair grabs the man.
5. (OR) The doctor that the bedridden patient accuses has black hair.
6. (P) The boy with the brown hair is grabbed by the man.
7. (A) The little boy with the big book yells at the soldier.
8. (SR) The cowboy that captures the Indian has on blue pants.
9. (OR) The boy that the girl chases is wearing a green shirt.

10. (OR) The man that the young boy grabs has brown hair.
11. (SR) The man that records the young woman has brown hair.
12. (P) The boy in the green shirt is chased by the girl.
13. (A) The girl with blonde hair photographs the nurse.
14. (P) The man in the red shirt is pushed by the boy.
15. (A) The little boy in the blue shirt instructs the teacher.
16. (SR) The doctor that accuses the bedridden patient has black hair.
17. (OR) The teacher that the little boy instructs has on a blue shirt.
18. (SR) The girl that photographs the nurse with the camera is blonde.
19. (P) The cowboy with blue pants is captured by the Indian.
20. (OR) The soldier that the little boy yells at has black hair.
21. (A) The young doctor with blonde hair questions the soldier.
22. (SR) The man that grabs the little boy has brown hair.
23. (A) The man in the red shirt pushes the little boy.
24. (A) The man with brown hair records the young woman.
25. (SR) The soldier that yells at the small boy has black hair.
26. (A) The bedridden patient with the black hair accuses the doctor.
27. (P) The boy with black hair is yelled at by the soldier.
28. (SR) The soldier in uniform that questions the doctor has blonde hair.
29. (SR) The teacher that instructs the boy is wearing a blue shirt.
30. (OR) The girl that the nurse photographs with the camera is blonde.
31. (OR) The man that the young woman records has brown hair.
32. (P) The boy in the blue shirt is instructed by the teacher.
33. (OR) The cowboy that the Indian captures has on blue pants.
34. (OR) The man that the boy pushes is wearing a red shirt.
35. (P) The patient with black hair is accused by the doctor.
36. (OR) The soldier in uniform that the doctor questions has blonde hair.
37. (A) The girl in the green shirt chases the small boy.
38. (P) The man with brown hair is recorded by the woman.
39. (P) The girl is photographed by the nurse with blonde hair.
40. (SR) The boy chases the girl is wearing a green shirt.

REFERENCES

Caplan, D. (2001). The measurement of chance performance in aphasia, with specific reference to the comprehension of semantically reversible passive sentences: A note of issues raised by Caramazza, Capitani, Rey and Berndt (2001) and Drai, Grodzinsky and Zurif (2001). *Brain and Language, 76,* 193–201.

Drai, D., Grodzinsky, Y., & Zurif, E. (2001). Broca's aphasia is associated with a single pattern of comprehension performance: A reply. *Brain and Language, 76,* 185–192.

Goodglass, H., & Kaplan, E. (1972). *The assessment of aphasia and related disorders.* Philadelphia: Lea and Febiger.

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders* (2nd ed.). Philadelphia: Lea and Febiger.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The assessment of aphasia and related disorders* (3rd ed.). Philadelphia: Lippincott Williams & Wilkins.

Hickok, G., & Avrutin, S. (1995a). Comprehension of Wh-questions in two Broca's aphasics. *Brain and Language, 50,* 10–26.

Hickok, G., & Avrutin, S. (1995b). Representation, referentiality, and processing in agrammatic comprehension: Two cases studies. *Brain and Language, 50,* 10–26.

Kertesz, A. (1982) The Western Aphasia Battery. New York: Grune & Stratton.

Porch, B. E., (1967). *Porch index of communicative ability: Theory and development.* Chicago, IL: Riverside Publishing Company

Schuell, H. (1965). *The Minnesota Test for Differential Diagnosis of Aphasia.* Minneapolis: University of Minnesota Press.

Zurif, E., Gardner, H., & Brownell, H. (1989). The case against the case against agrammatism. *Brain and Cognition, 10,* 237–255.