Cross linguistic aphasia testing: The Portuguese version of the Aachen Aphasia Test (AAT)

MARTIN LAUTERBACH,[1] ISABEL PAVÃO MARTINS,[1] PAULA GARCIA,[2] JOANA CABEÇA,[2]
ANA CRISTINA FERREIRA,[2] AND KLAUS WILLMES[3]

[1]Language Research Laboratory (LEL, Laboratório de Estudos de Linguagem), Institute of Molecular Medicine
 (IMM, Instituto de Medicina Molecular), Faculty of Medicine, Lisbon, Portugal
[2]High School for Health Sciences (ESSA, Escola Superior de Saúde), Alcoitão, Portugal
[3]Section Neuropsychology, Neurological Clinic, RWTH Aachen University, Aachen, Germany

Abstract

We report the adaptation of the Aachen Aphasia Test (AAT) to the Portuguese language (PAAT) and the results of its standardization in 125 persons with aphasia and 153 healthy controls. Patients with aphasia had a previous syndromic diagnosis, obtained through a Portuguese aphasia battery, which served as a reference. The control group was stratified by age and educational level. Hierarchical cluster analyses showed good construct validity. The increasing degree of difficulty and complexity throughout the item sets comprising subtests was confirmed. The discriminatory power of the PAAT for the selection of aphasic from non-aphasic persons proved to be as high as for the AAT versions in other languages. Classification of standard aphasic syndromes by means of discriminant analyses was good. Internal consistency, measured by means of Cronbach's alpha coefficient, was high to very high for the different PAAT subtests. Performance differences caused by age or educational level among the healthy control persons emphasized the need for correction factors. In conclusion, the PAAT showed robust psychometrical properties, comparable to the original German and to adaptations to other languages. It constitutes a useful tool for cross-linguistic and multicenter studies. (*JINS*, 2008, *14*, 1046–1056.)

Keywords: Neuropsychological test, Language, Multilingualism, Portugal, Age factors, Educational status

INTRODUCTION

The Aachen Aphasia Test (AAT) is an assessment tool for the diagnosis, classification, and quantification of aphasic disorders that was developed, validated, and standardized at the RWTH Aachen University (Huber et al., 1983, 1984) for the German language. The subtests were originally designed according to well defined linguistic criteria, varying in a systematic manner the modalities, linguistic units, and regularities of the German language. Its psychometric properties were extensively tested and showed high internal consistency, good construct validity, and re-test reliability (Poeck & Göddenhenrich, 1988; Willmes, 1985, 1990; Willmes et al., 1983). The discriminatory power of the test also proved to be high, as the AAT selects very well aphasic patients out of a population of brain damaged persons and probabilistically assigns patients to different standard aphasic syndrome groups. More interestingly, the performance profile can be analyzed with inferential statistical procedures in the individual case using methods of psychometric single case analysis (Huber, 1973; see also Willmes, 1985, 2003). This approach allows detecting statistically reliable differences in subtest performance as well as global and differential changes among subtests between two test occasions. In addition, the AAT provides information about the severity of impairments with reference to a standard syndrome. The test profile also provides clues for therapy planning.

Since the publication of the original version in 1983, the AAT has been adapted to English, Dutch, Italian, and Thai (Graetz et al., 1991, 1992; Luzzatti et al., 1987, 1991; Miller et al., 1998, 2000; Pracharitpukdee et al., 1998). Therefore, at present, the AAT is probably the most widely used aphasia battery in Europe.

Correspondence and reprint requests to: Martin Lauterbach, Laboratório de Estudos de Linguagem (LEL), UNIC, Instituto de Medicina Molecular (IMM), Hospital Santa Maria, 1649-028 Lisboa, Portugal. E-mail: mlauterbach@fm.ul.pt

1046

Cross-linguistic research of monolingual subjects proved to be an appropriate tool to investigate some of the universal mechanism of language. In order to study the neuronal bases for language, cross-linguistic research has followed two main approaches. The first takes language as a between-subjects variable and applies the same test material to speakers of different languages. The second approach looks at a single language, in order to describe its particularities. Cross-linguistic research contributed to the reformulation of certain aspects of psycholinguistic theory, (e.g., the concept of agrammatism) by analyzing data about languages other than English. It also emphasized the impact that the structure of a particular language imposes on its respective processing (Bates et al., 1991). Cross-linguistic research in aphasia revealed that brain lesions can cause quite uniform and reliable symptoms across languages, especially concerning how aphasic patients use morphological information (Benedet et al., 1998; Frederici et al., 1991). On the other hand, this research also demonstrated that the performance and competence of aphasic patients are clearly shaped by language specific properties (Bates et al., 1987). Many language specific patterns are retained under conditions of aphasia.

Yet, one major concern of cross-linguistic research in aphasia is to assure the comparability of patient groups. A mismatch of patient groups may mask cross-linguistic differences or exaggerate aphasic syndrome group similarities. Given the lack of a reliable procedure, group definitions in most studies were carried out on clinical-behavioral grounds (Bates et al., 1987; Benedet et al., 1998). Bates and colleagues dismissed a purely test-score based matching procedure. Instead, they proposed a procedure based on percentile matching (Bates et al., 1991). This requires sufficiently large patient databases in each country to establish percentile ranking of individual patients, a prerequisite that is not always available. Frederici and colleagues effectively used the classification results from the German and the Dutch version of the Aachen Aphasia Test to match patients, native of the respective language (Frederici et al., 1991). In Portugal, aphasic disorders have been evaluated for the last 30 years mostly by means of an adaptation of the Multilingual Aphasia Examination (Benton, 1969, 1989) labeled BAAL: Bateria de Avaliação de Afasia de Lisboa (Castro-Caldas, 1979; Damásio, 1973) The BAAL proved to have similar psychometric properties compared to the Western Aphasia Battery (Ferro & Kertesz, 1987). Yet, concerning the reliability of numerical classification systems, these two batteries are not sufficiently similar for an effective cross-linguistic comparison.

Taking into account the linguistic variability between languages, the adaptations of the AAT to other languages attempted to include comparable linguistic properties of the respective language, instead of a mere lexical translation of the test material. The aim of this article is to describe the construction of the Portuguese version of the AAT, its validation and standardization. The major criterion for a successful Portuguese version will be a preservation of the psychometric properties of the original and already translated versions.

METHOD AND MATERIALS

Structure of the PAAT

The Portuguese version of the AAT (PAAT for short) has the same general set-up as its original version (Table 1). The PAAT consists of a semi-structured interview after which spontaneous speech properties are rated, the Token Test, and four subtests, which vary linguistic modalities between subtests and linguistic units and regularities between individual item sets. The regularities of the Portuguese language were taken into account, in order to create a difficulty gradient for the chosen elements. Each subtest consists of 3 to 5 item sets, giving 21 sets all together. Each item set is composed of 10 items.

Adaptation into Portuguese

Because of linguistic differences between the German and the Portuguese language it was not possible to simply translate the original AAT material literally. To preserve the underlying rationale for the construction of the test, the Italian version (Luzzatti et al., 1987, 1991) was of great help, because Portuguese and Italian are both Romance languages with a similar grammatical structure.

To maintain the internal structure of the AAT described earlier, all items from the subtests Repetition and Written Language and most of the items from the subtests Naming and Comprehension had to be replaced. For the respective subtests that use words as stimuli, the Portuguese words were chosen in order to introduce increasing complexity, decreasing frequency and increasing articulatory and lexical difficulty for native speakers of Portuguese within each item set (e.g., in the Repetition subtest of the original version), words of increasing length with regular CVCV structure represent a particular difficulty for speakers of German, (e.g. *Hepatitis*), but this structure represents the default case in Portuguese vocabulary, as in Romance languages in general. On the other hand, consonant clusters pose a specific articulatory challenge for Portuguese (e.g., Volkswagen), which is a highly familiar sequence for native Germans.

Some of the line drawings used in the subtests Naming and Comprehension had to be substituted according to the constructional principles of the AAT mentioned above and the properties of the Portuguese language. Whereas in German the formation of the compound word dishwasher is effected by concatenation (Geschirrspülmaschine), this word is formed in Portuguese by means of a constituent phrase (máquina de lavar loiça). The first five items in the subtest Naming item set "compound words" are made up by following the sequencing principle. Some compound words in Portuguese are composed of a noun with an inflected verb form e.g. saca-rolha (corkscrew), from sacar a rolha (to pull the cork). For the Portuguese version of the AAT the items

Table 1. Composition of the PAAT

	Test part/Subtest		Description of Item set/Rating scale
1	Spontaneous language rating (6-point rating scales)	1	communicative behavior
		2	articulation and prosody
		3	automatized language
		4	semantic structure
		5	phonemic structure
		6	syntactic structure
2	Token Test (TT)	1–5	5 item sets
3	Repetition (REP)	1	single phonemes
		2	one–three syllabic words
		3	loan and foreign words
		4	morphologically complex words
		5	phrases of increasing length
4	Written Language (LE)	1	reading aloud words/phrases
		2	composing words/phrases from grapheme/morphemes
		3	writing words/phrases to dictation
5	Naming (NOM)	1	pictured objects (single nouns)
		2	colors
		3	pictured objects (compound nouns)
		4	pictured situations and actions
6	Comprehension (COM) (4-choice items)	1	auditory words
		2	auditory sentences
		3	reading words
		4	reading sentences

6–10 in the item set naming of compound words use such verb-noun combinations. This kind of words poses a particular difficulty for patients with Broca's aphasia, who tend to omit the verb component of the word. This kind of construction has already been used in the Italian version and allowed particularly well for the detection of naming difficulties for verbs.

The item sets that use phrases as stimuli (item set 5 of the subtest Repetition; item set 4 of the subtest Comprehension) were adopted from the Italian version of the AAT with almost no changes, given that Romance languages have nearly the same syntactic structure. For the fourth set of the subtest Naming, all drawings of the Italian version could be adopted, given that the cultural differences between Northern and Southern Europe had already been considered for the Italian version (e.g., changing Father and son are playing Cowboys and Indians to Father and son are playing soccer).

Participants

All data concerning the patient and the control sample were collected according to the regulations of the Helsinki Declaration, with written consent obtained from the participant or his/her family.

A sample of $n = 125$ persons with aphasia and $n = 153$ healthy subjects was examined with the PAAT. A summary of demographic data for the aphasic and the healthy control sample is provided in Table 2.

Data for the control sample was obtained from healthy urban populations of Lisbon metropolitan area and from rural populations of northern and southern Portugal. All participants older than 50 years were required to have a normal score in the Mini Mental State Examination (MMSE; Folstein et al., 1975), in order to exclude dementia.

In order to obtain normative data sensitive to age and educational level, the sample of healthy control participants was stratified by age and years of education. All illiterate persons were older than 65 years.

Controls were—on average—13 years younger than aphasic patients ($t(276) = 6.74$; $p < .001$). On average controls attended school for 2.6 additional years as compared to aphasic patients ($t(276) = -4.02$; $p < .001$). Post hoc testing after a significant ANOVA result revealed that only the group of global aphasic patients differed significantly from

Table 2. Demographic data of aphasic patients and control subjects

Group	n	Gender (n)		Age (yrs)		Schooling (yrs)	
		male	female	mean	SD	mean	SD
Aphasic	125	79	46	57.7	13.3	7.0	5.0
Control	153	58	95	44.5	19.6	9.6	5.5

the control group in terms of education (Tamhane post hoc test: mean difference 4.18; $p = .001$). The distribution of gender was not balanced between the patient and the control group. Whereas the patient group was composed of 63.2% men and 36.8% women, the opposite pattern was found for the control group, 38.6% men and 61.4% women. This difference was significant according to Pearson's $\chi^2(1) = 16.70$, $p < .001$.

In the control group there was a significant negative correlation between "age" and "schooling" ($r = -.507$; $p < .001$). Gender difference regarding age or schooling were not observed.

The assessment of the patients took place at least four weeks after onset of aphasia, in order to exclude more unstable and atypical aphasia symptomatology during the acute phase.

Etiology was mainly vascular: 77 patients suffered an ischaemic stroke and 23 a brain hemorrhage stroke. The remaining 25 patients had other etiologies such as traumatic brain injuries, tumors or arteriovenous malformations.

All patients with aphasia were also examined with the BAAL. Syndrome classification of the BAAL was taken as a reference, because this battery has been used during the last 30 years to carry out the classification according to aphasic syndromes in Portugal. The BAAL uses numerical criteria (non-overlapping score ranges) for the four subtests Fluency, Naming, Repetition and Comprehension to classify aphasic patients unequivocally into one of eight aphasic syndromes: the 4 standard syndromes of Broca, Wernicke, global, and amnesic aphasia, the 3 transcortical subtypes (sensory, motor, and mixed), and conduction aphasia. If a clinically aphasic patient scores too high so that he/she cannot be classified anymore into one of the eight syndrome categories, a residual aphasia is diagnosed.

For the purpose of the standardization of the PAAT we assigned each patient to one of the four standard syndromes or one of the four additional "non-standard" syndromes that the BAAL taxonomic system provides or to the category residual aphasia. The non-standard and residual aphasias were merged to form the "other"-group ($n = 24$). The results of the BAAL syndrome assignment are shown in Table 3.

The demographic variables age, schooling, and time since CVA of the five patient groups were compared by means of ANOVA. No statistically significant differences were found, except for the variable "age" ($F(4, 120) = 3.55$; $p = .009$). Post-hoc comparisons revealed that the only statistically significant pairwise group comparison was between the groups of non standard aphasia (other) and *Wernicke aphasia* (because of heterogeneous variances, Tamhane's post hoc test procedure was employed: mean difference 12.78; $p = .015$). On average Wernicke's aphasia patients were older than the patients that suffered from a non-standard aphasia. Gender differences for the patient groups were present for the variable "schooling". On average, women had less years of education than men (female: 5.8 years; male: 7.8 years; $t(123) = 2.21$, $p = .029$); this holds true only for the overall analysis. Gender comparisons by means of t-tests for the single diagnostic groups did not reveal any significant effect of age, schooling, or time since CVA.

Testing

The test was applied according to the instructions in the user's manual of the original version of the AAT (Huber et al., 1983). All aphasic patients were tested with the PAAT by one of the authors (ML), except for a sub-sample of 39 patients that were assessed by ACF, an experienced speech therapist with a specific training on administering the PAAT.

The BAAL testing, used for syndrome assignment of the patient group, was performed by experienced speech therapists from the centers that participated in the study. All aphasia testing was carried out individually in rooms dedicated to speech therapy.

The control group data were collected by PG and JC, two students of speech therapy in their final year. They were specifically trained in the application of the PAAT. Testing took place at the participants' homes.

The item scoring of all tests was checked by the first author. Equivocal scorings were discussed and changed whenever necessary, using the recordings of the test session.

Table 3. Demographic data and syndrome assignment results of the patient group

| Group | n | Gender (n) | | Age (years) | | Schooling (years) | | Time since CVA (month) | |
		male	female	mean	SD	mean	SD	mean	SD
All aphasic pts.	125	79	46	57.7	13.3	7.0	5.0	17.6	25.5
Only standard aphasic syndromes	101	63	38	59.4	12.6	6.6	4.6	17.9	26.3
Global aphasia	25	17	8	59.2	11.4	5.4	4.1	14.6	19.7
Wernicke's aphasia	28	20	8	64.1	12.3	6.5	4.8	13.2	17.2
Broca's aphasia	17	5	12	58.9	15.3	5.9	4.7	13.2	15.1
Amnesic aphasia	31	21	10	55.7	11.2	8.0	4.7	27.5	38.4
Others	24	16	8	51.3	14.6	8.8	6.1	16.4	22.1

RESULTS

Statistical analyses were carried out using the SPSS program package.

Internal Consistency

In order to examine internal consistency, Cronbach's α coefficient was calculated for all item sets and subtests for the 125 aphasic persons ($n = 109$ of them were literate). With the exception of item set 5 of the Token Test and the four item sets of the Comprehension subtest, all item sets showed a high consistency (Cronbach's α coefficient > 0.9).

Table 4 shows the consistency coefficients for the different subtests and their item sets.

Validity aspects of the PAAT

Results reported are based on different sample sizes, depending on whether only standard syndromes were considered or reading skills were necessary to perform subtests. In the latter case, illiterate subjects are not included. For each analysis, the number of subjects entering in the statistical analyses is indicated separately.

A complete linkage hierarchical cluster analysis of the 21 item sets (Fig. 1) included in the PAAT (from the 5 subtests with 3 to 5 item sets of 10 items each) and comprising all 88 literate aphasic patients with a standard syndrome was performed. Similarity among pairs of item sets was computed using the Euclidean distance measure. According to the complete linkage agglomeration criterion an item set is joined with an existing cluster only if its similarity to all elements of the cluster is larger than to any other item set in any other cluster. Data were standardized before computation of Euclidean distances since the score ranges for item sets of the Token Test are different from score ranges of all other item sets. The item sets basically cluster according to subtests. Only the item set "reading aloud" does not go together with the other two item sets of the subtest written language. Reading aloud is closer to item set 2 Naming of colors from the subtest Naming.

Table 4. Reliability information for the PAAT ($n = 125$ aphasic persons, $**n = 109$ literate aphasic persons)

Subtest item set	Cronbach's alpha	*SD*	Standard Error	95%-Confidence Interval for true score	Critical Difference* (5%, two-sided)
Token Test	0.972	14.45	2.42	±3.98	5.63
TT01	0.911	3.59	1.07	±1.76	2.49
TT02	0.894	3.56	1.16	±1.91	2.70
TT03	0.905	3.41	1.05	±1.73	2.45
TT04	0.910	2.99	0.90	±1.48	2.09
TT05	0.857	2.4	0.91	±1.50	2.11
Repetition	0.990	46.54	4.65	±7.66	10.83
REP01	0.946	8.7	2.02	±3.33	4.70
REP02	0.967	10.02	1.82	±2.99	4.23
REP03	0.968	10.61	1.90	±3.12	4.42
REP04	0.970	10.66	1.85	±3.04	4.30
REP05	0.966	9.49	1.75	±2.88	4.07
Written Lang.**	0.986	31.52	3.73	±6.14	8.68
LE01**	0.981	11.34	1.56	±2.57	3.64
LE02**	0.970	11.06	1.92	±3.15	4.46
LE03**	0.975	11.45	1.81	±2.98	4.21
Naming	0.986	40	4.73	±7.79	11.01
NOM01	0.953	11.44	2.48	±4.08	5.77
NOM02	0.956	11.04	2.32	±3.81	5.39
NOM03	0.938	9.38	2.34	±3.84	5.43
NOM04	0.973	10.1	1.66	±2.73	3.86
Comprehension	0.949	26.43	5.97	±9.82	13.89
COM01	0.814	6.25	2.70	±4.43	6.27
COM02	0.794	7.17	3.25	±5.35	7.57
C0M03**	0.884	8.16	2.78	±4.57	6.47
COM04**	0.882	8.44	2.90	±4.77	6.74
COMauditory	0.884	12.52	4.26	±7.01	9.92
COMlreading**	0.933	15.8	4.10	±6.73	9.51
COMwords**	0.903	13.07	4.07	±6.70	9.47
COMsentences**	0.907	14.48	4.42	±7.26	10.27

*Critical difference for comparison of two test administrations of the same subtest/item set when retesting the same patient.

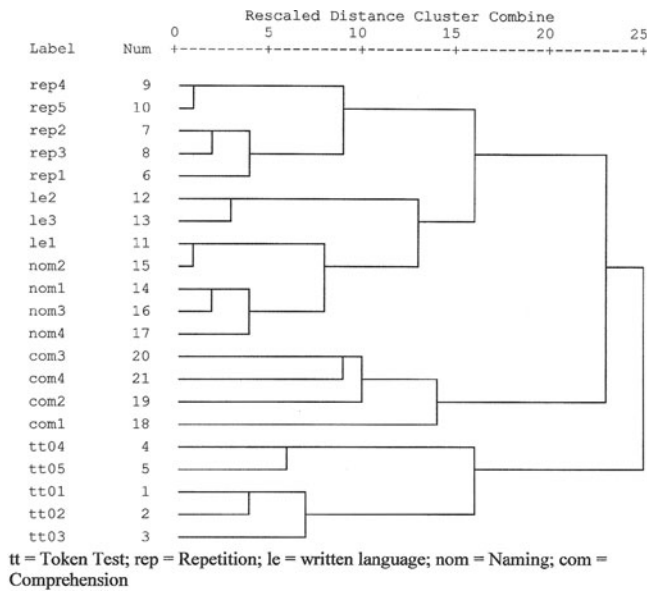tt = Token Test; rep = Repetition; le = written language; nom = Naming; com = Comprehension

Fig. 1. Complete linkage hierarchical cluster analysis of $n = 88$ literate aphasic patients with standard syndromes from the SPSS output (agglomeration criterion: complete linkage; similarity measure: Euclidean distance).

Table 5 shows descriptive statistics for the aphasic patient group and the control group for the different subtests and their respective item sets. In general, performance is lower, the higher the test number within an item set. This reflects the constructional idea of increasing difficulty across the subsets of items. The mean scores for the different subtests between the different diagnostic groups and the control group vary in the expected order from global aphasia with the worst performance to amnesic aphasia with the best performance among the patient groups. Mean item difficulty of the subtest written language does not vary systematically between the sets of items. Items had been chosen in order to be parallel with respect to phonological structure, word length and frequency.

Mean performance of healthy controls is shown in the last row of Table 5. The maximum score for all subset of items is 30 points, except for the Token Test, which has 10 points per item set only.

Selection Aphasia/No Aphasia

To evaluate the differential validity of the PAAT we first performed a discriminant analysis for all literate subjects,

Table 5. Mean scores and standard deviations for each diagnostic group and controls; number of participants; number of literate subjects in parenthesis

Item set	All aphasics $n = 125$ (109)		Global $n = 25$ (20)		Wernicke $n = 28$ (24)		Broca $n = 17$ (14)		Amnestic $n = 31$ (30)		Others $n = 24$ (21)		Controls $n = 153$ (139)	
	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
TT01	6.32	3.59	2.28	2.37	4.39	2.73	8.29	2.17	9.03	1.70	7.88	3.21	9.92	0.33
TT02	4.76	3.59	1.12	1.69	2.39	2.11	6.88	2.83	7.39	2.40	6.42	3.22	9.82	0.55
TT03	3.22	3.41	0.28	0.68	0.89	1.42	4.18	3.00	5.65	2.77	5.21	3.83	9.67	0.81
TT04	2.02	2.99	0.00	0.00	0.21	0.63	2.35	3.00	3.55	3.03	4.00	3.71	9.18	1.31
TT05	1.63	2.40	0.00	0.00	0.25	0.97	1.53	1.94	3.26	2.52	2.92	2.93	8.37	1.86
REP01	21.74	8.70	13.92	9.26	19.39	8.24	18.00	8.94	28.26	1.71	26.83	3.42	29.76	0.93
REP02	21.53	10.02	13.08	10.01	17.57	10.26	16.65	9.61	29.35	0.91	28.29	1.99	29.97	0.18
REP03	18.51	10.61	9.24	8.21	12.64	10.27	13.71	8.53	27.97	2.40	26.21	4.40	29.65	0.25
REP04	15.39	10.66	6.12	6.06	8.64	9.09	9.88	7.27	25.77	3.98	23.42	6.03	29.96	0.25
REP05	12.06	9.49	2.48	3.29	6.54	7.21	7.94	6.21	21.39	4.01	19.33	6.56	29.24	1.50
LE01	17.65	10.62	2.20	4.15	16.92	10.02	13.43	7.58	25.90	1.90	24.24	5.98	27.97	0.20
LE02	14.57	11.06	1.35	2.58	10.46	8.64	12.36	8.87	24.00	6.45	19.86	9.80	29.32	1.38
LE03	13.06	11.45	1.15	2.94	8.50	9.37	7.79	9.17	23.97	6.39	17.52	9.96	29.14	1.46
NOM01	17.02	11.44	4.24	6.52	13.14	9.71	16.65	11.17	26.87	3.96	22.38	9.28	29.24	1.50
NOM02	17.14	11.04	3.20	5.39	15.25	10.59	16.47	9.66	26.26	3.16	22.58	7.66	29.10	1.81
NOM03	11.12	9.38	1.44	3.18	6.71	6.45	9.65	8.19	20.23	5.56	15.50	8.17	27.40	3.21
NOM04	12.66	10.10	1.24	2.74	9.18	6.55	9.24	8.12	22.42	5.21	18.46	9.07	27.52	2.72
COM01	22.81	6.25	18.12	7.30	20.93	6.36	23.18	4.96	25.84	3.16	25.71	5.24	28.85	2.04
COM02	18.94	7.17	12.72	6.59	15.93	6.08	21.71	5.25	22.97	4.66	21.79	7.17	27.98	2.76
COM03	17.72	7.34	9.70	7.32	17.25	6.35	18.29	5.76	21.97	3.31	19.48	7.83	25.36	2.04
COM04	15.88	8.47	6.20	6.81	12.88	6.35	15.79	6.87	21.90	4.11	20.00	8.07	27.90	2.68

Note. TT = Token Test; REP = Repetition; LE = written language; NOM = Naming; COM = Comprehension.

Table 6. Aphasia selection properties of the PAAT.
Discriminant analysis for patients and controls, comprising only literate subjects ($n = 247$). Numbers in parentheses report doubtful classifications with a probability of allocation below 0.8. Discriminant analysis is based on the PAAT subtests Token Test, Repetition, Naming, Written Language and Comprehension.

		Group Membership	Predicted Group Membership (PAAT)		Total
			aphasia	no aphasia	
Cross-validated	Count	aphasia	86 (8)	23 (8)	109
		no aphasia	0	138 (2)	138
	%	aphasia	78,9	21,1	100,0
		no aphasia	,0	100,0	100,0

patients, and healthy controls, in order to describe the efficiency of the test in distinguishing aphasia. In the analysis only the total scores of the five subtests were entered, since evaluation of spontaneous speech was not performed for healthy persons. The majority (90.7%) of the 247 cases were correctly classified (Table 6). The classification results of this analysis are shown in Table 6.[1] Entering the variables in a stepwise fashion, the concordant classifications reached 91.9%. In this approach, the Token Test was the variable that provided the best discriminatory result for the separation of subjects with and without aphasia. None of the other subtests provided additional information to improve this distinction significantly.

Syndrome Assignment

Further discriminant analyses were performed, in order to test whether the original syndrome classification, established by the BAAL, would be concordant with the assignment based on the PAAT performance profile. Table 7 shows the classification results for the discriminant analysis of 88 literate aphasic subjects, originally diagnosed as belonging to a standard aphasia syndrome. In the analysis, all five subtests and the six spontaneous speech ratings were considered for classification. This resulted in 89.9% of the cases being classified in accordance with the initial diagnosis based on the BAAL. Entering the variables in a stepwise fashion, classification accuracy remains at 89.9%. Using the stepwise method, the seven variables that provided the best discriminatory power were the following: spontaneous speech ratings of articulation and prosody, communicative behavior, Token Test, spontaneous speech rating of automatized speech, repetition, spontaneous speech rating of syntactic structure, and of semantic structure. The remaining variables did not improve syndrome classification significantly.

---

[1]Regarding Tables 6 and 7. A substitute of proper cross validation is accomplished by the leaving-one-out method that guards against underestimating the classification error: each case is classified by the discriminant functions derived from all cases other than that particular case

The same type of analysis, without subtests or item sets that do involve reading or writing, for all 101 aphasic patients (literate and illiterate), originally diagnosed as belonging to a standard aphasia syndrome, yielded 89.1% of concordant classifications. Applying the stepwise method for the 101 aphasic patients yielded 91.1% of concordant classifications. The variables that constitute the discriminant functions were the same seven variables as in the analysis of the 88 literate aphasic patients, mentioned above.

DISCUSSION

The present study showed that, overall, the Portuguese version of the AAT is a reliable instrument to evaluate native Portuguese individuals with aphasia and for cross-linguistic research of aphasia in Europe. The following discussion will center on the psychometric properties of the PAAT and their comparison with the German original and the English and Italian versions. Finally, cross-linguistic issues arising from the adaptation of the AAT to Portuguese will also be analyzed.

Demographic Data

Considering the performance of the control group first, it is worth noting that age and education (schooling) had a significant effect on test results. Furthermore there was a negative correlation between age and schooling (i.e., elderly individuals tend to be less educated). Therefore normative data for the PAAT should include correction terms for age and years of schooling.

Differences concerning age, gender, and educational level between the control and the aphasic group derive from the fact that we intended to obtain normative data for age and educational level that are representative of the whole population. However, the composition of the patient group does not mirror that population. Clinical populations with aphasia compared to the general population tend to be older, and hence they are likely to have fewer years of education. There were more male than female subjects in the patient group.

Table 7. Syndrome assignment properties of the PAAT
Discriminant analysis for literate aphasic patients with standard syndromes ($n = 88$). Numbers in parentheses report doubtful classifications with a probability of allocation below 0.7. Discriminant analysis is based on the six spontaneous speech ratings together with the PAAT subtests Token Test, Repetition, Naming, Written Language and Comprehension.

			Predicted Group Membership (PAAT)				
			Global	Wernicke's	Broca's	Amnestic	Total
Cross-validated (BAAL)	Count	Global	18 (0)	1 (0)	1 (0)	0	20
		Wernicke's	0	21 (0)	0	3 (0)	24
		Broca's	2	0	12 (1)	0	14
		Amnestic	0	1 (0)	1 (0)	28 (2)	30
	%	Global	90,0	5,0	5,0	,0	100,0
		Wernicke's	,0	87,5	,0	12,5	100,0
		Broca's	14,3	,0	85,7	,0	100,0
		Amnestic	,0	3,3	3,3	93,3	100,0

This is because of the fact that stroke affects predominately the male population and that, probably for sociological and professional reasons, men are more often referred to language therapy. Similar reasons may explain the higher number of male patients in the adaptations of the AAT to the English (Miller et al., 2000) and the Italian language (Luzzatti et al., 1994). The demographic variables time since CVA, age and schooling did not differ significantly between the patient groups.

Internal Consistency

The existing versions of the AAT in other languages (Graetz et al., 1992; Huber et al., 1983; Luzzatti et al., 1994; Miller et al., 2000; Pracharitpukdee et al., 1998) showed high internal consistency for subtests and item sets. This property proved to hold true also for the PAAT. Cronbach's α coefficient ranged between .8 and .9. The only item set with Cronbach's α coefficient below .8 was auditory comprehension of phrases. This may be caused by the fact that the comprehension of complex verbal material is not that straightforward, requiring additional working memory, inhibitory and attentional resources. Performance in the comprehension subtest also profits from metalinguistic and more general problem solving strategies when having to handle multiple choice items. The items of the other test parts are much more homogenous in this respect. Miller and colleagues (2000) argued in a similar way concerning their findings of lower Cronbach's α coefficients for the comprehension item sets.

Validity aspects

The hierarchical cluster analysis for item sets supports the basic constructional ideas of the PAAT. The item sets of the individual subtests were allocated closer together than item sets from different subtests. The major branching of the cluster tree diagram is clearly according to a distinction between receptive and expressive tasks. In the English version this division had not been so clear-cut. Composing of Words and Phrases and Writing to Dictation correlated more highly with the receptive tasks.

In the PAAT, among the expressive tasks, the item set Reading Aloud grouped closest to the item set Color Naming. Alexia and color naming deficits are often related clinical symptoms. Two factors may contribute to this association: (1) the close anatomical neighborhood and the shared vascularization of the areas that support reading and color naming (Damásio & Damásio, 1983); and (2) the similarity between graphemes and color names from a structural point of view. The visual-verbal conjunction for graphemes and colors are arbitrary but singular. There are no real lexical alternatives for their designation (Stachowiak & Poeck, 1976). This may make the visual verbal conjunction of both in an analogous manner more susceptible to impairment. The results of the cluster analysis may reflect these conditions. In the English version of the AAT (Miller et al., 2000) a similar pattern was observed.

Increasing Complexity

The AAT was constructed in a way that the individual item sets within a subtest are of increasing difficulty. Average performance for the individual item sets of the Token Test and the subtest Repetition reflect this constructional principle consistently (cf. Table 5). This property holds true for the whole group of aphasic patients as well as for all individual syndrome groups.

As the item sets of the subtest *Written Language* were designed in a linguistically parallel way, using items with nearly the same phonological structure and comparable frequency, one would have expected a similar performance for the three item sets, reading aloud words and phrases, composing word/phrases from graphemes/morphemes and writing words and phrases to dictation. This held only true for the control group and the patients with amnesic aphasia.

The other four patient groups performed always better in the reading task compared to the writing performance. Paresis of the dominant hand can account for this fact only partially, because this difference was also observed in the item set composing word/phrases from graphemes/morphemes. For these patients, the areas involved in the production of written language seem to be more severely impaired.

The results for the Naming subtest are not that homogenous regarding the increasing complexity criterion of the test construction. Color Naming and Description of Situations and Actions yielded a better score for the groups of Wernicke's and amnesic aphasic patients as compared to the preceding item sets of the Naming subtest. In fact, the Description of Situations and Actions requires additional syntactic competencies that seem to compensate for the naming deficit of Wernicke's and amnesic aphasic patients. Naming of colors seems to be as easy as naming pictured objects with a single noun for global, Broca's and amnesic aphasic patients. Wernicke's aphasics performed even better on the Color Naming item set, probably because the naming of terms that belong to one well established semantic category leads to less semantic interference.

Two aspects of the findings for the subtest Comprehension are worth noting. All syndrome groups performed better on the auditory as compared to the reading comprehension tasks. The results for the comprehension of single words are always superior to those for phrases. The higher semantic and syntactic complexity of sentence comprehension leads to lower scores for those item sets. Processing complex tasks also requires more cognitive control and analysis in general no matter the language being processed. Consequently, the same pattern of results was observed in the original and the other translated versions. (Huber et al., 1983; Luzzatti et al., 1994; Miller et al., 2000)

Discriminatory Power

The capacity of the PAAT to distinguish aphasic from non-aphasic persons was, by means of a first discriminant analysis (DA, Table 6), a little bit lower (90.7%) than in the English version, in which the discriminant analysis yielded a concordant assignment of 93.9% of the cases, as compared to the clinical classification (Miller et al., 2000). For the Italian version the classification results were even better: 98.7% of the aphasic persons and 93.2% of the control persons were classified according to the clinical evaluation (Luzzatti et al., 1994). For the English version, the Token Test and Repetition proved to be the best predictors of aphasia; in the Italian version the subtest that separated best the aphasic from the non-aphasic group was the subtest Naming, followed by the subtests Repetition, Written Language, Token Test, and Comprehension. In the Portuguese Version, the strongest and only discriminatory variable between the two groups was the Token Test. One reason for these differences may be the fact that in our analyses, we used a stepwise discriminant function procedure. In general, hier-

archical analysis, in which variables have been entered in a hypothesis driven order, is preferable because stepwise analysis may generate unreliable results. However, in the absence of hypotheses to drive variable entry order (as in our study), leaving variables of low explanatory value in a model may increase noise to the extent that estimates of the effect of truly related factors may be biased or their relationship to the dependent variables obscured. Stepwise analysis can address this problem. However, further studies will be needed before firm conclusions can be drawn. These discriminatory differences are on one hand probably influenced by the properties of the languages themselves. According to the Competition Model (MacWhinney & Bates, 1989; Bates et al., 1991) cross-linguistic differences can be explained by different cue-validity and different cue costs of the grammatical structures of the respective language. These cross-linguistic differences are robust under brain damage, turning the distinction between non-aphasic and aphasic persons into a relationship sui generis for a given language.

On the other hand even aging, fatigue, or lower educational level *per se* can impair linguistic performance (Bates et al., 1987; Reis et al., 2001). The control persons from the Portuguese normative group did not show a clear ceiling effect in all subtests. Probably because of the large variance in educational level there was performance overlap between poorly educated healthy persons and mildly impaired aphasic patients, thus decreasing the discriminatory power of the test. For the final version of the PAAT, the separation between aphasic and non-aphasic persons should be improved by introducing correction factors for educational level and age.

A second discriminant analysis (DA, Table 7) that was used for the classification of patients with standard syndromes yielded good concordance between the BAAL and the PAAT. Thus, the adaptation of the AAT to Portuguese has a similar classificatory power as the already existing adaptations to other languages. The fact that all spontaneous language evaluation categories entered into the discriminant analysis clearly shows that the evaluation of spontaneous language production is of outstanding value for classification purposes. Among the formal language tests, only the Token Test and the subtest Repetition contributed to the Discriminant Analysis.

Cross-linguistic Issues

The purpose of this study was to contribute to the generation of a unitary clinical diagnostic instrument for the assessment of aphasic language impairment in Europe, rather than to add language specific data for Portuguese to the cross-linguistic data corpus. From this point of view this study is in line with language as a between-subjects variable approach, described by Bates et al. (2001) and Bates et al. (1991). Reliable classification of aphasic syndromes is a fundamental prerequisite for cross-linguistic studies. The PAAT uses a threefold approach to ascertain a reliable classification according to clinical syndromes. Firstly, it pro-

vides numerical scores for the language performance of aphasic patients regarding the different modalities of language. Secondly, severity of symptoms can be accessed, comparing the individual performance with reference to the normative data of standard syndromes, including correction for age and educational level. This procedure converges on Bates' and colleagues' proposal of composing patient groups by percentile matching (Bates et al., 1991). Lastly, the evaluation of the spontaneous language production also contributes to syndrome classification. Classification based on behavioral data was used in many former cross-linguistic studies (Bates et al., 1987; Benedet et al., 1998). As mentioned earlier, the evaluation of spontaneous speech contributes significantly to the discriminant power of the PAAT. We believe that the classification procedure of the PAAT provides reliable results for matching patient groups.

One last point deserves attention. Following statistical figures provided by the United Nations Organization, Portuguese is the fifth most spoken language of the world, with variants in four continents. For the use of the PAAT outside the European Portuguese speaking population slight changes would be necessary in order to adapt the test material cross-culturally.

CONCLUSION

The PAAT maintains the psychometric properties of the original German version. It has good sensitivity and specificity for the diagnosis of aphasia. It showed a good correspondence with the BAAL used as the reference for the diagnosis of aphasia in Portugal. Because the original version showed very high inter-rater agreement and good test-retest correlation, we refrained from repeating these studies for the Portuguese version, arguing that because of the identical scoring system and the very similar test construction, the findings for the original version may be extrapolated to the PAAT. Low educational level, and to some extent advanced age of the persons tested have to be taken into account as covariates, as they add to aphasia in lowering performance on certain test parts.

The psychometric properties make the test a useful tool for diagnostic assignment to aphasia syndrome categories; a detailed quantification of impairments that can support a detailed plan for therapy; and for the evaluation of efficacy of speech therapy. In a European perspective, the PAAT enables Portuguese language researchers to participate in international studies where the AAT is the reference tool. The PAAT could be adapted to the other existing variants of Portuguese with some slight changes for some of the stimuli.

ACKNOWLEDGMENTS

REFERENCES

Bates, E., Devscovi, A., & Wulfeck, B. (2001). Psycholinguistics: A Cross-Language Perspective. *Annual Review of Psychology*, *52*, 369–396.

Bates, E., Frederici, A., & Wulfeck, B. (1987). Comprehension in Aphasia: A Cross-Linguistic Study. *Brain and Language*, *32*, 19–67.

Bates, E., Wulfeck, B., & MacWhinney, B. (1991). Cross-linguistic research in aphasia: An overview. *Brain and Language*, *4*, 123–148.

Benedet, J.M., Christiansen, J.A., & Goodglass, H. (1998). A cross-linguistic study of grammatical morphology in Spanish and English speaking agrammatic patients. *Cortex*, *34*, 309–336.

Benton, A.L. (1969). Development of a Multilingual Aphasia Battery. Progress and Problems. *Journal of Neurological Sciences*, *9*, 39–48.

Benton, A.L. (1989). *Multilingual aphasia examination* (2nd ed.). Iowa City, IA: AJA Associates.

Castro-Caldas, A. (1979). Diagnóstico e evolução das afasias de causa vascular. PhD Thesis, Faculty of medicine of Lisbon.

Damásio, A.R. (1973). Perturbações neurológicas da linguagem e de outras funções simbólicas. PhD Thesis, Faculty of medicine of Lisbon.

Damásio, A.R. & Damásio, H. (1983). The anatomic basis of pure alexia. *Neurology*, *33*, 1573–1583.

Ferro, J.M. & Kertesz, A. (1987). Comparative classification of aphasic disorders. *Journal of Clinical and Experimental Neuropsychology*, *9*, 365–375.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Frederici, A., Weissenborn, J., & Kail, M. (1991). Pronoun comprehension in aphasia: A comparison of three languages. *Brain and Language*, *41*, 289–310.

Graetz, P., De Bleser, R., Willmes, K., & Heeschen, C. (1991). De Akense afasie test. *Logopedie en Foniatrie*, *63*, 58–68.

Graetz, P., De Bleser, R., & Willmes, K. (1992). *Akense Afasie Test*. Nederlandstalige Versie (Lisse: Swets and Zeitlinger).

Huber, H.P. (1973). *Psychometrische Einzelfalldiagnostik*. Weinheim: Beltz.

Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1983). *Der Aachener Aphasie Test*. Göttingen: Hogrefe.

Huber, W., Poeck, K., & Willmes, K. (1984). The Aachen Aphasia Test. In F.C. Rose (Ed.), *Advances in Neurology, Vol. 42: Progress in Aphasiology*. New York: Raven Press.

Luzzatti, C., Willmes, K., Bisiacchi, P., De Bleser, R., Faglia, L., Mazzucchi, A., Posteraro, L., & Taricco, M. (1987). L'Aachener Aphasie Test (A.A.T.)-II: Proprietà della versione italiana. *Archivio di Psicologia, Neurologia e Psichiatria*, *48*, 480–519.

Luzzatti, C, Willmes, K., & De Bleser, R. (1991). *L'Aachener Aphasie Test (A.A.T.)*. Fierencze, Italy: Organizzazioni Speciali.

Luzzatti, C., Willmes, K., De Bleser, R., Bianchi, A., Chiesa, G., De Tanti, A., Gonella, M.L., Lorenzi, L., & Pozzoli, C. (1994). Nuovi dati normativi per la versione italiana dell'Aachener Aphasie Test (A.A.T.) *Archivio di Psicologia, Neurologia e Psichiatria*, *55*, 1086–1130.

MacWhinney, B. & Bates, E. (Eds.). (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

Miller, N., De Bleser, R., & Willmes, K. (1998). The English language version of the Aachen Aphasia Test. In W. Ziegler & K. Deger (Eds.), *Clinical phonetics and linguistics*, 253–261. London: Whurr Publishers.

Miller, N., Willmes, K., & De Bleser, R. (2000). The psychometric properties of the English language version of the Aachen Aphasia Test (EAAT). *Aphasiology*, *14*, 683–722.

Orgass, B. (1976). Eine Revision des Token Tests: I. Vereinfachung der Auswertung, Itemanalyse, und einführung einer Alterskorrektur. II. Validitätsnachweis: Normierung und Standardisierung. *Diagnostica*, *22*, 70–87, 141–156.

Poeck, K. & Göddenhenrich, S. (1988). Standardized tests for the detection of dissociation in aphasic language performance. *Aphasiology*, *2*, 375–380.

Pracharitpukdee, N., Phanthumchinda, K., Huber, W., & Willmes, K. (1998). The Thai version of the German Aachen Aphasia Test (AAT): Description of the test and performance in normal subjects. *Journal of The Medical Thai Association*, *6*, 402–412.

Reis, A., Petersson, K.M., Castro-Caldas, A., & Ingvar, M. (2001). Formal schooling influences two- but not three dimensional naming skills. *Brain and Cognition*, *47*, 397–411.

Stachowiak, F.J. & Poeck, K. (1976). Functional disconnection in pure alexia and colour naming deficit demonstrated by facilitation methods. *Brain and Language*, *3*, 135–143.

Willmes, K., Poeck, K., Weniger, D., & Huber, W. (1983). Facet theory applied to the construction and validation of the Aachen Aphasia Test. *Brain and Language*, *18*, 259–276.

Willmes, K. (1985). An approach to analyzing a single subject's scores obtained in a standardized test with application to the Aachen Aphasia Test (A.A.T.). *Journal of Clinical and Experimental Neuropsychology*, *7*, 331–352.

Willmes, K. (1990). Statistical methods for a single-case study approach to aphasia therapy research. *Aphasiology 1990*, *4*, 415–436.

Willmes, K. (2003). The methodological and statistical foundations of neuropsychological assessment. In P.W. Halligan, U. Kischka, & J.C. Marshall (Eds.), *Handbook of Clinical Neuropsychology (Ch. 3)*. Oxford: Oxford University Press.