Journal of Neuroscience and Behavioural Health Vol. 3(2), pp. 16-26, February 2011
Available online http://www.academicjournals.org/JNBH
ISSN 2141-2286 ©2011 Academic Journals

Full Length Research Paper

Brief aphasia evaluation (minimum verbal performance): Psychometric data in healthy participants from Argentina

Vigliecca Nora Silvana[1]*, Peñalva Marisa Carola[2], Castillo Jorge Antonio[2], Molina Silvia Cristina[2], Voos Javier Alfredo[3], Ortiz Maria Marta[4] and Ribichich Marialicia[5]

[1]Consejo Nacional de Investigaciones Científicas y Técnicas de la Argentina (CONICET); - Argentina. Servicio de Neurocirugía del Hospital Córdoba; Centro de Investigaciones de la Facultad de Filosofía y Humanidades- (CIFFyH). Universidad Nacional de Córdoba, Argentina.
[2]Centro de Investigaciones de la Facultad de Filosofía y Humanidades (CIFFyH). Universidad-Nacional de Córdoba, Argentina.
[3]Grupo de Ingeniería Clínica de la Universidad Tecnológica Nacional (UTN) Regional Córdoba, Argentina.
[4]Colegio de Fonoaudiólogos. Buenos Aires, Argentina.
[5]Consultorio Psicopedagógico Privado. Buenos Aires, Argentina.

Accepted 22 January, 2011

Aphasia test psychometric studies are necessary, especially for brief instruments in Spanish. A test was designed to quickly detect the basic resources of verbal communication in aphasic patients. In a sample of healthy Spanish-speakers, it was studied: (a) their test performance according to demographic variables; (b) the inter-item-internal consistency; (c) the subtest-factor structure; (d) the test-retest reliability during an extended inter-test interval of 7 to 14 months; (e) the inter-rater reliability. Data were collected from 151 participants living in Cordoba or Buenos Aires, aged 6 to 80, 56% females, with an average 8-year education. Subsamples of 34 and 26 participants were evaluated for test-retest and inter-rater reliability studies. The total score was only affected by age. Age and education had a significant effect on the time required to perform the test. Satisfactory reliability coefficients were observed. An exploratory and unrestricted factor analysis indicated that 68% of the variance was explained by three factors. A one factor solution was also suitable. The test-score proved to be reliable, representative of a verbal homogeneous construct and sensitive to the effect of age. The administration of this test of free-distribution (with 72 items) required on average, between 11 and 17 min.

Key words: Aphasia screening test, minimum verbal performance, reliability, validity, norms, comparative city studies, age, education, gender, free-distribution tests.

INTRODUCTION

Background

Spanish is one of the most widely spoken languages in the world and Spanish-speaking populations are increasing worldwide. Therefore, developing valid and reliable psychological instruments for any Spanish-speaking community is necessary. Considering aphasia tests, although, there is a need to psychometric studies on these tests in general; Spanish instruments particularly lack these types of studies.

Healthy subject (HS) studies are crucial in

*Corresponding author. E-mail: nsvigliecca@hotmail.com. Tel: +54 3543 420877. Fax: +54 3514334061, ext. 102.

psychometrics. Nevertheless, there are no aphasia tests for Spanish speakers which had shown, in their psychometric bases, the test-performance according to the demographic variables as well as the inter-rater reliability in HS. The comprehensive Boston Diagnostic Aphasia Examination (BDAE)-Spanish version (Goodglass and Kaplan, 1996) and the short Bedside Assessment of Language in Spanish (Sabe et al., 2008) did not provide any kind of HS data or inter-rater reliability indices in their original studies. As a way of completing the data provided in the English and the Spanish BDAE manuals, Davis (1993) reported inter-rater reliability studies on the original BDAE and Pineda et al. (2000) analyzed the HS performance and the effect of demographic variables on a Spanish version of this test.

Considering brief aphasia tests in any language, not only is the information pertaining to measurement properties and clinical utility limited (Salter et al., 2006) but also and most probably as a result of ceiling effect, HS-reliability studies are rare. Test-retest reliability throughout long-term periods in HS has never been studied and factor analysis has not been commonly carried out. Besides, the effect of demographic variables as part of the norm studies is not clear (Kostalova et al., 2008).

In the present exploratory study, a new aphasia test was analyzed trying to respond to some of the questions described earlier for HS. The Brief Aphasia Evaluation (BAE) was developed to mainly assess the minimum verbal performance (that is, the basic verbal functioning) of patients with aphasia so as to begin, after that with a more complete and better organized neuropsychological or language evaluations. By means of this technique the patient's performance was attempted to be explored through a more operative way, which is imperative for professionals who have to administer neuropsychological tests in difficult conditions, particularly in public hospitals.

By assessing the minimum verbal performance in HS rather than just the impairment in patients, the patients' dysfunction was mainly inferred by the discrepancy observed with respect to HS. However, other comparisons are also needed for validation (see below).

In this laboratory, the presence and magnitude of a general verbal deficit (including the presence and magnitude of its symptoms or components) are considered more useful for the patient's diagnosis and treatment than try to find a perfect archetype for each combination of symptoms (Vigliecca, 2010). In spite of that, we do not intend to replace the theoretical and technical perspective of the professional involved, but simply to complement it.

Neuropsychological techniques are proposed to reach through psychological tests, an appropriate diagnosis of brain functions both in their quality and magnitude. Studies from our laboratory have demonstrated the concurrent validity of this test to differentiate between patients with left cerebral lesions (that is, patients injured in the verbal dominant hemisphere) and patients with right cerebral lesions, as well as patients with left cerebral lesions and HS (Vigliecca et al., 2011).

Factor-structure studies

Salter affirmed (Salter et al., 2006) that screening tools are useful to detect the presence or absence of a particular condition of construct. However, there are almost no studies which have verified through factor analysis, the structure of the components of the aphasia screening tests according to the expected construct. As far as we know, the factor structure of a brief aphasia test has only been studied in the Reitan-Indiana Aphasia Screening Examination (ASE) (Williams and Shane, 1986).

Regarding the BAE, the underlying or theoretically expected construct is the minimum verbal performance, which is shared by patients and HS. An unrestricted factor analysis carried out in a sample of patients and HS (Vigliecca et al., 2011) indicated that, all its subtests (with loadings of 0.65 or above) grouped in one verbal factor which explained 78% of the variance. The same pattern was observed in the sample of patients alone. However, as only 38 HS participated of that study, factor analysis could not be accomplished in the control group since some subtests showed no variance.

Reliability studies

Satisfactory reliability coefficients have been reported for the Western Aphasia Battery (WAB) (Shewan and Kertesz, 1980) and the BDAE (Goodglass and Kaplan, 1996) although; both studies have been performed just on aphasic patients.

Inter-rater reliability coefficients of 0.75 or above have been reported for the BDAE (Davis, 1993). The BDAE original manual does not provide test-retest reliability data. Shewan and Kertesz (1980) reported test-retest reliability coefficients of 0.88 or above for the WAB subtests only in a group of chronic aphasic patients; however, this test is worldwide used (many times as the gold standard) not only for chronic patients. Test-retest coefficients were studied by these authors during an extended inter-test interval of at least 6 months but the statistical significance of the changes observed between test and retest was not reported.

As regards brief aphasia tests, reliability studies have also been performed on patients. Test-retest and inter-rater reliability coefficients of 0.70 or above have been reported for the Acute Aphasia Screening Protocol and the Frenchay Aphasia Screening Test (FAST) (Salter et al., 2006; Crary et al., 1989; Enderby et al., 1987).

Theodoros et al. (2008) reported intra and inter-rater reliability coefficients of 0.59 or above for the short form of the BDAE which was assessed via an internet-based videoconferencing system. For Spanish language, Sabe et al. (2008) reported test-retest and internal-consistency coefficients of 0.62 or above in aphasic patients. Although, the time since onset of disease was not reported by these authors (Sabe et al. (2008), the changes observed between test and retest (46% of variation) were highly significant.

Aphasia versus language stability

In spite of the earlier said, the aphasic condition is not expected to be stable; great fluctuations can be observed in aphasic patients even overnight. On the contrary, it is the underlying verbal ability of non-brain-damaged participants which is expected to be stable thus, allowing the establishment of a normative performance against which to judge the aphasic performance.

A significant performance improvement has consistently been reported for aphasic patients when several weeks or months elapsed since onset of condition (Sabe et al., 2008; Shewan, 1986; Wade et al., 1986). Therefore, a reliable aphasia test should show a corresponding stability in HS during those several weeks or months. Otherwise, learning curves would be necessary for patient and control groups.

Effect of demographic variables

Considering the comprehensive BDAE in its Spanish version, Pineda et al. (2000) observed, in a sample of healthy Colombian workers living in Medellin, that gender and age had a significant effect on reading and writing subtests, while education had a significant impact on most of the BDAE subtests. In fact, a stepwise regression model showed that education was the best predictor of the variance in the BDAE scores.

Regarding the HS-effect of demographic variables in brief aphasia tests, Kostalova et al. (2008) reported an effect of education on the Mississippi Aphasia Screening Test- Czech version (MASTcz), which was coincident with the original MAST (Nakase-Thompson et al., 2005) and the ASE (Williams and Shane, 1986). Although, no effects of age (range 14 to 87) or gender were reported (Kostalova et al., 2008), an interaction between age and education was observed when age was coded below and above 60 years old. The lack of effects of age and gender (Kostalova et al., 2008) was not coincident with the original MAST (Nakase-Thompson et al., 2005). Meanwhile an inverse relationship between age (range 21to81) and score was observed for the FAST (Salter et al., 2006).

Minimum verbal performance, HS-variance and ceiling effect

These demographic contradictory results are interesting because a minimum verbal performance test involving over learned verbal tasks should supposedly be less prone to be affected by demographic variables in HS. Actually, a minimum neuropsychological performance which as a whole can be reached by all or almost all HS (but not by all brain-injured patients) is the rationale which underlies the very widespread mini mental test of Folstein, for example (Folstein et al., 1975). Nevertheless, not only should the poor variance but also the effect of demographic variables be corroborated in HS, before analyzing the factors which contribute to determine the variance among brain injured patients. As far as we know, the effect of demographic variables on brief aphasia tests has never been studied in Spanish-speaker samples.

On the other hand, ceiling effect (if it really exists in HS) is almost a necessary condition to be studied in this kind of tests considering that it represents both a superior limit and a control parameter for the patients' performance. As a consequence, it should be specifically analyzed, not avoided or assumed *a priori*.

Research questions

The BAE is part of a bigger research project which attempts to develop culture-free (or easily adaptable to different languages) tests under a multicenter approach. So, the BAE items were specially designed for this purpose. The words were selected in view of their high frequency of use and cultural or geographical dependent words or phonemes were excluded. The underlying and general scientific proposal of such research, aims at developing efficient tests, that is, brief and/or easy to apply neuropsychological techniques without neglecting the goals of accuracy and validity.

In the present study, we specifically wanted to analyze the performance which can be expected in different groups of HS according to the demographic variables as a way of getting a preliminary standardized reference (a norm) for the patients' performance interpretation. In order to analyze the effect of the recruitment place, a small subsample belonging to an additional city was evaluated.

With the aim of analyzing whether the different parts of the BAE can be considered as components of the same verbal construct (conceptual validity), an exploratory subtest-factor analysis was carried out in this bigger and more variable sample of HS than the one recruited in a previous study (Vigliecca et al., 2011).

Regarding the reliability indices, not only the internal consistency of the items which measure the same concept was deemed to be estimated, but also the stability and the objectivity.

In this exploratory research, no specific hypotheses were formulated in view of the lack of enough or consistent antecedents. However, and in addition to the need of exploring the HS psychometric properties and the test efficiency, it was expected that the test resulted as resistant as possible to the effect of demographic variables and (as any other norm referenced test) useful to be applied in health institutions belonging to the places from which the samples were recruited.

Additionally, the present methodological approach may represent a step forward (not a perfect one) in the difficult process of getting satisfactory psychometric properties in short tests.

Objectives

To study in a sample of healthy Spanish speakers; (a) their performance in the BAE according to demographic variables; (b) the inter-item-internal consistency; (c) the subtest-factor structure and, in a subgroup of those HS; (d) the test-retest reliability during an extended inter-test interval of 7 to 14 months, as well as (e) the inter-rater reliability.

By simultaneously analyzing several indices of HS performance, a better control parameter for the patients' performance may be obtained.

MATERIALS AND METHODS

Materials

The present test, which was freely offered by its authors in paper, informatics and English versions, (http://neurotests.frc.utn.edu.ar) was designed to bed-side and quickly detect the basic resources of verbal communication in aphasic patients. This aphasia instrument is part of the battery of "Neuropsychological Tests Abbreviated and Adapted for Spanish-Speakers" (Vigliecca, 2004; Vigliecca and Aleman, 2009) and it was composed of 72 items scored from 0 to 3 (maximum score 216). This test was developed ten years ago (Vigliecca, 2000), independently of the proposed in other aphasia screening tests.

The main functions to be studied with the test were: Comprehension, expression, naming, repetition, reading, writing, attention (phonemic analysis and synthesis), memory and praxia. All the tasks were verbal except maybe for praxia (see below).

Comprehension included items such as salutation, questions to be answered with a gesture, questions to be answered with "Yes/ No" or "I don't know" and "point at" questions; *expression* explored the interviewee's ability to describe his/her own disease, as part of the spontaneous speech and it also involved saying the numbers from 1 to 10 and saying the alphabet (until the letter "I"), as part of verbal automatisms; naming included the tasks of saying the names of single objects and completing sentences which were started by the interviewer; repetition included the tasks of saying the same words, word series and sentences which were uttered by the interviewer; reading involved the visual discrimination of numbers, letters and words which were said by the interviewer, the execution of written commands, the comprehension of written expressions by confrontation with pictures and reading aloud; writing included items such as writing the name, the numbers from 1 to 10 and the alphabet until the letter "I", writing numbers, letters, words and sentences by dictation, as well as writing spontaneously a sentence; attention was based on the tasks of phonemic discrimination (selection of a target among several options), phonemic synthesis (recognition of a spelled word), spelling (phonemic analysis) and reversed spelling; memory involved a delayed repetition (that is, the recall of words which had been previously said by the interviewer); praxia involved the imitation of a mouth movement made by the interviewer.

The first six functions (specially the first four and complementarily, reading and writing) are classically assumed as the main factors which determine the variance among brain injured patients, thus, hypothetically discriminating aphasia types. The remaining three functions were incorporated with exploratory purposes. Other conditions such as visual agnosia, visual field defects, orientation, prosody, gesture- comprehension and expression, etc. are also complementarily explored by this test, but they were not analyzed in the present work.

Subjects

Participants were community-dwelling Argentinean volunteers without any known neurological or psychiatric disease, adapted to their daily life demands. They were recruited from academic, cultural, recreational and retirement centers from Córdoba (77%) and Buenos Aires (23%). The centers were intentionally selected trying to get heterogeneity in the socioeconomic background. The institution authorities were interviewed and the research objectives as well as the characteristics required for the participants were explained. A list of potential volunteers was prepared. The test did not pose any risk to the participants who, on the contrary, were cognitively stimulated. We had the informed consent of the participants or the caregivers. The institutions and individuals who participated in the research remained anonymous. They were asked for other potential participants thus, trying to create a network of volunteers.

In order to include a subject in the sample, we also took into consideration the information provided by the initial interview of the neuropsychological battery. This interview evaluated risk factors, background and their probable incidence on cognition and behavior. We asked questions about genetic neurological or psychiatric disorders; hypoxia during labor; significant diseases; head trauma with loss of consciousness; seizures; tumors; hypertension; heart disease; faints; consumption (previous and current) of: drugs, alcohol, cigarettes, medicines, etc.; malnutrition; contact with toxic agents (frequency and duration); school performance (failures and merits); development (language, motor skills, sphincter control); relevant changes in: feeding behavior, sexual behavior, sphincter control, sleep-wake cycle, mood or character; presence of: unexpected responses, odd sensations, delusions, hallucinations, alterations in perception (illusions, distortions), etc.; memory, speaking, verbal comprehension, attention, orientation (personal, place and time), reading, and writing problems; sensory difficulties, etc. For younger than 11 year-old-subjects the interview questions were answered by the parents or caregivers.

We excluded subjects who showed symptoms of neurological or psychiatric disorders, psychotropic drug consumption including excessive alcohol, risk of neurological damage by disease or accident, any kind of medical illness which could affect neuropsychological performance or sensory or motor difficulties which prevented them from carrying out the tests fluently. Students who had done some academic year twice were also excluded. In case of any doubt from any member of the research team regarding the intervening variables, participants were excluded.

The effect of demographic variables, the internal consistency and the factor structure were studied in a sample of 151 HS, aged 6 to 80 (mean ± SD: 31.30 ± 24.25); 56% females and 44% males; 59% with a background of elementary education, 22% with a background of high school (secondary) education and 18% with a college or university background (education years: 7.81 ± 5.56). Only eight subjects were not right-handed (ambidextrous or left-handed). Only eight six-year-old children (6YOC), 3 girls and 5 boys attending the first grade of the elementary school, were in the process of acquiring the basic reading and writing skills. The rest of the subjects had at least completed one year of education.

From this total sample, a sub group of 31 right-handed subjects was re-evaluated for the test-retest reliability study. In this subsample of 7 to 80 years old (50.23 ± 24.03), there were 47% females and 53% males, of whom 35% had an elementary level, 33% a high school level and 32% a higher level of education (education years: 11.06 ± 5.74). The range of the inter-test interval was 7 to 14 months (10.85 ± 2.16).

From this subsample, another group was re-evaluated for the inter-rater reliability study. In order to carry out this study, two trained interviewers acting independently evaluated the same interviewee. The consistency between the two interviewers' evaluations was analyzed by means of both the interview recorded on an audiotape (in which written answers could not be seen) and the interviewee's written answers during the written subtests. For the audio-register, only those subjects who had either agreed to be recorded or had favorable recording conditions in their environments were studied. This subsample was made up of 26 HS of 7 to 80 years old (51.92 ± 25.12), 57% females and 42% males; 38% with elementary level of education, 31% with secondary level and 31% with higher level (education years: 10.61 ± 6.03).

Alternatively, the interpretation that the second rater made from the first rater's administration protocol was also analyzed. Such interpretation was done independent of the explicit interviewee's behaviors, that is, by just looking at the first rater's written-administration registers (evaluations, crosses, marks, comments, etc.) about the interviewee. For this paper register, the sample of the previously described 31 HS was studied.

Statistical analysis

When the effect of demographic variables was analyzed, the total score and the duration of the test were the dependent variables. The total score was the main variable of study because the verbal performance is primarily expressed by this index. The test duration was incorporated just as complementary information. Both variables were analyzed by a gender x education x age ANOVA (age-range: 12 to 80; age-intervals: 24 years). For younger than 12 year-old-subjects (age-range: 6 to 11) only a gender x age ANOVA was carried out since all these subjects only attended elementary school; besides, years of study were correlative to years of age for this group. One-way ANOVAs were carried out for those analyses in which only one independent variable was involved.

The dependent variables were quantitative; the adequacy of the underlying model for the ANOVAs was checked by the residual analysis (the primary tool for checking normality was the residual histogram (plots shown)); the plots of the residual versus the predicted values was mainly used for checking the equality of variances (plots not shown)); in case of inadequacy, data were reevaluated by their ranks to check consistency (results not shown). In case of inconsistency, data were reevaluated by nonparametric statistics (main effects only, results not shown). The coefficients of variation (CV) were also taken into account in the analysis of ANOVAs assumptions.

The internal consistency was analyzed by the standardized Cronbach's alpha coefficient (Vogt, 2005) and the rest of the reliability indices by the intra-class correlation coefficient (ICC), whereby the order of administration is randomized. Differences between test and retest and differences between both raters were analyzed by the dependent sample-Student t-test. An exploratory factor analysis (by using the principal component-method and the varimax raw-rotation) was carried out with the nine subtests of this instrument. The variance involved in each successive factor was analyzed (that is, the explained variance) and only factors with Eigen values greater than 1 (Kaisser criterion) were retained; factor loadings were shown as descriptive data.

RESULTS

Effect of demographic variables

Participants from Córdoba and Buenos Aires did not differ either in the total score or in the time required to perform the test (score (F (1, 149) = 0.78, p < 0.38: Buenos Aires: 210.23 ± 11.58, Córdoba: 211.89 ± 9.17, total sample 211.51 ± 9.77; test duration (F (1, 149) = 1.76, p < 0.19: Buenos Aires: 13.11 ± 5.23, Córdoba: 14.23 ± 4.07, total sample 13.97 ± 4.38). Participants of these two cities were also similar in their demographic characteristics (years of age and education (F (1, 149) ≤ 0.19, p ≥ 0.66; gender and manual preference ($\chi^2 \leq 0.97$; df: 1; p ≥ 0.32). Therefore, both samples were considered as a single one.

The score of non-right handed subjects was not significantly different from the score of right-handed subjects (F (1, 149) = 0.53, p < 0.47). As a consequence, manual preference was not considered a criterion to exclude subjects in the analysis of demographic variables. On the contrary, a significant difference was observed between the 6YOC and the rest of the sample (F (1, 149) = 225.38, p < 0.001: 6YOC: 179.62 ± 24.78; rest of the sample: 213.33 ± 3.13). (The 6YOC were also different from both the rest of the children and the seven-year-old children (results not shown)). As a consequence, we excluded the 6YOC from the analysis of demographic variables except otherwise indicated. The remaining sample was distributed according to age-categories in thefollowing way: <11: 31%; 12 to 35: 29%; 36 to 59: 19%; >60: 20% (N = 143).

For the total score, the gender x education x age ANOVA for older than 12-year-old subjects indicated only a significant main age effect (Table 1). In general terms, the oldest subjects showed a poorer performance than the younger ones (Figure 1). By pooling subjects with secondary and higher levels of education, results did not change (Table 1). The gender x age ANOVA for younger than 12-year-old subjects indicated that, girls showed a

Table 1. Score: Effect-summary of gender x education x age ANOVA.

Total score (part a)							Total score (part b)						
ANOVA: 1-Gender, 2- education (three levels), 3- age							ANOVA: 1-Gender, 2- education (two levels), 3- age						
	df Effect	MS Effect	df Error	MS Error	F	p-level		df Effect	MS Effect	df Error	MS Error	F	p-level
1	1	2.06	80	6.74	0.31	0.582	1	1	2.61	86	6.75	0.39	0.536
2	2	7.17	80	6.74	1.06	0.350	2	1	7.07	86	6.75	1.05	0.309
3	2	59.82	80	6.74	8.87	0.000 *	3	2	93.13	86	6.75	13.80	0.000 *
12	2	3.10	80	6.74	0.46	0.633	12	1	0.62	86	6.75	0.09	0.762
13	2	12.58	80	6.74	1.87	0.161	13	2	11.26	86	6.75	1.67	0.195
23	4	3.29	80	6.74	0.49	0.745	23	2	2.79	86	6.75	0.41	0.663
123	4	9.85	80	6.74	1.46	0.222	123	2	2.90	86	6.75	0.43	0.652

Part a: 1st, 2nd and 3rd education levels. Part b: 2nd and 3rd levels were pooled.* Significant effect.

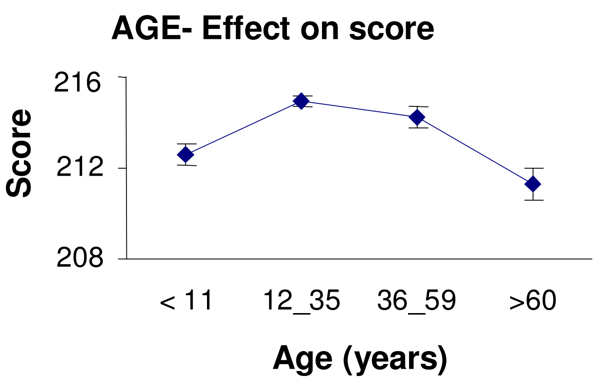

Figure 1. Score mean ± SE according to age-category, p < 0.001 (ANOVA).

trend to have a better performance than boys (results not shown) but, again, only a significant main effect of age was observed (gender: F (1, 41) = 3.01, p < 0.09; age: F (1, 41) = 5.73, p < 0.02; gender x age interaction: F (1, 41) = 1.09, p < 0.30). As expected for children, in this case the younger subjects showed a poorer performance than the older ones. Considering that both children and older subjects showed equivalent results, the two samples were pooled. Figure 1 shows the effect of age (alone) on score (F (3, 139) = 11.37, p < 0.001) for 7-to-80-year-old subjects. The average change between the age-category with the best performance and the age-category with the worst performance was less than 4 points, indicating less than 2% of variation.

Interestingly, by including the 6YOC, results did not change and the gender x age ANOVA also indicated just a main effect of age (gender: F (1, 143) = 0.75, p < 0.39; age: F (3, 143) = 5.97, p < 0.001; gender x age interaction: F (3, 143) = 1.82, p < 0.15). In this case, the mean score for <11-year-old subjects was 207.62 ± 15.30.

Just for informative purposes, when the test duration (see below) was incorporated as a covariate, none of the score-outputs changed (results not shown). Considering the test duration, the gender x education x age ANOVA for older than 12-year-old subjects indicated a significant main effect of both education and age without interaction among factors (Table 2). By pooling subjects with secondary and higher levels of education, results did not change (Table 2). Figure 2 shows the effect of age x education for 12-to-80-year-old subjects. As can be seen, older subjects showed a poorer performance (a longer test duration) than younger subjects and less educated subjects showed a poorer performance than more educated subjects. Nevertheless, the average change between the category with the best performance and the category with the worst performance was less than 4 min, indicating less than 4% of variation.

The gender x age ANOVA for younger than 12-year-old subjects in the test duration indicated neither main effects of gender or age nor interaction between factors (gender: F (1, 41) = 0.27, p < 0.61; age: F (1, 41) = 2.37, p < 0.13; gender x age interaction: F (1, 41) = 0.02, p < 0.87). The average time for the whole group of children was 15.19 ± 4.60 min, which was similar to the value observed in subjects with elementary level of education in the 12 to 80-year-old sample (F (3, 78) = 1.21, p < 0.31). By including the 6YOC in the sample of children, a significant age effect emerged (gender: F (1, 49) = 0.05, p<0.82; age: F (1, 49) = 6.11, p<0.02; gender x age interaction: F (1, 49) = 0.22, p<0.64). In this case, the average test-duration for 9-11-year-old children was 14.09 ± 4.32 minutes, while the average test-duration for <9-year-old children was 17.37 ± 4.91 minutes.

The histograms for the original three-way ANOVAs (Figures 3 and 4) indicated that, the residuals were fairly adjusted to normality (normality was shown by a curve). On the contrary, the plot of the residual versus the

Table 2. Test-duration: Effect-summary of gender x education x age ANOVA.

Test-duration (part a)							Test-duration (part b)						
ANOVA: 1-Gender, 2- education (three levels), 3- age							ANOVA: 1-Gender, 2- education (two levels), 3- age						
	df Effect	MS Effect	df Error	MS Error	F	p-level		df Effect	MS Effect	df Error	MS Error	F	p-level
1	1	3.72	80	4.63	0.80	0.373	1	1	2.31	86	4.47	0.52	0.474
2	2	43.58	80	4.63	9.42	0.000 *	2	1	76.51	86	4.47	17.13	0.000 *
3	2	35.90	80	4.63	7.75	0.001 *	3	2	34.20	86	4.47	7.66	0.001 *
12	2	0.36	80	4.63	0.08	0.925	12	1	0.29	86	4.47	0.06	0.799
13	2	7.97	80	4.63	1.72	0.185	13	2	7.40	86	4.47	1.66	0.197
23	4	1.17	80	4.63	0.25	0.907	23	2	2.94	86	4.47	0.66	0.520
123	4	1.66	80	4.63	0.36	0.837	123	2	3.20	86	4.47	0.72	0.492

Part a: 1st, 2nd and 3rd education levels. Part b: 2nd and 3rd levels were pooled. * Significant effect.

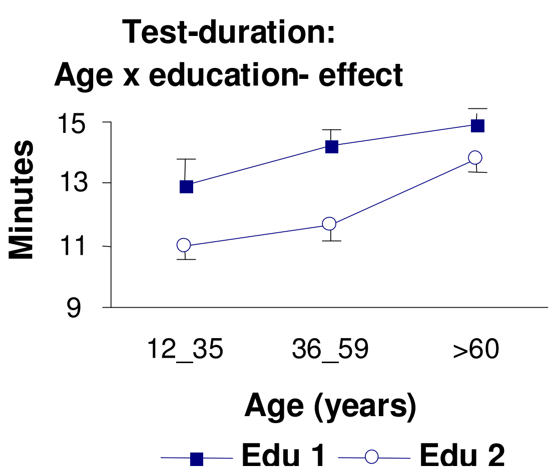

Figure 2. Time-mean ± SE according to age-category and education-level (Edu 1: First level, Edu 2: second and third level). ANOVA: Age effect: $p < 0.001$, education effect: $p < 0.001$. Age x education interaction: $p < 0.520$.

predicted values indicated a tendency for the variance of the residuals to decrease as the scores increased. The same plot did not reveal anything problematic for the test duration. The CVs were < 2% for the total score and 15% for the test duration which were considered satisfactory. Anyway, all the statistical analyses brought about identical outputs (results available upon request).

Additional information

Considering the total score, the 10th, 50th and 90th percentiles for the whole sample were 208, 214 and 216 respectively and by excluding the 6YOC, those percentiles were 209, 214 and 216.

Considering the test duration, the 10th, 50th and 90th percentiles for the whole sample were 10, 14 and 19 respectively and by excluding the 6YOC, those percent-tiles were 9, 14 and 17. (More information is available upon request).

Internal consistency and factor structure

By excluding items with zero variance, the Cronbach´s alpha coefficient indicated an internal consistency of 0.93. An unrestricted factor analysis indicated that, 68% of the variance was explained by three factors. Six of the verbal tests showed loadings of 0.50 or above in the first factor; naming and memory (with loadings of 0.72 and 0.74, respectively) grouped in the second factor, while *praxia* emerged as an independent third factor with a loading of 0.96. An internal consistency of 0.87 was observed among the subtests of the first factor, while in the rest of the factors the Cronbach´s alpha coefficient could not be calculated due to ceiling effect. A one factor solution was also acceptable with a subtest-internal consistency of 0.77 and an explained variance of 48%. Under this model, all the verbal subtests showed factor loadings of 0.22 or above.

Test-retest and inter-rater reliability

As regards the test stability, a satisfactory consistency between test and retest was observed for the total score without significant differences between both measures according to the Student t-test (Table 3). The average change between test and retest for the total score was 0.12 points, indicating a variation close to 0%. Consistency was lower for the test duration and a significant decrease in this measure was observed from the 1st to the 2nd administration. Nevertheless, the

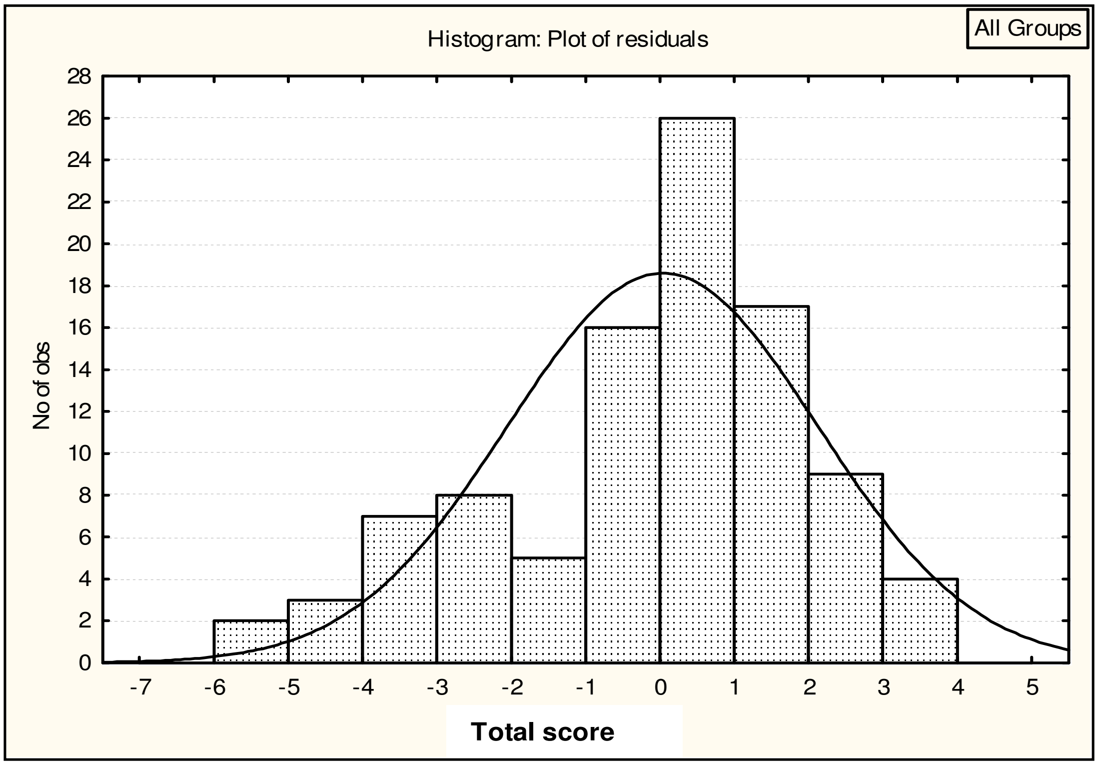

Figure 3. Histogram of the residuals for the gender x education x age ANOVA (Total score).

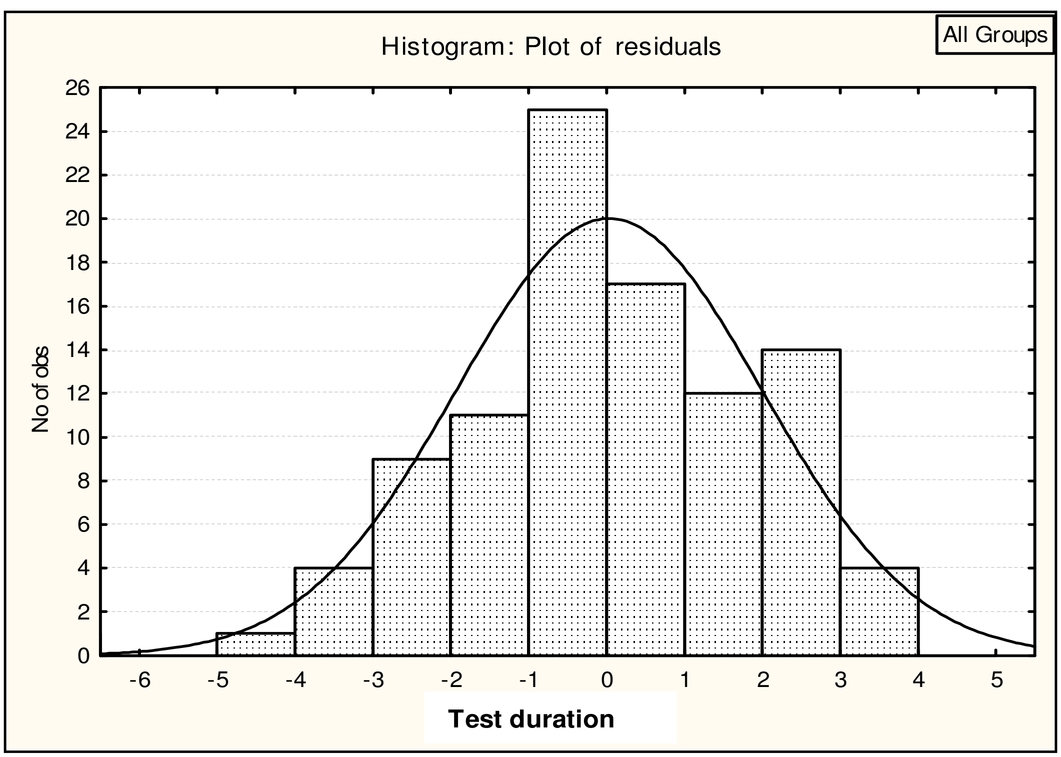

Figure 4. Histogram of the residuals for the gender x education x age ANOVA (Test-duration).

Table 3. Test-retest reliability.

	Test (1)	Retest (1)	ICC	Difference
Score	213.29 ± 2.61	213.18 ± 2.76	0.78	t = 0.40, df: 33, p < 0.689
Test-duration	13.03 ± 3.11	12.18 ± 3.27	0.74	t = 2.30, df: 33, p < 0.028 *

(1) Mean ± SD; * Significant effect.

Table 4. Inter-rater reliability.

Interpretation of the written and audio-taped interviewee's answers				
	Rater 1 (1)	**Rater 2 (1)**	**ICC**	**Difference**
Score	212.88 ± 2.76	212.69 ± 3.07	0.96	t = 1.31, df:25, p < 0.20
Time	12.73 ± 3.35	13.15 ± 3.74	0.93	t = -1.69, df:25, p < 0.11
Interpretation of the first rater's registers				
	Rater 1 (1)	Rater 2 (1)	ICC	Difference
Score	213.17 ± 2.76	213.15 ± 2.74	0.99	t = 0.44, df: 33, p < 0.66
Time	12.18 ± 3.27	12.21 ± 3.30	0.99	t = -1.00, df: 33, p < 0.32

(1) Mean ± SD.

average change between test and retest for the test duration was less than one minute. As regards the inter-rater reliability, a high consistency was observed among all the measures analyzed (that is, the second rater's interpretation of the interviewee's answers, and the second rater's interpretation of the first rater's registers in both the total score and the test duration). Significant differences between the two raters were not observed for any of the evaluations (Table 4).

DISCUSSION

The present age effect on score was coincident with the FAST (Enderby et al., 1987) and the original MAST studies (Nakase-Thompson et al., 2005). Nevertheless, the average change between the age-category with the best performance and the age-category with the worst performance in this work was less than 4 points, indicating less than 2% of variation.

Unlike the MASTcz (Kostalova et al., 2008), the MAST (Nakase-Thompson et al., 2005), the ASE (Williams and Shane, 1986) and even the extensive BDAE in its Spanish version (Pineda et al., 2000), we did not observe an effect of education on the present test score which surely had to do with the over learned tasks chosen for testing. This finding could be seen as an advantage for these kinds of tests which intentionally attempt to get a superior level of reference as resistant as possible to the effect of confounding factors. However, we did observe an effect of education on the test duration.

In previous aphasia test studies carried out on aphasic patients, the effect of demographic variables did not seem to be apparently a factor of interest. In the present HS-study, the coefficients obtained from the internal consistency and inter-rater reliability analyses were excellent, in agreement with previous aphasia test studies (Crary et al., 1989; Davis, 1993; Enderby et al., 1987; Goodglass and Kaplan, 1996; Sabe et al., 2008; Shewan and Kertesz, 1980). The score-test-retest ICC can also be considered satisfactory because the mean inter-test time was close to one year and as demonstrated with demographic variables, the scores of the most extreme age-categories tended to go in opposite directions throughout a year. The average change between test and retest for the total score indicated a variation close to 0%. Shewan and Kertesz (1980) observed a variation of 10% in chronic aphasic patients and Sabe et al. (2008) observed a variation close to 50% in apparently, non-chronic aphasic patients. In any case, present results set a precedent for the common and implicit assumption that healthy subject aphasia test scores do not significantly change during extended intervals. On the contrary, a significant decrease was observed between test and retest in the test duration, thus, surely representing an index of the effect of practice in this longitudinal design.

The subtest factor structure of the present test essentially showed one or two major verbal factors which clearly included eight of the nine subtests except for *praxia,* which apparently emerged as a non-verbal factor. Nevertheless, as this isolated subtest assesses an oral praxia related to verbal phonation (which is usually affected in aphasic patients), when the test was administered to patients, such subtest grouped with the rest of the verbal subtests (Vigliecca et al. 2011). Williams and Shane (1986) reported two major factors for the ASE; general language abilities factors band a sensorimotor coordination factor. However, the ASE intentionally includes non-verbal tasks.

Conclusions

Present results demonstrated that this test of minimum verbal performance (in its scored expression) can be considered a reliable instrument, representative of a homogeneous construct and in spite of dealing with over learned verbal tasks, sensitive to the effect of age in healthy subjects. According to the age factor, the inverted U-shaped figure obtained in the present study is consistent with most studies on long-term cognitive evolution. The poor variance observed on score also guarantees that a bigger number of cases are not necessary to confirm that pattern.

In general terms, the time required to administer the test was short according to the purpose of the test. In the older than 12-year-old sample, this time was affected by age and education but even considering the most extreme values achieved by the different groups, the time ranged on average between 11 and 15 min approximately. For children, this time was a little longer (between 14 and 17 min approximately). The test turned out to be efficient (valid, reliable and brief) although, it is well-known that comprehensive and extensive scales usually have better psychometric properties than brief scales. If a more subtle assessment of the aphasic symptoms is needed, a more complete language evaluation will have to be administered.

All the figures described in the present work shed light on the performance which can be expected from the different groups of HS analyzed, including children, adults, elderly people and to a certain extent, illiterates. As expected, most of the present subjects got a score near the maximum because this test was specially designed to detect pathology (not ability).

In principle, the test can be considered useful to be applied in Cordoba and Buenos Aires. The lack of effect of the recruitment place can be seen as a positive finding in the attempt of attenuating the strong impact that the different cultures usually have on verbal language tests. Further research would be necessary to see if the present results can be corroborated (or generalized to other samples and situations) considering that the study was carried out with an accidental sample of Spanish speakers.

If the usual cautions for norm referenced tests are taken into account (Haynes and Pindzola, 2008), the present psychometric studies and the methodological approach employed may represent helpful options to get better inferences in the aphasia topic.

ACKNOWLEDGEMENTS

This study was supported by funds from the National Council of Scientific and Technological Research in Argentina (CONICET) and the Ministry of Science and Technology of Cordoba Province. We are grateful to Cecilia Actis for her English assistance.

Abbreviations: BAE, Brief aphasia evaluation; **BDAE,** Boston diagnostic aphasia examination; **HS,** healthy subject; **ASE,** aphasia screening examination; **WAB,** western aphasia battery; **FAST,** frenchay aphasia screening test; **MASTcz,** Mississippi aphasia screening test- Czech version; **MAST,** Mississippiaphasia screening test; **6YOC,** six-year-old children; **ICC,** intra-class correlation coefficient; **SD,** standard deviation; **ANOVA,** analysis of variance; **df,** degrees of freedom; **CV,** coefficients of variation.

REFERENCES

Crary MA, Haak NJ, Malinsky AE (1989). Preliminary psychometric evaluation of an acute aphasia screening protocol. Aphasiol., 3: 611-618.

Davis GA (1993). A survey of adult aphasia and related language disorders (2nd edition). Englewood Cliffs, New Jersey: Prentice Hall Inc.

Enderby PM, Wood VA, Wade DT, Hewer RL (1987). The Frenchay Aphasia Screening Test: a short, simple test for aphasia appropriate for non-specialists. Int Rehabil Med., 8: 166-170.

Folstein MF, Folstein SE, McHugh PR (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12: 189-198.

Goodglass H, Kaplan E (1996). Evaluación de la afasia y trastornos relacionados. Adaptación española. Madrid: Panamericana.

Haynes WO, Pindzola RH (2008). Diagnosis and Evaluation in Speech Pathology (7th edition). Boston: Allyn and Bacon.

Kostálová M, Bártková E, Sajgalíková K, Dolenská A, Dusek L, Bednarík JA (2008) Standardization study of the Czech version of the Mississippi Aphasia Screening Test (MASTcz) in stroke patients and control subjects. Brain. Inj., 22: 793-801.

Nakase-Thompson R, Manning E, Sherer M, Yablon SA, Gontkovsky SLT, Vickery C (2005). Brief assessment of severe language impairments: Initial validation of the Mississippi aphasia screening test. Brain. Inj., 19: 685-691.

Pineda DA, Rosselli M, Ardila A, Mejia SE, Romero MG, Perez C (2000). The Boston Diagnostic Aphasia Examination-Spanish version: the influence of demographic variables. J. Int. Neuropsychol. Soc., 6: 802-814.

Sabe L, Courtis MJ, Saavedra MM, Prodan V, de Luján-Calcagno M, Melián S (2008). Desarrollo y validación de una batería corta de evaluación de la afasia: '*bedside* de lenguaje'. Utilización en un centro de rehabilitación. Rev. Neurol., 46: 454-460.

Salter K, Jutai J, Foley N, Hellings C and Teasell R (2006). Identification of aphasia post stroke: a review of screening assessment tools. Brain. Inj., 20: 559-568.

Shewan CM and Kertesz A (1980). Reliability and validity characteristics of the Western Aphasia Battery (WAB). J. Speech. Hear. Disord., 45: 308-324.

Shewan CM (1986). The Language Quotient (LQ): a new measure for the Western Aphasia Battery. J. Commun. Disord., 19: 427-439.

Theodoros D, Hill A, Russell T, Ward E, Wootton R (2008). Assessing acquired language disorders in adults via the Internet. Telemed J. E.

Health, 14: 552-559.
Viglïecca NS (2004). Neuropsychological tests abbreviated and adapted to Spanish speakers: Review of previous findings and validity studies for the discrimination of patients with anterior vs. posterior lesions. Rev. Neurol., 39: 205-212.
Viglïecca NS (2000). Towards a more agile and operative aphasia evaluation. Rev Neurol Arg. 25 (Suppl. Section Neuropsychol., p.52.
Viglïecca NS, Aleman GP (2010). A novel neuropsychological assessment to discriminate between ischemic and nonischemic dementia, J. Pharmacol. Toxicol. Methods., 61: 38-43.
Viglïecca NS, Peñalva MC, Molina SC, Voos JA (2011). Brief Aphasia Evaluation (minimum verbal performance): Concurrent and Conceptual Validity Study in Patients with Unilateral Cerebral Lesions. Brain Inj. 25: 394-400 (doi:10.3109/02699052.2011.556106).
Viglïecca NS (2010). Aphasia: what are we measuring? Revista de Neuropsiquiatría y Neurociencia Cognitiva. 9: 54-55. (See also http://www.saneurociencias.org.ar/descargas/Libro_Congreso_de_Neurociencia_IIRCN.pdf, 234-235, accessed Feb 23, 2011).
Vogt WP (2005). Dictionary of statistics and methodology: a nontechnical guide for the social sciences (3rd edition). Thousand Oaks, California: Sage Publications, Inc.
Wade DT, Hewer RL, David RM, Enderby PM (1986). Aphasia afterstroke: natural history and associated deficits. J. Neurol. Neurosurg. Psychiatry, 49: 11-16.
Williams JM, Shane B (1986). The Reitan-Indiana Aphasia Screening Test: scoring and factor analysis. J. Clin. Psychol., 42: 156-160.