Cross-cultural adaptation and validation of the Italian Aachener Aphasie Bedside Test (I-AABT), a tool for Aphasia assessment in the acute phase.

Muo, Rossella; Raimondo, Simona; Martufi, Francesca; Cavagna, Nicoletta; Bassi, Marta; Schindler, Antonio

Published in:
Aphasiology

DOI:
10.1080/02687038.2020.1819953

Published: 24/09/2020

Document Version
(Peer reviewed version when applicable)

Document License
Publisher rights policy

Link to publication

Please cite the original version:
Muo, R., Raimondo, S., Martufi, F., Cavagna, N., Bassi, M., & Schindler, A. (2020). Cross-cultural adaptation and validation of the Italian Aachener Aphasie Bedside Test (I-AABT), a tool for Aphasia assessment in the acute phase. *Aphasiology*. https://doi.org/10.1080/02687038.2020.1819953

CROSS-CULTURAL ADAPTATION AND VALIDATION OF THE ITALIAN AACHENER APHASIE BEDSIDE TEST (I-AABT), A TOOL FOR APHASIA ASSESSMENT IN THE ACUTE PHASE

Validation of the Italian version of the I-AABT

Authors: Rossella Muò, Simona Raimondo, Francesca Martufi, Nicoletta Cavagna, Marta Bassi, Antonio Schindler.

Corresponding author:

Simona Raimondo

SLP, SCU ORL, Città della Salute e della Scienza, Ospedale Molinette – Università degli Studi di Torino, Turin, Italy

Via Genova 3, 10127 Turin, Italy

+39 0116336786

e-mail: sraimondo@cittadellasalute.to.it

Co-Authors:

Rossella Muò

SLP - ASL Città di Torino, S.C. Recupero e Rieducazione Funzionale Territoriale – Università degli Studi di Torino, Turin, Italy

Via Silvio Pellico 28, 10125, Torino, Italy

Department of Speech and Language Pathology, Abo Akademi University, Turku, Finland

+39 3479043723

e-mail: rossella.muo@unito.it

Francesca Martufi

SLP - Centro Didattico “Libro Aperto”

Via Bignone 83 M, Pinerolo, Turin, Italy

+39 349-5763118

e-mail: francesca.martufi@gmail.com

Nicoletta Cavagna

SLP - SSD NPIA ASST Bergamo Ovest - Polo territoriale di Verdello

Via XI Febbraio 4, Verdello, Bergamo, Italy

+3903541870

e-mail: nicoletta_cavagna@asst-bgovest.it

Marta Bassi

PhD - Dipartimento di Scienze Biomediche e Cliniche “L. Sacco” – Università degli Studi di Milano, Milan, Italy

Via G.B. Grassi 74, 20154 Milan, Italy

+39 0239043526

e-mail: marta.bassi@unimi.it

Antonio Schindler

MD - Dipartimento di Scienze Biomediche e Cliniche “L. Sacco” – Università degli Studi di Milano, Milan, Italy

Via G.B. Grassi 74, 20154 Milan, Italy

+39 0239043526

e-mail: antonio.schindler@unimi.it

Abstract

Background. Early recognition of post stroke aphasia is thought to be crucial to plan effective rehabilitation. The Aachener Aphasie Bedside Test (AABT) is a widely-used tool for aphasia assessment in acute phase. The AABT was translated into Italian (I-AABT) in 2011, but its psychometric properties had not yet been studied.

Aims. The aims of the study were to assess the I-AABT 1) short-term test-retest and inter-rater reliability, 2) concurrent and construct validity, 3) responsiveness and 4) to develop preliminary Italian normative scores to stage aphasia severity.

Methods & Procedures: Participants were recruited from three Italian hospitals and divided into four groups: 1) patients with acute aphasia (PwAA; n=116), 2) patients with post-acute aphasia (PwPA; n=54), 3) patients with right-hemisphere damage (PwRHD; n=48) and 4) patients without neurological disorders (PwND; n=30). The I-AABT was administered to all participants. The Aachner Aphasie Test (AAT) was administered to assess concurrent validity. Spearman's correlations and Intraclass Correlation Coefficient (ICC) were used to calculate reliability. Concurrent and construct validity were assessed through Spearman's correlations between I-AABT and AAT subscales and Mann-Whitney test, respectively. Wilcoxon signed test was used to assess responsiveness. Both normalized z and T scores were calculated to produce Italian normative scores. Finally, ROC curves were drawn to determine the diagnostic accuracy of the I-AABT Comprehension and Production subtests.

Outcomes & Results. Test-retest and inter-rater ICCs were highly significant and strong (ICC > .837, ICC > .698, p<.001, respectively). Correlations between I-AABT and AAT Comprehension and Production subtests were significant and strong (r_s >.611, p <.001). Mann-Whitney test confirmed statistically significant difference between PwAA, and both PwRHD and PwND for I-AABT Spontaneous Language, Comprehension and Production subtests. A significant improvement in the I-AABT items Semantic, Phonemic and Syntactic Spontaneous Language, Oral praxis, Object identification, Automatic Language and Naming ($p <.05$) was detected after intensive language therapy. Preliminary conversion tables were devised to classify comprehension and production impairment levels. I-AABT's AUC-ROC values for Comprehension and Production parts were significant to detect patients with aphasia with a cut value of 139.5 (sensitivity = 72.9%, specificity = 79.5%9) and 94.5 (sensitivity = 75.2%, specificity = 74.4%.), respectively.

Conclusions. The I-AABT proved to be a reliable, valid and responsive tool for the assessment of aphasia in acute stroke patients; its use is recommended in everyday clinical practice.

Introduction.

Acute stroke is recognized as a medical emergency requiring immediate hospitalization and early intervention in highly specialized Stroke Units. The Stroke Unit team should ensure prompt medical and neurological care, timely rehabilitation (from 24-48 hours post acute stroke), and continuity of care over time. Above all, early rehabilitation might be crucial in reducing the functional limitations that are a common consequence of stroke (Powers et al., 2018).

Aphasia is one of such consequences, as approximately one-third of those affected by stroke show aphasic symptoms (Maas et al., 2012). Therefore, early involvement of patients in appropriate intensive rehabilitation therapy was shown by Worrall et al. (2017) to promote the development of useful strategies to enhance communication and, consequently, to both reduce patients' isolation and increase their participation in the rehabilitation process. Since early identification and diagnosis of language deficits in acute care settings are recognized as important steps towards maximizing rehabilitation gains, assessment procedures and tools have progressively become more important over the last few years (Brady et al., 2012; Coleman et al., 2017; Royal College of Physicians, 2012; Salter et al., 2006). The treatment for post-stroke aphasia should start as soon as possible, even though consistent evidence for a beneficial effect of early language therapy has not yet been published (Brady et al., 2012; Brady et al., 2016; Godecke et al., 2014).

Aphasia assessment in the acute phase.

During the acute post-stroke phase, patients are usually characterized by a rapid evolution of both their clinical condition and their neuropsychological impairment as well as experiencing remarkable fatigue, with fluctuation of their cognitive abilities. Nevertheless, the assessment of communication disorders is generally recommended from the third day post stroke, whenever the patient's state of consciousness allows it, in order to establish a well-timed treatment plan and maximize recovery gains (Godecke et al., 2011; National Clinical Guideline Centre, 2013). A short bedside examination is reported to be useful in order to understand the severity of the language impairment, to monitor recovery and the changes in the patient's clinical condition, to plan the rehabilitation program and to provide prognostic insight into the recovery rate (Bley, Wagner, & Berrouschot, 2002; Salter et al.,2006; Spreen & Risser, 2003).

Assessment of aphasia in acute settings requires appropriate tools. Complete language test batteries, such as the Aachener Aphasie Test (AAT) (Huber, Poeck & Willmes, 1984), are usually too complex and excessively time-consuming to be performed with severely impaired patients, just as early post-stroke individuals generally are. Within these settings, the assessment should include tools that are short, simple, quick, standardized andeasy to be administered at the bedside. Moreover, the assessment should be performed by specialized clinicians (Boulanger et al., 2018).

In Europe, several standardized tools to assess post stroke acute language impairment and communication abilities are available. Nevertheless, it is difficult to accurately distinguish between screening and assessment tools. Screening tools are supposed to be short and straightforward enough to be administered by different health professionals (for example, nurses). The main aim of a screening tool should be to detect an impairment (for example a language impairment), in order to select those patients who are candidates for a more accurate assessment provided by a specialized health professional. For people with aphasia in the acute stage, a screening tool is needed to reveal language and communication impairment reliably and accurately in order to select those patients requiring a more complete assessment by a speech-language therapist (SLT) (National Health and Medical Research Council (NHMRC) Centre for Clinical Research Excellence (CCRE) in Aphasia Rehabilitation, 2019). Aphasia screening tools include the Frenchay Aphasia Screening Test (FAST; Enderby & Crow, 1996), the Language Screening Test (LAST; Flamand-Roze et al., 2011), the Mississippi Aphasia Screening Test (MAST; Nakase-Thompson, 2004), the Ullevaal Aphasia Screening Test (UAS; Thommessen et al., 1999), and the Acute Aphasia Screening Protocol (AASP;

Crary, Haak, & Malinsky, 1989). These tests are often used in clinical practice as screening tools to distinguish between stroke patients with or without aphasia. However, during the acute stages SLTs often use many of these tests not only for screening purposes, but also for an early assessment of language and communication skills, as they include items appraising motor functions, speech production, spoken and written language production and comprehension, and communication-related cognitive skills.

Unlike screening tests, assessment tools in the acute post-stroke phase should be more complete so as to provide SLTs with relevant information for the diagnostic process, set the main early rehabilitation aims, and give suggestions for counseling activity.

Italian tools for aphasia assessment.

Several tools for the assessment of aphasia during the post-acute and chronic phase are available for the Italian population. A recent review paper (Ivanova & Hallowell, 2013) mentions the Italian version of the Aachener Aphasie Test (Luzzatti, Willmes & De Bleser, 1996), the Bilingual Aphasia Test (Paradis, Canzanella & Baruzzi, 1987), the Boston Naming Test (D'Agostino, 1985), the Communicative Abilities of Daily Living (CADL, Pizzamiglio et al., 1984), the Clinical test of Lexical Retrieval and Production (Test clinici di ricerca e produzione lessicale. Novelli et al., 1986, and the Italian Battery for Assessing Aphasic Deficits (Batteria per l'Analisi dei Deficit Afasici, B.A.D.A. Miceli et al., 1994). The former list can be updated with more recent tools, like the Neuropsychological Exam for Aphasia (Esame Neuropsicologico per l'Afasia, E.N.P.A., Capasso & Miceli, 2001), the Italian version of the American Speech Language and Hearing Association - Functional Assessment of Communication Skills for Adults (I-ASHA-FACS, Muò et al., 2015), the Italian version of the Communication Outcomes After Stroke Scale (COAST-IT) and Carer-Communication Outcomes after Stroke Scale (Carer-COAST-IT) (Bambini et al., 2016), the Italian version of the Kagan's Scales Measure of skills in Supported Conversation (I-MSC) and Measure of

Participation in Conversation (I-MPC) (Muò et al., 2018), and the Stroke and Aphasia Quality of Life Scale (SAQOL-39, Posteraro et al., 2004). Nevertheless, there is a paucity of standardized tools designed for the assessment of aphasia in the acute phase. In 2009, the "*Esame del linguaggio al letto del malato*" (ELLM) was published (Allibrio et al., 2009). In 2014, the Frenchay Aphasia Screening Test was translated into Italian (I-FAST) and the validation process is still ongoing (Bonelli et al., 2015). Finally, in 2011, the AABT was translated and adapted into Italian (Iosa et al., 2011), but it has never been validated. To the best of our knowledge, no other published tools are available.

The Aachener Aphasie Bedside Test.

The Aachener Aphasie Bedside Test (AABT; Biniek et al., 1992) was originally developed for this purpose. The AABT is a short tool, characterized by easy-to-use administration at the bedside. The original version of the AABT was developed in German and published in 1991 (Biniek et al.) with the aim of assessing communication and language skills in severe neurologic patients. In fact it can be administered in the acute phase with minimum prerequisites of cognitive and language skills. In comparison to other tools that are only focused on language impairment, the AABT allows the examiner to score nonverbal communication skills as well. In this way, it may be a useful tool to informally distinguish people with both severe language impairment and strongly reduced nonverbal communication skills from people with severe language impairment but with either preserved or mildly reduced nonverbal communication skills.

The psychometric properties of the AABT were established by 1) repeated examination of 82 acute stroke patients, rated by 20 raters on the basis of 10 videotapes; 2) repeated examination of 28 participants with chronic aphasia at three 2 day intervals between assessments; and 3) parallel examination of 47 participants affected by chronic aphasia with both the AABT and the AAT administered on the same day. Objectivity, reliability and validity of the AABT were highly rated, so that it can be considered an adequate tool to assess aphasia in the acute phase. Data on the 82 acute stroke patients showed that an initial prognosis of language disorder can be made as early as the fourth day post stroke (Biniek et al., 1992).

A recent review by El Hachioui et al. (2017) underlined that several tests had never been validated in stroke patients with and without aphasia, and they recommended that future studies should focus on better validation of the available acute aphasia screening and assessment tests in a large stroke population. Although the AABT can be considered a well-known test for aphasia, widely used in clinical practice, no studies addressing its validation among stroke patients with or without aphasia could be found. The authors concluded that the AABT psychometric properties had not been studied properly and they recommended further studies to focus on a better psychometric investigation in a large stroke sample (El Hachioui et al., 2017).

The importance of the present study lies in the fact that reliable and valid assessment tools, tested on large acute stroke samples, are needed in Italy, in order to assess the aphasia characteristics during the acute phase, provide data for both the diagnosis and the prognosis and plan the rehabilitation path and strategies. Moreover, a validated test, already available in German, will represent a vital contribution to advancing cross-linguistic studies and international collaborations.

Aims of the study.

As very few validated assessment tools for aphasia under acute stroke conditions are currently available in Italian, the main aim of this study was to analyze the major I-AABT psychometric properties, including inter-rater and short term test-retest reliability, and both concurrent and construct validity. Since no substantial changes were introduced during the translation process, the authors expected to replicate the original AABT properties. The second aim was to study the I-AABT responsiveness. The authors expected the I-AABT to show good sensitivity to change in monitoring recovery and changes after intensive language therapy. Finally, the authors aimed to develop preliminary Italian normative scores that could be applied in clinical practice to classify patients' aphasia severity and monitor patients' linguistic performance over time.

Methods

The study was carried out according to the Declaration of Helsinki. All participants enrolled in the study gave their written informed consent, when their clinical conditions were satisfactory to understand and sign. Techniques to support conversation were used accordingly both to describe the characteristics of the research project, including the aims of the study and the demands for the participants' direct involvement, as well as enabling participants to express their refusal or consent to participation. As neurological deficits related to acute stroke often cause inability to give consent, the responsibility was given to the patients' main partner or relative for all those people who were unable to receive information, according to their capacity of understanding. All procedures used in this study are not invasive and the proposed assessment is commonly used in clinical practice so participants are not exposed to any significant risks.

Participants.

For each phase of the study, different groups of patients were consequently recruited from either two large Italian acute care hospitals (the "Città della Salute e della Scienza - Molinette" Hospital in Turin and the "Luigi Sacco" Hospital in Milan) or an Italian post-acute rehabilitation unit ("Casa di Cura Privata del Policlinico" in Milan). Overall, data from 116 participants with acute aphasia, 54 participants with post-acute aphasia, 48 participants with right hemisphere damage and 30 control participants without neurological disorders, were collected. Demographic and clinical characteristics of participants included in each phase of the study as well as the detailed number of subjects included for each study, are shown in Table 2.

(Table 2 about here)

More details about the inclusion criteria of the participants for each part of the study are presented in the procedure section.

All aphasia diagnoses were made by a specialized physician (neurologist or phoniatrician) on the basis of the clinical examination and neuroimaging results (CT and MRI), in accordance with an SLT (Powers et al., 2018). Patients were considered suitable to enter the study if their state of consciousness allowed alertness for at least ten consecutive minutes.

The AAT and AABT assessment was performed by three SLTs, with over five years' experience in aphasia management.

Materials.

The Italian version of the Aachener Aphasia Bedside Test (I-AABT; Iosa et al., 2011) was administered to all the recruited patients. The Italian version of the Aachener Aphasia Test (AAT; Luzzatti, Willmes & De Bleser, 1996) was used to assess concurrent validity.

The AABT can be administered from the third day post acute onset of brain damage. Its use is recommended during the entire acute phase, defined as the period from the acute event until 4 to 6 weeks later. Minimum prerequisite for administration is the presence of 10 consecutive minutes of alertness. Its administration takes between 15 and 40 minutes, but interruptions reducing the administration time up to a maximum of 15 minutes are allowed, whenever necessary. The AABT contains items to assess both aphasia and dysarthria symptoms. As the acute phase is characterized by rapid continuous changes in both the patients' clinical conditions and communication skills, the AABT aims to investigate only the three main features that the scientific community considers as the most significant for this setting: 1) the presence of communicative intention and both the presence and the effectiveness of spontaneous speech, 2) comprehension of short, simple and contextual sentences, and 3) basic speech and language abilities, suchasautomatic language, repetition and naming. Further language skills that are significant for the assessment of aphasia (such as reading, writing, describing an image or summarizing a text) were not included in the protocol, because it would have increased the administration time.

The AABT was translated and culturally adapted into Italian in 2011 (I-AABT; Iosa et al., 2011). The adaptation process required some changes related to the Italian context.

Both the AABT and the I-AABT consist of three main parts: 1) preliminary part (Global communication behaviour in greetings and farewell and Spontaneous speech), 2) comprehension, and 3) production. Each part is divided into subtests.

Preliminary part. Global communication behaviour in greetings and farewell and Spontaneous speech.

In the *Global communication* section, the examiner is asked to observe the patient's global communicative behaviour in greetings and farewell and to evaluate the adequacy of the patient's characteristics on the following items: Eye contact, Shaking hands, Verbality, Facial expression, Gesture and Emotional response. Each item from the *Global communication* section is scored between 1 and 4 points, depending on whether the behaviour is absent, hard to stimulate, easy to stimulate, or spontaneous, respectively.

In the *Spontaneous speech* section, the examiner is asked to lead a conversation about either the patient's family, job or hobbies, or about the history of their disease. The clinician is to evaluate Communicative behaviour, Phonological level, Semantic level, Morphosyntactic level, Automatic language and Articulation and prosody. The scoring system of this section ranges from 0 to 5 for each of the 6 parameters included (higher scores are indicative of better skills).

Comprehension.

The *Comprehension* section investigates the patients' ability to understand simple requests, such as performing some movements with either their head or eyes (Simple request), or with their mouth (Oral motor skills and oral-facial apraxia), or identifyingfour objects belonging to the same semantic category (Identification of objects). Each of the three subtests includes 10 items, with each item being scored firstly as pass or fail (score 1 or 0, respectively) and subsequently, the score is defined by the way the behaviour is elicited. Depending on whether the patient requires either a verbal request or repetition of the verbal request, imitation, or repetition of the imitation, the score ranges from 5 to 1 respectively. The total score of the *Comprehension* section ranges from 0 to 150 (higher scores are indicative of better skills).

Production.

The *Production* section is formed by three subtests: 1) Articulation and phonation, assessing both oral diadochokinesia, and the Maximum Phonation Time (MPT); 2) Automatic language and repetition, evaluating the automatic linguistic function through tasks such as producing the musical scale, counting, and repeating greetings; 3) Naming objects.

Both the Automatic language and repetition subtest and the Naming objects subtest comprise 10 items. Each item can be scored from 1 to 5 respectively, based on the way the answer is elicited (verbal request, repetition of the verbal request, imitation, repetition of the imitation) or if the patient names or makes a description of the object or needs a semantic or phonemic cue, or presents dysarthria. The total score of the *Production* section, excluding the subtest Articulation and phonation which is strictly related to speech, ranges from 0 to 100 (higher scores are indicative of better skills).

In the process of adaptation to the Italian language, the song "*Hänschenklein*" used in the original subtest Automatic language and repetition was substituted with the Italian folk song, "*Quel mazzolin di fiori*". The expression "*Nachsprechenbzw. Mitsprechen von Floskeln – Guten Morgen*" was replaced with the request to repeat "*Buongiorno*" (i.e. "*Good morning*"). The objects of the Naming subtest were replaced by others according to their frequency of use in the Italian language.

Procedures.

The study consisted of four phases: 1) reproducibility analysis, 2) validity analysis, 3) responsiveness analysis, according to the COSMIN taxonomy (Mokkink et al., 2010) and 4) ROC analysis and preliminary normative scores. The procedures used for each phase are summarized in table 1.

(Table 1 about here)

The demographic and clinical characteristics of participants recruited for each phase are summarized in table 2.

Phase 1 - Reproducibility of the I-AABT analysis.

The aim of the first phase of the study was to evaluate the I-AABT short term test-retest and inter-rater reliability. For short term test-retest reliability analysis and for the inter-rater reliability analysis two different groups of patients with aphasia in the acute phase (from 3 to 7 days post stroke) were recruited. During the acute phase, quick clinical changes can often be observed, and patients' learning abilities proved to be impaired. The authors thus considered it a priority to repeat the assessment within a short period of time in order to reduce the influence of spontaneous recovery, compared to the need for separating the two assessments by a considerable amount of time to avoid learning.

Phase 2 - Validity of the I-AABT analysis.

The aim of the second phase of the study was to calculate if the I-AABT assesses the construct it claims to measure (validity). Both concurrent and construct validity investigation were considered. Concurrent validity was defined as the way the I-AABT scores relate to results obtained from other validated or gold standard measures. In order to study concurrent validity, patients with aphasia in the post-acute phase (between 15 days up to six months after the acute event) were recruited. Construct validity is defined as the I-AABT capacity to identify aphasic symptoms. In order to analyze construct validity, the I-AABT data from patients with left-hemisphere stroke and aphasia in the acute phase were compared with the I-AABT data from right-hemisphere stroke patients in the acute phase and acute in-patients without cerebral lesions.

Phase 3 - Responsiveness.

Responsiveness refers to the property of the assessment tool to detect changes over time in the construct to be measured. In order to analyze the I-AABT responsiveness, patients with left hemispheric stroke in the post-acute phase were assessed pre and post a period of intensive speech and language therapy.

Phase 4 – Diagnostic accuracy and normative scores calculation.

In order to determine diagnostic accuracy of the I-AABT all patients with aphasia in the acute phase (n=116), patients with right-hemisphere damage (n=48) and participants without neurological disorders (n=30) were included. Preliminary normative scores for I-AABT scales measuring comprehension and production performance were computed based on the overall sample of patients with aphasia in the acute phase who were recruited in the present study. For participants involved in short-term test-retest and inter-rater reliability analyses (n=46), the first I-AABT assessment was taken into account.

Statistical analysis.

Statistical analyses were conducted with SPSS 25. Descriptive analysis was used for demographical and clinical data. Either mean and standard deviation or median and minimum-maximum range were calculated for I-AABT subscales, as appropriate. The short term test-retest and inter-rater reliability were assessed with both Spearman's correlation coefficient (r_s) and Intraclass Correlation Coefficient (ICC). A two-way random-effects, absolute agreement model, single measure ICC was calculated for inter-rater reliability analysis. A two-way mixed-effects, absolute agreement model, single measure ICC was used for short term test-retest analysis. For concurrent validity analysis, Spearman's correlation coefficient between I-AABT and AAT ratings was calculated. The correlation was considered strong for r_s values greater than *0.5*, moderate for r_s values ranging between *0.3* and *0.5* and weak for r_s values of less than *0.3* (Vittinghoff et al., 2005). The Mann-

Whitney test was used to study construct validity. The I-AABT scores obtained from patients with aphasia in the acute phase were compared with the ratings obtained from both the patients with right-hemisphere stroke and the participants without neurological diseases. The Wilcoxon signed test was used to assess responsiveness. To calculate preliminary normative scores in aphasia severity assessment with I-AABT, the following subscales were taken into consideration: Simple requests, Oral praxias, and Objects identification for Comprehension, Automatic language and repetition, and Naming, for Production. After inspection of skewness and kurtosis values, participants' raw scores in each subscale were first transformed into normalized *z* scores (Chiorri, 2011), that is, standard scores were computed from the percentile ranks of raw scores. Percentile ranks were calculated using the following formula (Chiorri, 2011): $PR = (f_{l\_cum} + 0.5f) / n$, where f_{l_cum} is the cumulative frequency of the score immediately below the score of interest, f is the frequency of the score of interest, and n is the total sample size. Normalized *z* scores were then converted into T scores (with M=50 and SD=10). To facilitate interpretation, Receiver operating characteristic (ROC) curves with area under the curve (AUC), sensitivity and 1-specificity were created to determine the diagnostic accuracy of the Comprehension and Production part of the I-AABT. In case an AUC was significantly larger than 0.5, the optimal cut-off point was determined. Means, standard errors (SD), and 95% confidence interval are reported.

Results

I-AABT results.

The I-AABT scores in the different samples recruited in the study are summarized in Figure 1 and presented in greater detail in Appendix 1, Appendix 2 and Appendix 3.

(Figure 1 about here)

The I-AABT results from the three Comprehension subtests (Simple requests, Oral Motor skills and Identification of objects) showed the widest distribution in the group of participants with acute aphasia, followed by the group of post-acute aphasia patients. The groups of participants with right-hemisphere damage and participants without neurological disorders showed narrower distributions that were at or close to ceiling. The same decreasing trend can be observed for the I-AABT results from the Production section subtests Automatic language and repetition and Naming objects. The presence of some outliers and far outliers can also be noticed for the Identification of objects and Naming objects subtests, in particular. Although a slightly wider distribution can be observed for participants with aphasia, the results obtained from the Oral diadochokinesia and MPT subtests appear to be more equally distributed among the four groups of participants and as a consequence of the characteristics of the items, no ceiling effect is observed.

Data from the I-AABT Preliminary part results (Global communication behaviour in greeting and farewell), and Spontaneous language reported in Appendix 1 and Appendix 2, suggested that all the participants maintainedspontaneous or easily stimulable pragmatic abilities. Patients with right-hemisphere stroke obtained slightly less severe scores than participants with aphasia. On the contrary, most patients with aphasia showed moderate-severe language impairment. In fact, they could not communicate the message they wanted to and needed some support from their communication partner, produced several semantic paraphasias and/or had several anomias, several phonemic paraphasias, and their sentences were mainly either short and incomplete or long with interruptions and with several syntactic errors. Articulation and Automatic language seemed to be more preserved as patients showed minimal speech impairment and only some stereotypes and/or perseverations. As expected, right-hemisphere damaged patients' median scores reached the maximum score possible (median = 5), although the minimum-maximum range scores showed that some patients obtained lower results.

AAT results.

The AAT scores were available for patients with aphasia recruited for phases 2 and 3 (Appendix 2 and Appendix 3). The AAT T-scores showed a generally moderate impairment for the participants recruited for phase 2 (T-scores ranged from 46 to 52), while participants recruited for phase 3 were generally more impaired and showed a generally severe linguistic impairment (T-scores ranged from 26 to 43).

Phase 1 - Reproducibility analysis.

Short-term test-retest and inter-rater reliability analyses are shown in Table 3.

(Table 3 about here)

The short term test-retest reliability results showed highly significant, strong Spearman's correlation and ICCs for all the I-AABT items ($r_s > .841$, $p < .001$; ICC > .837); the inter-rater reliability analysis confirmed highly significant, strong correlations and ICC for all the I-AABT items ($r_s > .791$, $p < .001$; ICC > .698). Among the observations stated for the preliminary part, the item Emotional responses showed perfect test-retest and inter-rater reproducibility and the item Gestures showed a perfect inter-rater reproducibility.

Phase 2 - Validity analysis.

In order to analyze concurrent validity, Spearman's correlations between the I-AABT Comprehension and Production subtests and the AAT Token test, Comprehension and Naming subtests were performed. Results are shown in Table 4. All the correlations were strong (r_s>.611, p < .001).

(Table 4 about here)

In order to assess construct validity, the I-AABT results obtained from the participants with aphasia were compared with the results obtained from either the participants without neurological disorders or the participants with right-hemisphere brain damage (Table 5).

(Table 5 about here)

Results showed generally high significant differences for all language items of the I-AABT (Spontaneous Language, Comprehension and Production subtests, p < .002), and for the Maximum Phonation Time (p = 0.05). The I-AABT Preliminary part showed significant differences for all the items (p < .008), excluding Shaking hands.

Comparisons between the scores of participants with aphasia and the participants with right-hemisphere brain damage scores showed significant results for Verbality and Emotional responses (p < .021) and high significant differences for all the items of the I-AABT Spontaneous Language, Comprehension and Production (p < .005). The Maximum Phonation Time did not reach statistical significance.

Phase 3 - Responsiveness analysis.

The responsiveness results are shown in Appendix 3. Comparisons between the results pre- and post speech and language treatment assessed with the AAT were all significant (p < .021) with the only exception of the Reading and Writing item (p =.092). Although our I-AABT median data reported a general improvement between the scores pre- and post speech and language treatment, comparisons were significant only for the I-AABT Spontaneous Language Semantic, Phonemic and Syntactic items ($p < .037$), for the I-AABT Comprehension Oral praxis and Objects identification ($p < .043$) and for the I-AABT Production Automatic Language and Naming ($p < .029$).

Phase 4 - Normative scores calculation.

Table 6 illustrates the descriptive statistics of the comprehension and production subtests administered to the overall sample of participants with aphasia in the acute phase (n=116).

(Table 6 about here)

Except for skewness of Naming and kurtosis of Simple Requests, most skewness and kurtosis values were not within the acceptable ±2 range, calculated as the ratio of the parameter estimate divided by its standard error (George & Mallery, 2010). It was thus necessary to normalize z scores prior to T score transformation. The conversion from raw into T scores across the entire 0-50 possible range of raw scores is detailed in table 7. Based on the standard deviation (SD=10) of the T scores distribution, a moderate ($45 < T \leq 55$) impairment range can be identified. Scores falling above this range ($T>55$) can be an indication of mild or no impairment whereas scores falling below this range ($T \leq 45$) can be an indication of severe impairment.

(Table 7 about here)

I-AABT's AUC-ROC values for Comprehension and Production parts are significantly larger than 0.5 and are reported in Figure 2. The cut-off value for the I-AABT Comprehension part was 139.5 showing a sensitivity of 72.9% and a specificity of 79.5%; the cut-off for I-AABT Production was 94.5 with a sensitivity of 75.2% and a specificity of 74.4%.

(Figure 2 about here)

Discussion

In post-stroke patients, early recognition of aphasic symptoms is highly recommended in order to guarantee adequate referral and timely rehabilitation (Powers et al., 2018). Early diagnosis of language impairment requires a bedside administration of specifically structured tests. Since 1991, the AABT has been successfully introduced into clinical practice for its properties in recognizing patients with severe aphasia and it has been recently translated and adapted for the Italian population (Iosa et al., 2011). Our study aimed to explore the I-AABT psychometric properties, accepting the recommendation by El Hachioui et al. (2017) for research studies that included large stroke populations for the AABT validation. Following El Hachioui et al. suggestions, other methodological recommendations were also carefully considered in the study. In particular, a reliable reference diagnosis was included for all participants recruited and patients were enrolled consequently so that they were representative of the general stroke population. Detailed information on the number of participants with and without aphasia were reported to provide a correct classification, and a good description of the I-AABT subtests was added, to reduce the risk of bias as much as possible.

Moreover, Rohde et al. (2018) commented that no study among those examined in their recent review, had reported diagnostic data on patients with or without aphasia in stroke populations. This study offers useful data to meet this need, since participants with right hemisphere stroke were included.

Comparison of the I-AABT scores in different population.

Results obtained from the patients with right-hemisphere stroke were in line with the well-known characteristics of this population (Lehman Blake, Frymark & Venedictov, 2013), whose linguistic function is generally preserved. Not all participants without brain lesions reached the maximum score for all the I-AABT subtests, showing that the I-AABT does not present a relevant ceiling effect. Particularly, participants without brain lesions showed a broad range of performance on the I-AABT Simple requests, Oral motor skills, and Automatic language and repetition subtests. In contrast, performance on the I-AABT Identification of objects and Naming objects subtests was at or near ceiling in this group. Together with impairment of participants with aphasia on these two subtests, these findings underscore the utility of I-AABT Identification of objects and Naming objects subtests for specifically detecting an underlying language impairment. This datum is of clinical interest as it has never been reported before and suggests that language performance in patients in acute care settings even without brain lesions can be suboptimal. This finding is not surprising, if we consider that participants included are mainly aged or elderly people recruited in a hospital setting. It is known that hospitalization may affect cognitive performance in some elderly people and that some difficulties in cognitive domains, such as attention and memory, can contribute to a mild impairment in linguistic function (Covinsky et al., 2003). That is why the I-AABT results that are suggestive of a mild impairment should be interpreted with caution, as different factors (linguistic impairment, other cognitive domains impairment) may contribute to low scores.

While providing data on several aspects, data on reading, writing and written comprehension abilities were not supplied. The data on more complex linguistic abilities can be useful to distinguish mild language impairment, consequently, the I-AABT should not be used or it should be used together with the other measures, in patients with mild aphasia. Instances of dysarthria and apraxia can be informally observed while administering the I-AABT, but they need to be investigated using more detailed and specific assessment tools (Richter, Wittler & Hielscher-Fastabend, 2006).

Reproducibility of the I-AABT.

The I-AABT short term test-retest and inter-rater reliability results were satisfactory. In line with the expectations, the short term test-retest scores were higher than the inter-rater ones. Our short term test-retest reliability data replicated Biniek et al. (1992) reliability results. The inter-rater reliability results cannot be compared to the original study results, as those data were not analyzed in the original study. However, those data should be interpreted cautiously because they were based on two assessments performed during the same day and two sources of unreliability could affect the results, i.e. fluctuations in the patients' behaviour and possibly different views of the raters. To reduce the impact on the patients' fluctuation on performance, the raters checked that no significant clinical changes had occurred between the two assessments. In order to reduce a potential learning bias from repeated assessments, two different groups of patients for the short term test-retest and for the inter-rater reliability analysis were recruited.

Concurrent and construct validity of the I-AABT.

The concurrent validity analysis results confirmed strong correlations between the I-AABT and the AAT scores. For both Comprehension and Production subtests strong correlations between scores in the I-AABT and in the AAT werefound. Strong correlations were also found between the I-AABT Comprehension subtest and the AAT Naming subtest and between the I-AABT Production subtest and the AAT Token Test and Comprehension subtests. Those data point out the limitation any screening test has in distinguishing between comprehension and production impairments. Clinical reality confirmed that these impairments often appear together in the acute stage. Our results suggest stronger correlations compared to Biniek et al. (1992), as the authors reported correlations ranging from .25 to .82. Those results should be interpreted with caution, because the time window for a proper usage of both tests, I-AABT and AAT, never overlaps. Nevertheless, the authors decided to follow the same procedure as in the original study by Biniek et al. (1992) since no other test could be used to compare the I-AABT results and no gold standard test for the acute phase was available for the Italian population.

The construct validity analysis results confirmed a significant difference in the I-AABT scores in patients with aphasia compared to patients without neurological conditions. Score comparisons between the patients with aphasia and the patients with right brain damage showed significant differences only in the I-AABT Spontaneous Speech, Comprehension and Production sections. Those results were expected, as these I-AABT sections mainly explored language abilities, which are known to be compromised by aphasia. On the other hand, the pragmatic aspects that are considered in the I-AABT Global communication subtest are usually not only linked to the left hemisphere damage and indeed difficulties after a right brain damage have been reported (Lehman Blake, Frymark & Venedictov, 2013). Comparisons with Biniek et al. (1992) results were not possible as this aspect had not been investigated in the original study.

Responsivness of the I-AABT.

The I-AABT also proved to be sensitive in detecting changes over time (responsiveness), in patients with aphasia assessed pre and postlanguage therapy in the subacute phase, when improvement is expected. Administration of AAT pre and post therapy showed a significant improvement. The I-AABT also showed a general improvement. In particular comparisons were significant for the I-AABT Spontaneous Language Semantic, Phonemic and Syntactic items, for the I-AABT Comprehension Oral praxis and Objects identification and for the I-AABT Production Automatic Language and Naming. To our knowledge, no other studies investigated the adequacy of the AABT to monitor recovery after intensive language therapy in the subacute phase.

Diagnostic accuracy and normative scores of the I-AABT.

Diagnostic accuracy of the I-AABT showed AUC values higher than 0.8 for both the Comprehension and Production part. AUC values higher than 0.8 are often considered to be of moderate accuracy (Swets, 1988), thus suggesting that the I-AABT should be used with caution to distinguish patients with and without aphasia.

The I-AABT proved to be an adequate tool for patients with severe aphasia, particularly as it is made up of short simple tasks. Moreover, it proved to be a useful tool to highlight the patients' best communication abilities, including both the linguistic and the extra-linguistic abilities. During the assessment the examiner can provide both items at different levels of difficulties and different types of stimulation. The examiner is thus allowed to take into consideration all the possible reactions and the preferred communication strategies to obtain a response to the requested task. Those data are useful for the planning of the rehabilitation strategies. In planning rehabilitation for patients with aphasia in the acute phase, the examiner may also take into consideration the preliminary conversion table of raw scores into T-scores. This might be useful in clinical practice as it allows the comparison of single patients' scores with data from a normative sample, identifying impairment profiles in specific comprehension and production abilities, and tracking performance changes over time.

Limitations of the study.

The main limitation was the inability to carry out a secured blinded assessment, as suggested by El Hachioui et al. (2017), thus an assessment bias due to the repeated measures by the same examiner is to be taken into consideration while examining the results. Moreover, the authors chose to repeat the assessment twice during the same day in order to reduce the impact of spontaneous recovery on the results which could interfere with the second assessment. However, this choice may facilitate the patients' spontaneous learning, thus improving their performance on the second administration of the test. To avoid this kind of assessment bias, the production of parallel forms of every test is recommended. Unfortunately this is not easy to be found for speech and language assessment tests. The authors largely discussed the possibility of scoring the patients' performance through a video observation, but performing a video recording would have raised several other ethical and practical complications related to the specific setting of the study and the severe clinical conditions of the patients. In particular, performing a video recording in the acute-phase setting could have had an impact on both the patient's performance and the possibility to recruit all the patients within the time period planned for the study, because of the possible need to exclude some of the most severe patients, thus affecting the results. Moreover, it would have been necessary to carefully consider more relevant ethical issues related to the need to obtain informed consent from patients with severe communication difficulties and cognitive deficits.

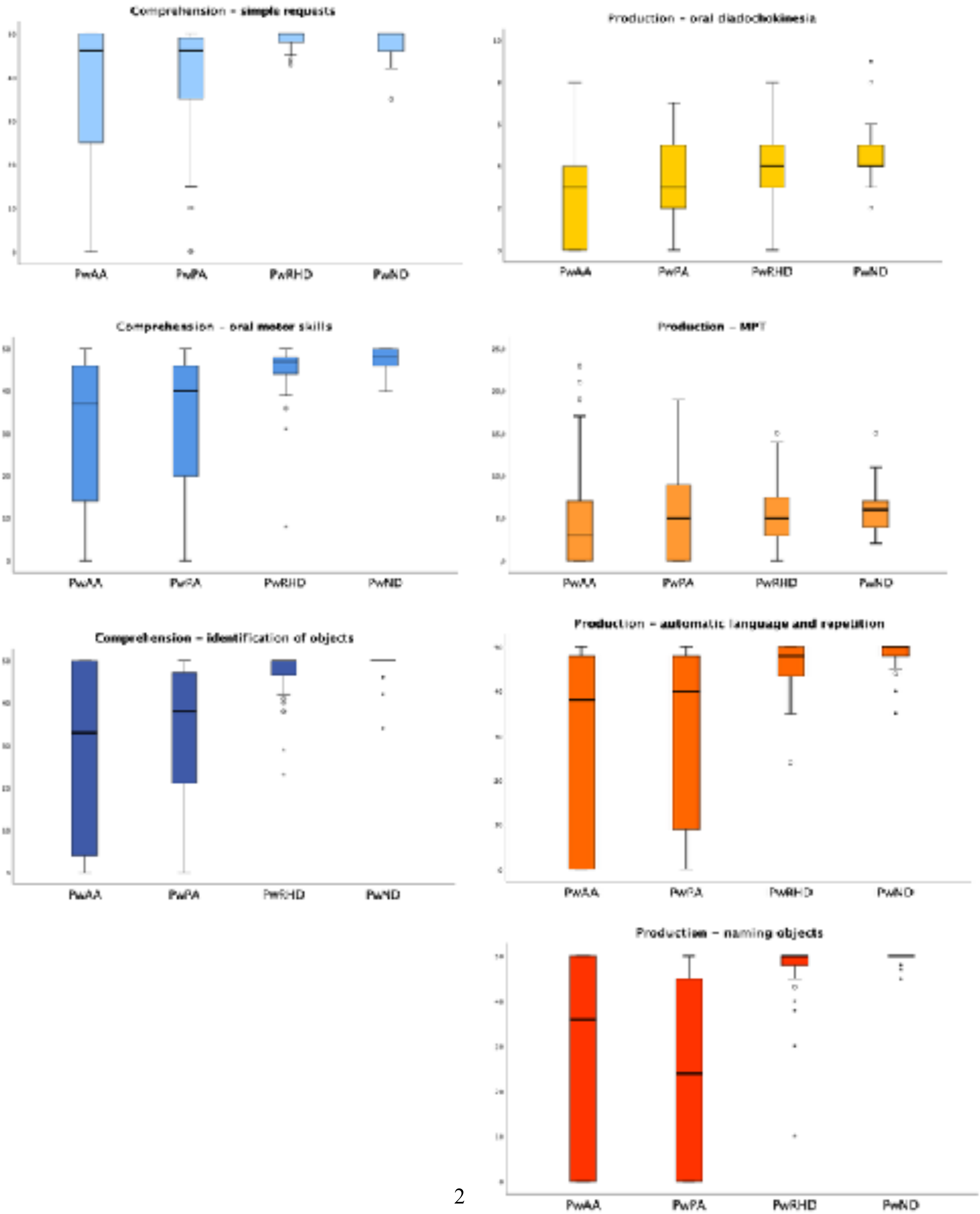
Another limitation concerns recruitment settings. It would have been ideal to study responsiveness documenting improvement in the early post-stroke stage. However, hospital stay duration for patients in the acute phase is generally too short to allow this kind of study and the organization of the rehabilitation path did not even allow a small group of patients to be followed for a period after discharge.

A final limitation concerns sample size. Even though it can be considered as large in light of the patient population under investigation, a higher number of participants should be enrolled in future studies in order to devise more accurate normative parameters for the identification of impairment levels. In addition, the use of stratified sampling techniques would allow researchersto weigh normative data based on socio-demographic characteristics - such as age, gender and education level - that may affect linguistic performance thus helping to better tailor rehabilitation intervention. Until then, the present normative scores – though helpful in current clinical practice – should be considered with caution.

Conclusions.

Results appear promising, supporting the use of the I-AABT in clinical practice with Italian patients affected by acute aphasia. The I-AABT proved to be a reliable and valid tool for a standardized assessment of aphasic symptoms in patients with acute stroke. The I-AABT also proved to detect important changes over time (responsiveness) in the construct measured and to distinguish the patients with aphasia from the patients without aphasia (either participants without cerebral lesions or patients with right hemisphere stroke). Due to its characteristics, the I-AABTmay be useful in Italian clinical practice, where there are currently no sufficient standardized tools for assessing aphasia in the acute phase, especially for patients with more severe stroke outcomes. The I-AABT may be of use in allowing an initial diagnostic hypothesis in the acute stage to indicate the first therapeutic approach and to monitor the evolution of patients' communication. The application of a standardized assessment tool in daily clinical practice, as well as in epidemiological, efficacy and outcome research is therefore recommended since it could enable the comparison of results.

Figure 1 - Distribution of the I-AABT scores for Comprehension subtests (simple request, oral motor skills, identification of objects) and Production subtests (oral diadochokinesia, MPT, automatic language and repetition, naming objects) measures in patients with acute aphasia, post-acute aphasia, right hemisphere brain damage and participants without neurological disorders.

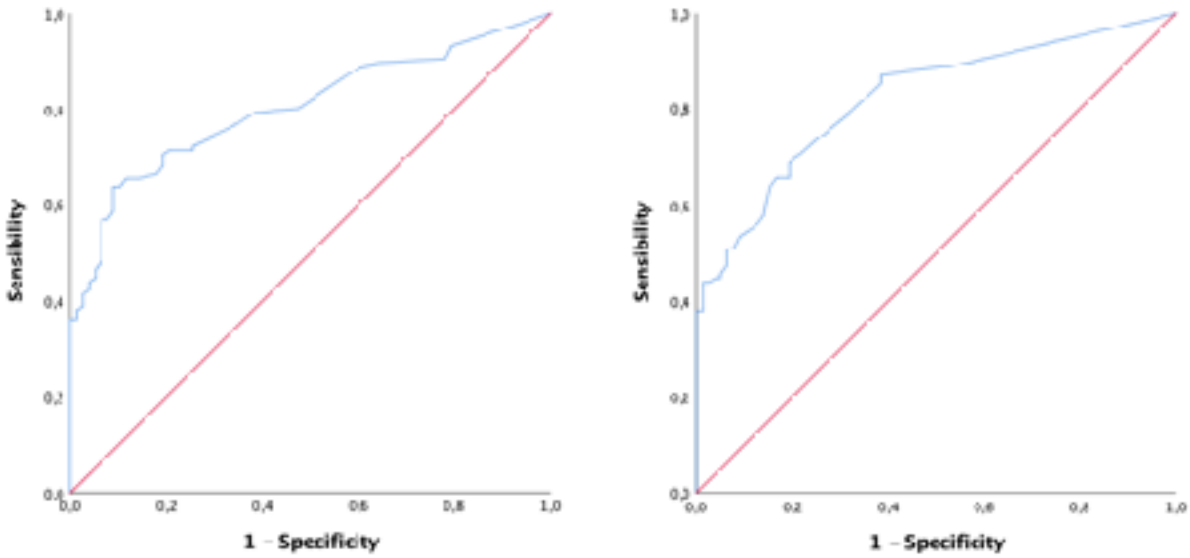


Boxes = interquartile range, whiskers = range not including outliers, thick horizontal lines = medians, PwAA = participants with acute aphasia (n=116), PwPA= participants with post-acute aphasia (n=54), PwRHD = participants with right hemisphere damage (n=48), PwND = participants without neurological disorders (n=30).

MPT = maximum phonation time

Figure 2 - Receiver operating characteristic (ROC) plot for the I-AABT Comprehension (A) and Production (B) subtests for participants with aphasia in the acute phase (n=116) and without aphasia (patients with right-hemisphere damage [n = 48] and participants without neurological disorders [n = 30]).

A B



A) AUC = 0.802 B) AUC = 0.822

Table 1 - Summary of the main characteristics of the participants recruited, the procedures used and the statistical analysis performed, for each phase of the study.

Phase	Analysis	Participants	Procedure	Statistical analysis
1	**Reproducibility**			
	Test-retest	25 PwAA	I-AABT was administered twice within the same day by the same examiner	r_s ICC
	Inter-rater	21 PwAA	I-AABT was administered twice within the same day by two different examiners	r_s ICC
2	**Validity**			
	Concurrent	30 PwPA	I-AABT and AAT were administered on two separate days by the same examiner.	r_s
	Construct	70 PwAA 48 PwRHD in acute phase 30 PwND acute in-patients	I-AABT scores were compared between the three populations	Mann-Whitney test
3	**Responsiveness**			
		24 PwPA	P assessed through both the I-AABT and the AAT before and after 40 days of intensive speech and language therapy	Wilcoxon signed test
4	**Diagnostic accuracy and normative scores calculation**			
		116 PwAA 48 PwRHD in acute phase 30 PwND acute in-patients	Preliminary normative scores for I-AABT scales measuring comprehension and production performance were computed based on the overall sample of patients with aphasia in the acute phase	T-scores ROC curves

PwAA = participants with aphasia in the acute phase, PwPA = participants with aphasia in the post-acute phase, PwRHD= participants with right-hemisphere ischemic stroke , PwND= participants without neurological disorders

Table 2- Demographic and clinical characteristics of participants recruited for short-term test-retest and inter-rater reproducibility analysis (phase 1), concurrent and construct validity (phase 2) and for responsiveness (phase 3) analysis.

	Reproducibility				**Validity**				**Responsiveness**	
	test-retest		**inter-rater**		**concurrent**		**construct**			
	M(SD) min-max	**n(%)**	**M(SD) min-max**	**n(%)**	**M(SD) min-max**	**n(%)**	**M(SD) min-max**	**n(%)**	**M(SD) min-max**	**n(%)**
Age										
PwA	70.6(.5) 33-88	25(100)	74.6(9.9 75-91	21(100)	62.2 (18.2) 24-90	30(100)	73.3 (12.7) 33-95	70(47.3)	75.4(12.7) 28-91	24(100)
PwRHD							68.6(14.4) 39-87	48(32.4)		
PwND							74(8.7) 56-91	30(20.3)		
Gender										
PwA male		13(52)		6(28.6)		19(63.3)		32(45.7)		9(37.5)
PwRHD male								25(52.1)		
PwND male								18(60)		
Diagnosis										
LHS		19(76)		12(57.1)		17(56.7)		60(40.5)		16(66.7)
LHeS		5(20)		5(23.8)		5(16.7)		6(4.1)		3(12.5)
TBI		1(4)		2(9.5)				1(0.7)		
CT				2(9.5)		6(20)		4(2.7)		
RHD-LH						1(3.3)				
RHD								39(26.4)		1(4.2)
RHeS								8(5.4)		1(4.2)
NND								30(20.3)		
Others						1(3.3)				3(12.5)

PwA = participants with aphasia, PwRHD= participants with right-hemisphere stroke, PwND= participants without neurological disorders; LHS = left-hemisphere ischemic stroke, LHeS= left-hemisphere hemorrhagic stroke, TBI = traumatic brain injury, CT= cerebral tumor (after surgery), RHD-LH= right-hemisphere ischemic stroke in left-handed patient, RHD= right-hemisphere ischemic stroke, RHeS= right-hemisphere hemorrhagic stroke, NND = no neurological disorders.

Table 3 – I-AABT results for short-term test-retest (n= 25) and inter-rater (n = 21) reproducibility analysis: Spearman rank correlations and intra-class correlation coefficients with confidence interval are reported.

	Test-retest		**Inter-rater**	
	n=25		n=21	
	r_s	**ICC(CI_{95})**	**r_s**	**ICC (CI_{95})**
Global communication behavior in greeting and farewell				
Eyecontact	.841***	.837(.665-.925)	.791***	.698(.390-.865)
Shaking hand	.949***	.980(.955-.991)	.893***	.842(.651-.933)
Verbality	.986***	.985(.967-.994)	.972***	.965(.916-.986)
Facial expression	.989***	.978(.951-.990)	.958***	.961(.906-.984)
Gestures	.971***	.986(.968-.994)	1***	1(1-1)
Emotional reactions	1***	1(1-1)	1***	1(1-1)
Comprehension				
Simple requests	.910***	.965 (.924-.985)	.844***	.944 (.868-.977)
Oralpraxias	.948***	.949(.888-.977)	.967***	.972 (.924-.989)
Objects identification	.850***	.874(.735-.943)	.957***	.965 (.917-.986)
Production				
Oraldiadochokinesia	.943***	.919(.825-.964)	.928***	.950 (.881-.979)
MPT	.934***	.954(.899-.979)	.981***	.954 (.890-.981)
Automatic language and repetition	.980***	.991(.981-.996)	.975***	.987(.967-.995)
Naming	.983***	.991(.980-.996)	.969***	.987 (.970-.995)

MPT = maximum phonation time
*** $p < .001$

Table 4 – Spearman rank correlations between the I-AABT Comprehension and Production subtests and the AAT Token test, Comprehension and Naming subtests for participants (n=30) recruited for concurrent validity analysis (phase 2).

	AAT		
I-AABT	**TT**	**Comprehension**	**Naming**
	r_s	r_s	r_s
Comprehension			
Simple requests	-.903***	.871***	.825***
Oralpraxias	-.911***	.885***	.922***
Objects identification	-.948***	.944***	.907***
Production			
Oraldiadochokinesia	-.624***	.611***	.617***
MPT	-.763***	.749***	.792***
Automatic language and repetition	-.837***	.844***	.903***
Naming	-.879***	.884***	.942***

TT=Token Test

MPT= Maximum Phonation Time

*** $p < .001$

Table 5 – I-AABT Mann-Whitney U test p-scores for comparisons between the participants with aphasia in the acute phase (n=116) and either the participants without neurological disorders (n=30) or the participants with right-hemisphere damage (n=48) recruited for construct validity analysis (phase 2).

	PwAA vs PwND M-W U	**PwAA vs PwRHD** M-W U
Global communication behavior in greeting and farewell		
Eye contact	1006.5**	1228.5
Shaking hand	1242	1682
Verbality	827.5***	875.5***
Facial expression	969**	1357.5
Gestures	692***	1305.5
Emotional reactions	2115***	1973**
Spontaneous Language		
Communicative behavior	2640***	2307***
Articulation	2371***	2016.5**
Automatic language	2415***	2385.5***
Semantic	2700***	2700***
Phonemic	2445***	2402.5***
Syntactic	2655***	2592***
Comprehension		
Simple requests	2479.5***	3981.5***
Oralpraxias	2778.5***	4129.5***
Objects identification	2864***	4316***

Production		
Oraldiadochokinesia	2579.5***	3759.5***
MPT	2332.5***	3493.5*
Automatic language and repetition	2818***	4056.5***
Naming	2902***	4349***

PwAA = participants with aphasia in the acute phase, PwND = participants without neurological disorders, PwRHD = participants with right-hemisphere damage.

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 6. Descriptive statistics of I-AABT Comprehension and Production subtests for the whole sample of participants with aphasia in the acute phase (n=116).

	M	SD	Range	Skewness (S.E.)	Kurtosis (S.E.)
Comprehension					
Simple requests	37.31	16.79	0-50	-1.17 (.22)	-.16 (.45)
Oral praxias	31.42	17.87	0-50	-.70 (.22)	-1.06 (.45)
Objects identification	30.21	19.63	0-50	-.51 (.22)	-1.33 (.45)
Production					
Automatic language and repetition	31.28	19.98	0-50	-.73 (.22)	-1.18 (.45)
Naming	28.80	21.47	0-50	-.40 (.22)	-1.64 (.45)

S.E. Standard Error.

Table 7. Conversion of raw into T scores based on the overall sample of participants with aphasia in the acute phase (n=116).

Raw scores	Comprehension T scores			Production T scores	
	Simple requests	Oral praxias	Objects identification	Automatic language	Naming
0	31	34	37	38	40
1	34	37	42	43	45
2	34	38	42	43	45
3	34	39	42	44	45
4	35	40	42	44	45
5	36	41	43	44	45
6	37	41	43	44	45
7	37	41	43	44	45
8	38	41	43	44	45
9	39	42	43	44	45
10	40	42	43	44	45
11	40	42	43	44	45
12	40	42	43	44	45
13	41	42	43	44	46
14	41	43	43	44	46
15	41	43	44	44	46
16	41	43	44	44	46
17	41	44	44	44	46
18	41	44	44	44	46
19	41	44	45	44	47
20	41	45	45	44	47
21	42	45	45	44	47
22	42	45	46	44	47
23	42	45	46	45	47
24	42	45	47	45	47
25	42	45	47	45	47
26	42	45	47	45	48

27	43	46	47	45	48
28	43	46	47	45	48
29	43	46	47	46	48
30	43	47	48	46	48
31	44	47	49	46	48
32	44	47	49	47	48
33	44	47	50	47	48
34	44	47	50	47	48
35	45	48	50	48	49
36	45	49	50	48	49
37	45	49	50	48	49
38	45	49	50	49	50
39	46	50	50	49	50
40	47	51	51	50	50
41	47	51	51	50	51
42	47	52	52	51	51
43	48	53	52	51	51
44	48	54	52	52	51
45	49	55	53	53	53
46	50	57	54	54	54
47	52	59	56	56	55
48	52	59	56	57	55
49	54	61	56	58	55
50	59	65	61	63	61

T scores in the medium severity range ($45 < T \leq 55$) are marked in grey colour. T scores above this range ($T>55$) indicate mild impairment whereas scores below this range ($T\leq45$) signal severe impairment.

References

- Allibrio, G., Gori, M. A., Signorini, G., & Luzzatti, C. (2009). Un esame del linguaggio per l'analisi dei deficit afasici al letto del malato. *Giornale di Psicologia, 3* (1), 7-22.
- Bambini V., Arcara G., Aiachini B., Cattani B., Dichiarante M.L., Moro A., Cappa S.F., Pistarini C. (2016). Assessing functional communication: validation of the Italian versions of the Communication Outcome after Stroke (COAST) scales for speakers and caregivers. Aphasiology. doi: 10.1080/02687038.2016.1225275.
- Biniek, R., Huber, W., Glindemann, R., Willmes, K., & Klumm, H. (1992). The Aachen Aphasia Bedside Test-criteria for validity of psychologic tests. *Der Nervenarzt, 63*(8), 473-479.
- Biniek, R., Huber, W., Willmes, K., Glindemann, R., Brand, H., Fiedler, M., &Annen, C. (1991). A test for the detection of speech and language disorders in the acute phase after stroke. Development and clinical application. *Der Nevernarzt, 62*(2), 108-115.
- Bley, M., Wagner, A., & Berrouschot, J. (2002). Diagnosis of aphasia on the stroke unit. *Der Nervenarzt, 73*(4), 336-341.
- Bonelli, J., Centorrino, S., Ferretti, F., & Carlucci, G. (2015). A standardization study of the Italian version of Frenchay Aphasia Screening Test in aphasic patients and control subjects. *9th European CPLOL Congress Book of abstracts*.
- Brady, M. C., Kelly, H., Godwin, J., & Enderby, P. (2012). Speech and language therapy for aphasia following stroke. *The Cochrane Database of Systematic Review,16*(5). doi: 10.1002/14651858.CD000425.pub3.
- Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *The Cochrane Database of Systematic Review, 1*(6). doi: 10.1002/14651858.CD000425.pub4.

- Boulanger, J. M., Lindsay, M. P., Gubitz, G., Smith, E. E., Stotts, G., Foley, N., … Butcher, K. (2018). Canadian Stroke Best Practice Recommendations for Acute Stroke Management: Prehospital, Emergency Department, and Acute Inpatient Stroke Care; *International Journal of Stroke, 13*(9), 949-984. doi:10.1177/1747493018786616.
- Capasso R. & Miceli G. (2001). Esame Neuropsicologico per l'Afasia (E.N.P.A.). Springer Verlag, Milano.
- Chiorri, C. (2011). Teoria e tecnica psicometrica. Costruire un test psicologico [Theory and psychometric techniques. The construction of a psychological test]. Milano: McGraw-Hill.
- Coleman, E. R., Moudgal, R., Lang, K., Hyacinth, H. I., Awosika, O. O., Kissela, B. M., & Feng, W. (2017). Early Rehabilitation After Stroke: a Narrative Review. *Current Atherosclerosis reports, 19*(12), 59. doi:10.1007/s11883-017-0686-6.
- Covinsky, K.E., Palmer, R.M., Fortinsky, R.H., Counsell, S.R., Stewart, A.L., … Landefeld. C.S.(2003). Loss of independence in activities of daily living in older adults hospitalized with medical illnesses: increased vulnerability with age. J AmGeriatrSoc. 51(4):451-8. doi:10.1046/j.1532-5415.2003.51152.x
- Crary, M., Haak, N., & Malinsky, A. (1989). Preliminary psychometric evaluation of an acute aphasia screening protocol. *Aphasiology, 3* (7), 611-618. doi: 10.1080/02687038908249027.
- D’Agostino, L. (1985). *Taratura su soggetti normali di prove di denominazione per l’afasia*. Finaldissertation, Medicalschool, Neurologic Clinic, Modena University.

- ElHachioui, H., Visch-Brink, E. G., de Lau, L. M., van de Sandt-Koenderman, M. W., Nouwens, F., Koudstaal, P. J., &Dippel, D. W. (2017). Screening tests for aphasia in patients with stroke: a systematic review. *Journal of Neurology, 264*(2), 211-220. doi: 10.1007/s00415-016-8170-8.
- Enderby, P. & Crow, E. (1996). Frenchay. Aphasia Screening Test: validity and comparability. *Disability and Rehabilitation, 18*(5): 238-240.
- Flamand-Roze, C., Falissard, B., Roze, E., Maintigneeux, L., Beziz, J., Chacon, A., … Denver, C. (2011). Validation of a new language screening tool for patients with acute stroke: Language Screening Test (LAST). *Stroke, 42*, 1224-1229. doi:10.1161/STROKEAHA.110.609503.
- George D. & Mallery M. (2010). SPSS for Windows step by step: a simple guide and reference, 17.0 update, 10th edition. Pearson, Boston.
- Godecke, E., Hird, K., Lalor, E. E., Rai, T., & Phillips, M.R. (2011). *Very early post stroke aphasia therapy: a pilot randomized controlled efficacy trial*. *International Journal of Stroke, 7*(8), 635-644. doi:10.1111/j.1747-4949.2011.00631.x.
- Godecke, E., Ciccone, N. A., Granger, A. S., Rai, T., West, D., Cream, A., Cartwright, J., & Hankey, G. J. (2014). A comparison of aphasia therapy outcomes before and after a very early rehabilitation programme following stroke. *International Journal of Language and Communication Disorders, 49*(2), 149-161. doi: 10.1111/1460-6984.12074.
- Huber, W., Poeck, K., & Willmes, K.(1984). The Aachen Aphasia Test. *Advances in Neurology, 42*, 291-303.

- Iosa, C., Raimondo, S., Miletto, A. M., & Schindler A. (2011). La valutazione del paziente afasico in fase acuta – Adattamento all'italiano dell'Aachener Aphasie Bedside Test (AABT)**,** *Logopedia e comunicazione,* 7(2).
- Ivanova, M.V. & Hallowell, B. (2013). A tutorial on aphasia test development in any language: Key substantive and psychometric considerations. Aphasiology, 1;27(8):891-920. doi:10.1080/02687038.2013.805728.
- Lehman Blake, M., Frymark, T., & Venedictov, R. (2013). An evidence-based systematic review on communication treatments for individuals with right hemisphere brain damage. *American Journal of Speech Language Pathology, 22*(1), 146-160. doi: 10.1044/1058-0360(2012/12-0021).
- Luzzatti, C., Willmes, K., & De Bleser, R. (1996). *Aachener Aphasie Test*. Firenze: Giunti - OS Edizioni.
- Maas, M. B., Lev, M. H., Ay, H., Singhal, A. B., Greer, D. M., Smith, W. S., … Furie, K.L. (2012). The prognosis for aphasia in stroke. *Journal of stroke and cerebrovascular diseases: the official journal of National Stroke Association, 21*, 350-357. doi: 10.1016/ j.jstrokecerebrovasdis.2010.09.009.
- Miceli, G., Laudanna, A., Burani, C., & Capasso, R. (1994). *Batteria per l'analisi dei deficit afasici*. Roma: Centro di Promozione e Sviluppo dell'Assistenza Geriatrica.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., … de Vet, H. C. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 19*(4), 539-549. doi: 10.1007/s11136-010-9606-8.

- Muò R., Cancialosi P., Galimberti L., Carrubba Cacciola B., Gilardone M., & Schindler A. (2015). Validation of the Italian version of the American Speech-Language and Hearing Association - Functional Assessment of Communication Skills for adults (I-ASHA-FACS). Aphasiology, 29(9):1110-1130. DOI: 10.1080/02687038.2015.1010475.
- Muò R., Rinaudo M., Rabino G., Massari E., Schindler A., Steni P., &Iacomussi T. (2018). Validation of the Italian version of the Kagan scales for people with aphasia and their conversation partners. Aphasiology,DOI:10.1080/02687038.2018.1482402.
- Nakase -Thompson, R. (2004). *The Mississippi Aphasia Screening Test*. The Center for Outcome Measurement in Brain Injury. Retrived from http://www.tbims.org/combi/mast.
- National Clinical Guideline Centre (UK) London: Royal College of Physicians, (2013). Stroke Rehabilitation. Long-Term Rehabilitation after Stroke. Clinical Guideline 162.
- National Health and Medical Research Counsil (NHMRC) Centre for Clinical Research Excellence (CCRE) in Aphasia Rehabilitation (2019). *Australian Aphasia Rehabilitation Pathway*. Retrived from http://www.aphasiapathway.com.au.
- Nouwens, F., de Lau, M. L., Visch-Brink, E. G., van de Sandt-Koenderman, W. M. E, Ligsma, H. F., Goosen, S., …Dippel, S. W. J. (2017). Efficacy of early cognitive-linguistic treatment for aphasia due to stroke: a randomised controlled trial (Rotterdam Aphasia Therapy Study-3). *European Stroke Journal, 2*(2), 126–136. doi: 10.1177/2396987317698327.
- Novelli, G., Papagno, C., Capitani, E., Laiacona, M., Vallar, G., & Cappa, S. F. (1986). Tre test clinici di ricerca e produzione lessicale: Taratura su soggetti normali. *Archivio Di Psicologia, Neurologia E Psichiatria*, *4*, 477–506.
- Pizzamiglio, L., Laicardi, C., Appicciafuoco, A., Gentili, P., Judica, A., Luglio, L., . . . Razzano, C. (1984). Capacita communicative di pazienti afasici in situazioni di vita quotidiana: Addattamento italiano. Archivio Di Psicologia, Neurologia E Psichiatria, 45, 187–210.

- Posteraro L., Formis A., Bidini C., Grassi E., Curti M., Bighi M., Agosti M., & Francescini M. (2004). Aphasia quality of life: reliability of the Italian version of SAQOL-39. Europa Medicophysica, 40(4):257-62.
- Powers, W. J., Rabinstein, A. A., Ackerson, T., Adeoye, O. M., Bambakidis, N. C., Becker, K., … Tirschwell, D.L. (2018). Guideline for the early management of patients with acute ischemic Stroke: A guideline for healthcare professionals from the American Heart Association / American Stroke Association. *Stroke, 49*(3), e46-e110. doi: 10.1161/STR.0000000000000158.
- Richter, K., Wittler, M., &Hielscher-Fastabend, M. (2006). Bielefelder Aphasie Screening (BiAS) – Zur Diagnostikakuter Aphasien. *Hofheim: NAT- Verlag.*
- Rohde, A., Worrall, L., Godecke, E., O'Halloran, R., Farrell, A., & Massey, M. (2018). Diagnosis of aphasia in stroke populations: A systematic review of language tests. *PLoS One*; 13(3): e0194143. doi: 10.1371/journal.pone.0194143.
- Royal College of Physicians (2012). National clinical guideline for stroke. Retrived from http://www.rcplondon.ac.uk/.
- Salter, K., Jutai, J., Foley, N., Hellings, C., & Teasell, R. (2006). Identification of aphasia post stroke: a review of screening assessment tools. *Brain Injury, 20*(6), 559-568. doi: 10.1080/02699050600744087.
- Spreen, O. & Risser, A. H. (2003). Assessment of aphasia. New York: Oxford University Press.
- Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science*; 240: 1285-93. doi 10.1126/science.3287615
- Thommessen, B., Thoresen, G. A., Bautz-Holter, E., & Laake, K. (1999). Screening by nurses for aphasia in stroke - the Ullevaal Aphasia Screening (UAS) test. *Disability and Rehabilitation*;*21*(3):110-115.

- Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. (2005). *Regression Methods in Biostatistics. Linear, Logistic, Survival, and Repeated Measures Models*. New York: Springer-Verlag.
- Wallesch, CW. & Bak, T. (1991) Acute Aphasia. In: Hartmann A., Kuschinsky W., Hoyer S. (eds) Cerebral Ischemia and Dementia. Springer, Berlin, Heidelberg.
- Worrall L. E., Hudson K., Khan A., Ryan B., & Simmons-Mackie N. (2017). Determinants of living well with aphasia in the first year post stroke: a prospective cohort study. *Archives of Physical Medicine and Rehabilitation, 98*, 235-240. doi: http://dx.doi.org/10.1016/j.apmr.2016.06.020.

Table list.

Table 1 - Summary of the main characteristics of the participants recruited, the procedures used and the statistical analysis performed, for each phase of the study.

Table 2 -Demographic and clinical characteristics of participants recruited for short-term test-retest and inter-rater reproducibility analysis (phase 1), concurrent and construct validity (phase 2) and for responsiveness (phase 3) analysis.

Table 3 – I-AABT results for short-term test-retest (n= 25) and inter-rater (n = 21) reproducibility analysis: Spearman rank correlations and intra-class correlation coefficients with confidence interval are reported.

Table 4 – Spearman rank correlations between the I-AABT Comprehension and Production subtests and the AAT Token test, Comprehension and Naming subtests for participants (n=30) recruited for concurrent validity analysis (phase 2).

Table 5 – I-AABT Mann-Whitney U test p-scores for comparisons between the participants with aphasia in the acute phase (n=116) and either the participants without neurological disorders (n=30) or the participants with right hemisphere damage (n=48) recruited for construct validity analysis (phase 2).

Table 6 - Descriptive statistics of I-AABT Comprehension and Production subtests for the whole sample of participants with aphasia in the acute phase (n=116).

Table 7 - Conversion of raw into T scores based on the overall sample of participants with aphasia in the acute phase (n=116).

Figure 1 - Distribution of the I-AABT scores for Comprehension subtests (simple request, oral motor skills, identification of objects) and Production subtests (oral diadochokinesia, MPT, automatic language and repetition, naming objects) measures in patients with acute aphasia, post-acute aphasia, right hemisphere brain damage and participants without neurological disorders.

Figure 2 - Receiver operating characteristic (ROC) plot for the I-AABT Comprehension (A) and Production (B) subtests for participants with aphasia in the acute phase (n=116) and without aphasia (patients with right hemisphere damage [n=48] and participants without neurological disorders [n=30]).

Appendix 1 - I-AABT median and range scores for participants with aphasia in the acute phase recruited for short-term test-retest (n = 25) and inter-rater (n = 21) reproducibility analysis.

Appendix 2 - I-AABT median and range scores for participants with aphasia in the post acute phase recruited for concurrent validity analysis (n=30) and participants recruited for construct validity analysis (participants with aphasia in acute phase [n=70], participants with right-hemisphere damage [n=48] and participants without neurological disorders [n=30]). AAT median, range and T scores for participants recruited for concurrent validity are also reported.

Appendix 3 - I-AABT pre-post intensive language therapy median scores and p-scores for the Wilcoxon test for participants with aphasia in the post acute phase recruited for responsiveness analysis (n=24). For the AAT pre-post intensive language therapy mean scores, T scores and p-scores for the Wilcoxon test are reported.