A SCREENING FOR APHASIA IN NEURODEGENERATION FOR THE DIAGNOSIS OF PATIENTS WITH PRIMARY PROGRESSIVE APHASIA. CLINICAL VALIDITY AND PSYCHOMETRIC PROPERTIES

VALIDITY OF SAND IN PRIMARY PROGRESSIVE APHASIA

Petronilla Battista[1,2,3], Eleonora Catricalà[4], Marco Piccininni[1,2], Massimiliano Copetti[5], Valentina Esposito[6,7], Cristina Polito[3], Antonio Miozzo[8], Elena Gobbi[7,12], Sofia Cuoco[9], Veronica Boschi[4], Marina Picillo[9], Sandro Sorbi[3,10], Paolo Barone[9], Sandro Iannaccone[6], Peter Garrard[11], Giancarlo Logroscino[1,2], Stefano F. Cappa[4,12]

1 Department of Basic Medicine, Neuroscience, and Sense Organs, University of Bari Aldo Moro, Bari, Italy;
2 Neurodegenerative Disease Unit, Department of Clinical Research in Neurology, University of Bari Aldo Moro, "Pia Fondazione Cardinale G. Panico", Tricase, Italy;
3 NEUROFARBA- Dipartimento di Neuroscienze, Psicologia, Area del Farmaco e Salute del Bambino, Università di Firenze, Firenze, Italy;
4 Center for Neurocognition and theoretical Syntax (NEtS), Scuola Universitaria Superiore IUSS Pavia, Italy;
5 Biostatistics Unit, IRCCS-Ospedale Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy
6 Department of Specialistic Rehabilitation of Neurological, Cognitive and Motor Disorders, IRCCS San Raffaele Hospital and Vita-Salute University, Milan, Italy
7 Università Vita-Salute San Raffaele, Milano
8 Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy;
9 Dipartimento di Medicina e Chirurgia, Università di Salerno, Italy
10 IRCCS Don Gnocchi Firenze, Italy
11 Neuroscience Research Centre St George's - University of London
12 IRCCS S. Giovanni di Dio Fatebenefratelli, Brescia, Italy

Corresponding author:
Prof. Stefano F. Cappa,
Institute for Advanced Study-IUSS Pavia, Palazzo del Broletto – Piazza Vittoria 15, 27100 Pavia, Italy.
stefano.cappa@iusspavia.it
and
IRCCS S. Giovanni di Dio Fatebenefratelli, via Pilastroni 4 I-25125, Brescia, Italy
Phone +39 030 3501327

Abstract

Background: We evaluated the psychometric proprieties of the Screening for Aphasia in NeuroDegeneration (SAND) battery in Italian Primary Progressive Aphasia (PPA) and movement disorder (MD) patients. **Methods**: The sample included 30 consecutive PPA and 45 MD patients who completed the SAND battery together with a clinical interview and a neurological/neuropsychological examination and 134 healthy subjects (HC). **Results**: The SAND battery showed good internal consistency and good convergent and divergent validity. ROC analysis revealed an AUC of 0.978 for PPA vs HC and 0.786 for PPA vs MD. A cut-off of ≥3 gave a Sensitivity 0.933 and a Specificity 0.946 for discriminating PPA vs HC; whereas a cut off ≥5 gave a Sensitivity of 0.767% and Specificity 0.667 for discriminating PPA vs MD. **Conclusions**: These results indicate that the SAND battery is an adequate, reliable and valid diagnostic tool for PPA.

Keywords: language impairment, screening instrument, dementia, diagnostic accuracy, speech-language assessment.

INTRODUCTION

The language network can be selectively affected by neurodegeneration, leading to progressive language dysfunction (primary progressive aphasia, PPA). After the seminal description of slowly progressive aphasia by Mesulam [1], many studies have been dedicated to this heterogeneous group of clinical conditions, due to different pathologies, in which language impairment is the main sign at onset, and remains the prominent clinical feature during many years of progression. The consensus paper on the definition of PPA and its clinical subtypes recognizes three different variants: non-fluent/agrammatic (nf/a-PPA), semantic variant PPA (sv-PPA) and logopenic PPA (lp-PPA) [2]. Each of these variants presents with a different pattern of speech and language deficits and brain imaging features [3]. The diagnosis of PPA is predominantly based on the clinical examination, including detailed history from patient and informant, and neurological and neuropsychological evaluation. Most language tests in common use, such as the Aachener Aphasie Test (AAT) [4] and the Boston Diagnostic Aphasia Examination [5], have been developed for the evaluation of aphasia due to stroke, which, besides having a different prognosis, is characterized by language profiles that can be very different from PPA [6]. Given the dearth of specific tools aimed at assessing language impairments in neurodegenerative diseases [7], we developed a Screening Battery for Aphasia in NeuroDegeneration (SAND) [8], based on the Mini Standard Language Examination proposed by Garrard and Ahmed [9]. The SAND battery aims at the detection of language disorders through the assessment of different components of language. It includes nine subtests: picture naming, word and sentence comprehension, word and sentence repetition, reading, semantic association, writing and picture description. It has been developed on the basis of the recommendations of current diagnostic guidelines [2] and of a comprehensive review of the language deficits in PPA subtypes. The claim that the SAND battery is adequate to detect language disorders requires the assessment of its construct, clinical validity and of its diagnostic accuracy. To this aim, we studied a group of participants with PPA, a group of healthy controls (HC) and a group of patients with movement disorders-MD (Parkinson's disease-PD and Progressive Supranuclear Palsy-PSP). We expected the PPA patients to be the most impaired group, as aphasia is the major symptom in these patients. Healthy individuals were expected to perform close to ceiling.

In the case of MD, PD patients were expected to show normal or mildly affected language performance [10]. A similar prediction could be made for PSP patients, since those presenting with a clinical picture of nf/a-PPA were excluded. The psychometric properties of the SAND battery and the discriminatory power in detecting PPA language dysfunction were assessed by comparing PPA with HC and with MD, estimating the Area Under the Curve (AUC), the Sensitivity and the Specificity.

MATERIAL AND METHODS

Participants

Two patient samples were included in the study:

Patients with a clinical diagnosis of PPA according to published guidelines [2], based on audiotapes of language and cognitive testing, history, and review of imaging (magnetic resonance imaging or positron emission tomography). Those patients not fulfilling the diagnostic criteria for a specific variant were assigned the label “unclassifiable”. These patients did not present the typical features of any variant, but only anomia or were characterized by features of more than one variant (e.g., non-fluent and logopenic).

Patients with a clinical diagnosis of Parkinson’s disease (PD) or Progressive Supranuclear Palsy (PSP). Patients with PD [11], and with probable or possible PSP [12] were clinically diagnosed. Motor impairment severity was assessed using Unified Parkinson's disease rating scale U-PDRS [13] for PD patients and Natural History and Neuroprotection in Parkinson Plus Syndromes-NIPPS-PPS [14] for PSP patients. All patients were recruited in a movement disorder unit. Features of language impairment could be present but were not the primary cause of referral for any of these patients. The SAND data were collected as part of a comprehensive study of cognition in MD.

Enrollment started in May 2015 and ended in February 2017 Participants were selected among the outpatients in hospital located in different Italian areas (Milan, Florence, Bari and Salerno).

Specifically, criteria for inclusion were: (i) availability of a audiotaped language examinations in order to allow an offline analysis of connected speech; (ii) Mini-Mental State Examination-MMSE [15] of at least 10; (iii) Italian native speaker; (iv) sufficiently intelligible speech, such that the intended target could be determined for the majority of words; (v) intact or corrected auditory and visual functions; (vi) successful completion of the experimental task.

Patients were excluded in the case of: (1) major psychiatric disorders [16]; (2) organic illness affecting the brain according to the International Classification Disorder [17]; (3) significant history of head injury; (4) major systemic diseases or medical complications, including thyroid disorders and sensory disorders (i.e., blindness or deafness); and (5) history of drug or alcohol addiction.

Informed written consent was obtained from all participants. The local Hospital ethical committee approved the study protocol. Information collected during the course of the research was kept confidential, the subjects’

names and details were removed to prevent identification of participants and data storage was password protected with data access restricted to study personnel.

All study participants underwent extensive studies at baseline, following a diagnostic protocol that included medical, neurological, neuropsychological and neuroimaging investigations (with magnetic resonance imaging or positron emission tomography). Further examinations, such as cerebrospinal fluid markers (β-amyloid and tau protein concentration) were obtained for a small group of patients in order to improve the accuracy of the diagnosis. Neurologists and neuropsychologists with experience in cognitive disorders and/or movement disorders examined all patients.

A group of HC was enrolled from a convenience sample of volunteers recruited in the centers included in the study. The group of HC was the same recruited for the collection of normative data. Subjects who reported a history of neurological or psychiatric illnesses or with a corrected score of less than 24 at MMSE were excluded [15].

A total of 92 patients were recruited (43 patients with PPA, 49 patients with MD). We excluded thirteen patients with PPA because they were too severely affected to perform a full neuropsychological assessment, 3 patients with MD (2 PSP and 1 PD) who fulfilled the criteria for dementia and 1 PSP patient with nf/a-PPA. Therefore, the global sample (75 patients, 43 males and 32 females, age range 42–85 years) thus included three different diagnostic groups: 30 PPA patients, including 8 sv-PPA, 12 lp-PPA, 6 nf/a-PPA, 4 unclassifiable PPA, and 45 MD patients, 24 with PD and 21 with PSP. The HC group included 130 native Italian speakers (54 males), aged over 45 years – both sexes, with an educational level of >1 year, with a mean age of 63.30 (SD =11.30, range 45–85 years) and a mean of 10.5 years of education (SD =4.89; range 2–25).
Data on demographic and clinical features of all patients and HC included in the study are reported in Table 1.

[Insert here Table 1]

SAND procedure

All participants were given the SAND battery, whose normative values have already been published [8]. The entire battery takes less than 20 minutes to administer and yields scores for each of the nine subtests. Picture description and written description analysis yields additional subscores, resulting in a total of twenty-three task-related scores (see Supplementary Materials Figure 1S).

Cognitive assessment

In addition to the SAND battery, PPA patients were administered an extensive battery of standardised neuropsychological tests in a 2-hr session. The tests covered the following cognitive domains: attention, executive functions, memory, language, and visuospatial processing (see Supplementary Materials). For each test score, we report the corresponding adjusted score obtained on the basis of the normative data (i.e., according to age and education corrections).

STATISTICAL ANALYSIS

Demographics and clinical variables

Quantitative variables are reported as mean, standard deviation, median and range, while qualitative variables are reported as frequencies and percentages. Comparisons among the diagnostic groups for demographic and clinical variables was made using Kruskal Wallis test by ranks and Nemenyi post-hoc tests for quantitative variables, or Pearson's chi-squared test for qualitative variables. The disease duration of the two patient groups (PPA and MD) was compared using the Mann-Whitney U Test.

Global diagnostic score

Before proceeding with statistical analysis, a Global Score of the SAND battery was calculated. The Global Score includes the twenty-three task-related-scores. Three steps were followed:

1) the raw scores were adjusted by adding or subtracting the influence of age, sex and education and corrected using normative data;
2) corrected scores were compared with the corresponding cut-off values obtained from HC;
3) the sum of the twenty-three dichotomous variables (1=pathological 0=normal) represents the global score, with higher scores indicating a more severe impairment (range 0-23).

Reliability and Construct validity

The reliability of the Internal Consistency of the SAND battery was estimated using Cronbach's alpha coefficient [18]. Construct validity was assessed considering only PPA patients and using corrected scores of the task-related-scores. In order to assess construct validity, a correlation analysis between corrected task-related-scores and the other language/non-language tests was carried out. Tables 2 and 3 display the correlations results. All the correlations were estimated using the Spearman's rank correlation coefficient, in order to deal with the non-normal distributions of the scores.

ROC analyses

The ROC curve for SAND Global Score was used to detect the optimal cut-off score maximizing both sensitivity and specificity. The AUC, sensitivity and specificity were calculated along with their 95% bootstrapped confidence interval (CI), considering 2000 sampling replications. Positive predictive value (PPV) and Negative predictive value (NPV) were also reported for the optimal cut-off. The ROC analysis, to assess the discriminatory power of the SAND battery (global score), was used for both comparisons: PPA vs HC and PPA vs MD. All analyses were performed using R (v 3.3.1) and Rstudio (v 1.0.153). A p value lower than .05 was considered as statistically significant.

RESULTS

Demographics

No significant differences were detected between groups in demographic data, with the exception of age, where HC were younger than the PPA group ($p < 0.001$); however, this variable does not influence the results,

since all performances have been adjusted for age, sex and education effects. MD patients had a significant longer disease duration than PPA patients (p=0.007).

Global diagnostic score

Global scores of the SAND battery for HC vs PPA vs MD are displayed in Supplementary Materials Figure 2S. As expected, PPA patients performed worse than HC (p <0.001) and MD (p=0.011). HC performances were close to ceiling. A significantly worse performance was observed in MD patients compared to HC (p<0.001).

Reliability and Construct Validity

The Cronbach's alpha coefficient of the global score of the SAND battery was 0.864 as computed from the whole sample.Table 2 shows the correlation analysis among the sub-scores of the SAND battery and the language tests. All SAND sub-scores were correlated with the conventional language test (all p<0.005). The strongest correlations were found for sentence repetition with the ENPA sentence repetition task (r=0.85) and for the naming subtest with the CAGI naming task (r=0.82). Table 3 shows the correlations between the global score of the SAND battery and non-language tests, which were not significant, with the exception of tasks with a relevant linguistic component (MMSE, digit span).

[Insert here Table 2 and Table 3]

ROC analyses

The ROC curve showed that the optimal cut-off score for the comparison of the PPA sample vs HC was 3, achieving a high sensitivity of 0.933 (95% CI 0.833 to 1.000), a high specificity of 0.946 (95% CI 0.908 to 0.985), a positive predictive value (PPV) of 0.800 and a negative predictive value (NPV) of 0.984 (Figure 1a). The overall discriminatory power (AUC) for the SAND battery was 0.978 (95% CI 0.945 to 0.999). ROC analyses are shown in Figure 1b. A score of 2 classified PPA vs HC with 0.933 sensitivity (95% CI 0.833 to 1.000) and 0.854 specificity (95% CI 0.792 to 0.908). A score of 4 had a sensitivity of 0.833 (95% CI 0.700 to 0.967) and a specificity of 1.000 (95% CI 1.000 to 1.000).

For the comparison between PPA and MD (PD + PSP) patients, the ROC curve showed that the optimal cut-off was 5, achieving a high sensitivity of 0.767 (95% CI 0.600 to 0.900), and a specificity of 0.667 (95% CI 0.533 to 0.800), a PPV of 0.605 and a NPV of 0.811 (Figure 1c). The overall discriminatory power (AUC) was 0.786 (95% CI 0.670 to 0.882). ROC analyses are shown in Figure 1d. A score of 6 accurately classified PPA vs MD with 0.667 sensitivity (95% CI 0.500 to 0.883) and 0.689 specificity (95% CI 0.555 to 0.822). A score of 4 had a sensitivity of 0.833 (95% CI 0.667 to 0.967) and specificity of 0.578 (95% CI 0.444 to 0.711).

[insert here Figure 1a-d]

DISCUSSION

Language assessment has a critical role in the clinical diagnosis of neurodegenerative diseases, in particular in the case of PPA. However, only a few attempts have been made to develop specific tools to diagnose, clinically classify and follow up the heterogeneous group of PPA patients. Recently, a systematic review

identified the few neuropsychological tests available for the assessment of speech and language disorders in PPA and discussed their limitations [7]. This is the main reason leading to the development of the SAND battery [8]. In this study, we explored its clinical validity and psychometric characteristics. Our findings suggest that the SAND battery provides useful information in the clinical diagnosis of PPA patients.

Consistency and validity. The good internal consistency, with a Cronbach coefficient of .864 for the global score, means that the SAND scores are consistent with each other for the content that they measure. The data about construct validity showed that the SAND is a valid measure of language functions. Specifically, the strength of the correlation among the nine scores of the battery and the other measures assessing comparable components of language performance (i.e., naming, comprehension etc.) was high, compatible with a shared general language function dimension. Divergent validity was also assessed, analyzing the correlation between SAND global score and non-language tasks. Two measures were significantly associated with the global score: 1) the MMSE, which includes language subtests (e.g., naming, writing and repetition of sentence); 2) the Digit Span backwards, which can be expected to correlate with the two tasks with a relevant working memory load e.g., repetition of single word/non-word and of sentences.

Clinical usefulness. As a screening battery, the main goal of the SAND is to determine whether an individual has language impairment. Our results clearly indicate that an impaired performance on the test can be used to identify individuals with clinically significant language impairments. The diagnostic accuracy for specific conditions needs to be confirmed using a reliable gold standard, i.e. neuropathological confirmation, positivity of reliable biomarkers [19], or presence of pathogenetic mutations [20], associated with FTD or with Alzheimer's Disease (AD) [21]. A diagnosis by independent experts using the current criteria [2] is at the moment the only available option in the case of PPA patients.

Speech and language disorders may occur in corticobasal degeneration and PSP. Patients with these diagnoses can actually fulfill the criteria for nf/a-PPA and Apraxia of Speech [22], [23], and were excluded from the present MD sample. The SAND battery is thus also able to discriminate non-aphasic MD patients from controls, providing a good evidence for its sensitivity in detecting subtle language deficits in neurodegenerative diseases in general.

The SAND battery represents a first step towards a concise multilingual standard language examination, as a fast and simple tool to help clinicians and researchers in the clinical diagnosis of PPA. A further development of this battery will lead to identify the combinations of tests and test items that most reliably and accurately classify patients based on neuropsychological/linguistic features. The use of machine learning algorithms, as reported by other several studies, may represent a useful approach to this aim (e.g., [24- 26]), when a sufficiently large dataset allowing the identification of which test/test item leads to the best discrimination of deficits in each variant will become available.

Limitations. A limitation is that the global score may be biased by the use of cut-off values obtained from the same control group used in the validation process. The group of MD patients was included in order to compare the performance of a different clinical population. Another limitation is the small sample size. Therefore, our findings require replication in larger populations to properly assess generalization. A minor point concerns PSP

patients, which were included in the MD group even if they may clinically present as the non-fluent/agrammatic variant PPA [27]. The present sample, however, was recruited from MD clinics, and we excluded cases fulfilling the criteria for PPA.

CONCLUSION

Taken together, our findings encourage the use in clinical practice of the SAND battery, which has been shown to be able to identify patients with PPA. Overall, the SAND showed good validity as a screening instrument to detect language impairment in patients with neurodegenerative disorders. The satisfactory PPVs and NPVs and the limited length of this cognitive and language measure makes this battery especially useful in clinical contexts [28]. Cognitive and language tools are inexpensive and low-tech additions to the diagnostic process that can be used in a variety of clinical settings: primary, secondary or tertiary care and research studies. The SAND battery may be especially useful in settings where a complete neuropsychological/language assessment cannot be proposed because of lack of time or insufficient neuropsychological resources.

ACKNOWLEDGEMENTS

We acknowledge Alessandra Marcone, Chiara Cerami, Rosanna Tortelli, Rosa Capozzo for their help in patient recruiting. We wish to thank Monica Consonni and Cristiano Chesi for their helpful comments, Caterina Balduzzi for her help in scoring data and Laura Sampietro for assistance in data management.

FUNDING

This study was supported by the MRC Research Grant: Ref MR/N025881/1, and by the “AIRAlzh” Onlus and ANCC-COOP Italia to P. Battista.

COMPETING INTERESTS None.

BIBLIOGRAPHY

1. Mesulam MM, Spectrum of primary progressive aphasia. Baillieres Clin Neurol. 1992.
2. Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, Ogar JM, Rohrer JD, Black S, Boeve BF, Manes F, Dronkers NF, Vandenberghe R, Rascovsky K, Patterson K, Miller BL, Knopman DS, Hodges JR, Mesulam MM, Grossman M. Classification of primary progressive aphasia and its variants. Neurology 2011; 76: 1006–1014.
3. Cerami C, & Cappa SF. Primary progressive aphasia. in Hodges’ Frontotemporal Dementia 2016.
4. Luzzatti C, Willmes K, De Bleser R., Bianchi A, Chiesa G, De Tanti A, Gonella ML, Lorenzi L, Pozzoli C. New normative data for the Italian version of the Aachen Aphasia Test (A.A.T.). [Nuovi dati normativi per la versione italiana dell’Aachener Aphasie Test (A.A.T.)]. Arch. di Psicol. Neurol. e Psichiatr. 1994; 55: 1086–1131.
5. Goodglass H. Boston diagnostic aphasia examination: Short form record booklet. Lippincott Williams & Wilkins 2000.
6. Mesulam MM & Weintraub S. Is it time to revisit the classification guidelines for primary progressive

aphasia? Neurology 2014; 82: 1108–1109.
7. Battista P, Miozzo A, Piccininni M, Catricalà E, Capozzo R, Tortelli R, Padovani A, Cappa SF, & Logroscino G. Primary progressive aphasia: a review of neuropsychological tests for the assessment of speech and language disorders. Aphasiology 2017; 31: 1359–1378.
8. Catricalà E, Gobbi E, Battista P, Miozzo A, Polito C, Boschi V, Esposito V, Cuoco V, Barone P, Sorbi S, Cappa SF, & Garrard P. SAND: a Screening for Aphasia in NeuroDegeneration. Development and normative data. Neurol. Sci. 2017; 38: 1469–1483.
9. Garrard P, & Ahmed S. An abbreviated examination for the assessment of linguistic impairment in primary progressive aphasia. Eur. J. Neurol. 2012;19:561.
10. Boschi V, Catricala E, Consonni M, Chesi C, Moro A, & Cappa SF. Connected speech in neurodegenerative language disorders: a review. Frontiers in psychology, 2017; *8*, 269.
11. Gelb DJ, Oliver E, & Gilman S. Diagnostic criteria for Parkinson disease. Arch. Neurol. 1999; 56: 33–9.
12. Litvan I, Agid Y, Calne D, Campbell G, Dubois B, Duvoisin RC, Goetz CG, Golbe LI, Grafman J, Growdon JH, Hallett M, Jankovic J, Quinn NP, Tolosa E, Zee DS. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): Report of the NINDS-SPSP International Workshop. Neurology 1996; 47: 1–9.
13. Fahn S, & Elton R. Unified Parkinson's Disease Rating Scale. in Recent developments in Parkinson's disease 1987; 153–63.
14. Payan CA, Viallet F, Landwehrmeyer BG, Bonnet AM, Borg M, Durif F, Lacomblez L, Bloch F, Verny M, Fermanian J, Agid Y, Ludolph AC, Leigh PN, Bensimon G; NNIPPS Study Group . Disease severity and progression in progressive supranuclear palsy and multiple system atrophy: Validation of the NNIPPS - PARKINSON PLUS SCALE. PLoS One 2011; 6.
15. Measso G, Cavarzeran F, Zappala G, Lebowitz BD, Crook TH, Pirozzolo FJ, Amaducci LA, Massari D, Grigoletto F. The mini-mental state examination: Normative study of an Italian random sample. Dev. Neuropsychol. 1993; 9: 77–85.
16. Association American Psychology. Diagnostic and Statistical Manual of Mental Disorders, (DSM IV). Washingt. DC, APA Fourth Ed. 1994; 915.
17. Organization WH. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. WorldHealthOrganization 1992; 1–267
18. Tavakol M, & Dennick R. Making sense of Cronbach's alpha. International journal of medical education, 2011; 2: 53–55.
19. Grossman M, Biomarkers in the primary progressive aphasias. Aphasiology 2014; 28: 922–940.
20. Rohrer JD. The genetics of primary progressive aphasia. Aphasiology 2014; 7038: 1–7.
21. Chare L, Hodges JR, Leyton CE, McGinley C, Tan RH, Kril JJ, & Halliday GM. New criteria for frontotemporal dementia syndromes: Clinical and pathological diagnostic implications. J. Neurol. Neurosurg. Psychiatry 2014; 85: 866–871.
22. Kertesz A, McMonagle P, Blair M, Davidson W, & Munoz DG. The evolution and pathology of frontotemporal dementia. Brain 2005; 128:1996–2005.
23. Hodges JR, Davies RR, Xuereb JH, Casey B, Broe M, Bak TH, Kril JJ, Halliday GM. Clinicopathological correlates in frontotemporal dementia. Ann. Neurol. 2004; 56: 399–406.
24. Battista P, Salvatore C, Castiglioni I. Optimizing neuropsychological assessments for cognitive,

behavioral, and functional impairment classification: A machine learning study. Behav. Neurol. 2017.
25. Fraser KC, Meltzer JA & Rudzicz F. Linguistic Features Identify Alzheimer ' s Disease in Narrative Speech. Journal of Alzheimer's Disease 2016; 49(2): 407-422.
26. Tunvirachaisakul C, Supasitthumrong T, Tangwongchai S, Hemrunroj S, Chuchuen P, Tawankanjanachot I, Likitchareon Y, Phanthumchinda K, Sriswasdi S, Maes M. Characteristics of Mild Cognitive Impairment Using the Thai Version of the Consortium to Establish a Registry for Alzheimer's Disease Tests: A Multivariate and Machine Learning Study. Dementia and geriatric cognitive disorders 2018; 45(1): 38-48.
27. Merello M, & Starkstein SE. Movement disorders in dementias. Movement Disorders in Dementias 2014.
28. Prince M, Bryce R, & Ferri C. World Alzheimer Report - The benefits of early diagnosis and intervention World Alzheimer Report. Alzeheimer's Disease International (ADI) 2011.

Figure captions

Fig. 1a) Summary of the diagnostic accuracy of the SAND battery for the comparison PPA vs HC. **1b)** ROC curve for the global score of the SAND battery to detect patients with language dysfunction evaluated in the sample of PPA vs HC (SEN, Sensitivity, SPE, Specificity, AUC, Area Under the receiver operating characteristic Curve; ROC, receiver-operating characteristic). **1c)** Summary of the diagnostic accuracy of the SAND battery for the comparison MD vs. PPA. **1d)** ROC curve for the global score of SAND battery to detect patients with language dysfunction evaluated in the comparison PPA vs MD.

Table 1 Demographic characteristics and clinical data of enrolled subjects. HC=Healthy Controls; PPA=Primary Progressive Aphasia; MD= movement disorders; PD=Parkinson's Disease; PSP=Progressive Supranuclear Palsy. The p-value for the overall differences test (p) and p-values for the post-hoc analysis (pairwise comparisons: PPA vs HC; MD vs HC and PPA vs MD) are reported.

Variables	HC	PPA	MD (PD + PSP)	P	PPA vs HC	MD vs HC	PPA vs MD
Sample size	130	30	45 (24+21)	-	-	-	-
Age mean (± st.dev.)	63.30 (±11.30)	70.9 (±6.04)	66.98 (± 8.05)	<0.001	<0.001	0.127	0.196
Education mean (± st.dev.)	10.92 (±4.90)	11.67 (±4.94)	11.11 (±4.90)	0.700	-	-	-
Sex (%m)	41.54%	53.33%	60.00%	0.078	-	-	-
Disease duration in months mean (± st.dev.)	-	31.63 (±17.62)	54.36 (±37.97)	0.007	-	-	-
MMSE mean (± st.dev.)	28.49 (±1.52)	20.48 (±6.50)	25.04 (±4.81)	<0.001	<0.001	<0.001	0.022
SAND global score mean (± st.dev.)	0.65 (±0.86)	7.87 (±3.87)	3.82 (±3.50)	<0.001	<0.001	<0.001	0.011
UPDRS for PD	-	-	15.80 (±7.97)	-			

NNIPPS-PPS for PSP	87.15 (±38.16)

Table 2 Correlation analysis among the sub-scores of the SAND battery and the state-of art language tests. Note. AAT= Aachener Aphasie Test, ENPA= Esame Neuropsicologico dell'Afasia, PPT= Pyramids and Palm Tree Test. Negative values refer to those tests for which the higher the score the worse the performance. *|0.30| < r < |0.50| moderate correlation; **r > |0.50| strong correlation. The following correlations were performed: Naming with CAGI naming; Auditory Sentence Comprehension with ENPA Auditory Sentence Comprehension; Word Comprehension with CAGI word-picture matching; Word/non-word repetition with ENPA Word/non-word repetition; Sentence Repetition with AAT Sentence Repetition, AAT Total score repetition and ENPA Sentence repetition; Reading with TIB total errors score, Writing – syntactic structure/Sentences with ENPA Sentence writing; Semantic Association with PPT total score; Picture description – nouns/words with Category Fluency and CAGI naming.

Tasks SAND	Language Tests	Spearman	P-value	Sample size
Naming	CAGI naming	0.82	**<0.001**	28
Sentence Comprehension	Token Test	0.63	**<0.001**	24
	Auditory sentence comprehension (ENPA)	0.45	**0.007**	17
Word Comprehension	CAGI word-picture matching	0.67	**<0.001**	25
Word/non-word repetition	Word/ non-word repetition (ENPA)	0.54	**0.009**	22
Sentence Repetition	Sentence repetition (AAT)	0.74	**0.002**	14
	Total score repetition (AAT)	0.84	**<0.001**	16
	Sentence repetition (ENPA)	0.85	**<0.001**	22
Reading	TIB	-0.59	**0.012**	17
Writing – syntactic structure/Sentences	Sentence writing (ENPA)	0.49	**0.020**	22
Semantic Association	PPT	0.50	**0.040**	17
Picture description – nouns/words	Category Fluency	0.55	**0.003**	26
	CAGI naming	0.62	**<0.001**	28

Table 3 Correlation analysis among the sub-scores of SAND battery and the non-language tests. Note. RAVLT = Rey Auditory Verbal Learning Test; RCF = Rey Complex Figure; SCWT = Stroop Colour-Word Test, TMT-AB = Trail Making Test part A and B; FAB = Frontal Assessment Battery; CDT = Clock Drawing Test. Negative values refer to those tests for which the higher the score the worse the performance. *|.30| < r < |.50| moderate correlation; **r > |.50| strong correlation. The Global Score of the SAND battery was correlated with the following tests: MMSE, RAVLT (immediate, delayed recall, recognition and false alarms), Digit Span (backward and forward), Corsi Span backward, RCF (copy and delayed recall), SCWT time and errors), TMT-AB, FAB and CDT.

	Non Language Tests	Spearman	P-value	Sample size
Global Score SAND battery	MMSE	-0.56	**0.001**	29
	RAVLT immediate	-0.44	0.059	19
	RAVLT recall	-0.24	0.319	19
	RAVL recognition	-0.51	0.073	13
	RAVLT false allarms	0.34	0.304	11
	Digit Span backward	-0.40	**0.034**	28
	Digit Span forward	-0.29	0.144	27
	Corsi Span backward	-0.19	0.341	27
	RCF Copy	-0.12	0.578	25
	RCF Recall	-0.36	0.084	24
	SCWT time	0.30	0.323	13
	SCWT errors	0.34	0.251	13
	TMT-A	0.34	0.108	23
	TMT-B	0.49	0.060	15
	TMT-AB	0.47	0.076	15
	FAB	-0.33	0.132	22
	CDT	-0.13	0.585	19