



HAL
open science

Improving language evaluation in neurological disorders: The French Core Assessment of Language Processing (CALAP).

Charlotte Jacquemot, Christophe Lalanne, Agnès Sliwinski, Page Piccinini,
Emmanuel Dupoux, Anne-Catherine Bachoud-Lévi

► To cite this version:

Charlotte Jacquemot, Christophe Lalanne, Agnès Sliwinski, Page Piccinini, Emmanuel Dupoux, et al.. Improving language evaluation in neurological disorders: The French Core Assessment of Language Processing (CALAP).. *Psychological Assessment*, 2019, 31 (5), pp.622-630. 10.1037/pas0000683 . hal-02326543

HAL Id: hal-02326543

<https://hal.science/hal-02326543>

Submitted on 19 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Improving Language Evaluation in Neurological Disorders: The French Core Assessment of Language Processing (CALAP)

- AQ: 1 **Charlotte Jacquemot**
PSL Research University; INSERM U955-E01, IMRB,
Neuropsychologie Interventionnelle; and Université Paris-Est
- AQ: 2 **Christophe Lalanne**
Université Paris and Hôpital Hôtel-Dieu, Paris, France
- AQ:3, 4 **Agnès Sliwinski and Page Piccinini**
PSL Research University; INSERM U955-E01, IMRB,
Neuropsychologie Interventionnelle; and Université Paris-Est
- Emmanuel Dupoux**
PSL Research University and Laboratoire de Sciences
Cognitives et Psycholinguistique, EHESS, CNRS, Paris, France
- Anne-Catherine Bachoud-Lévi**
PSL Research University; INSERM U955-E01, IMRB, Neuropsychologie Interventionnelle; Université Paris-Est; and Hôpital Henri Mondor-Albert Chenevier, Créteil, France

Aphasia is a devastating brain disorder, detrimental for medical care and social interaction. The early diagnosis of language disorders and accurate identification of patient-specific deficits are crucial for patients' care, as aphasia rehabilitation is more effective when focused on patient-specific language deficits. We developed the Core Assessment of Language Processing (CALAP), a new scale combining screening and detailed evaluation to rapidly diagnose and identify patient-specific language deficits. This scale is based on a model of language processing distinguishing between the comprehension, production, and repetition modalities, and their different components: phonology (set of speech-sounds), morphology (how the sounds combine to form words), lexicon (words), syntax (how words combine to form sentences), and concept (semantic knowledge). This scale was validated by 189 participants who underwent the CALAP, and patients not unequivocally classified as without aphasia by a speech-language pathologist underwent the Boston Diagnosis Aphasia Evaluation as the gold standard. CALAP-screening classified patients with and without aphasia with a sensitivity of 1 and a specificity of 0.72, in 3.14 ± 1.23 min. CALAP-detailed evaluation specifically assessed the language components in 8.25 ± 5.1 min. Psychometric properties including concurrent validity, internal validity, internal consistency and interrater reliability showed that the CALAP is a valid and reliable scale. The CALAP provides an aphasia diagnosis along with the identification of patient-specific impairment making it possible to improve clinical follow up and deficit-based rehabilitation. It is a short and easy-to-use scale that can be scored and interpreted by clinicians nonexpert in language, in patients with fatigue and concentration deficits.

Public Significance Statement

This validation study showed that the Core Assessment of Language Processing (CALAP), a new scale for assessing language disorders based on a functional language model and combining screening and detailed evaluation, is a valid and reliable tool for language assessment in patients with neurological disorders. CALAP-screening can be used to diagnose aphasia and CALAP-detailed evaluation efficiently identifies the patient-specific language deficits, making it possible to improve both clinical follow up and deficit-based rehabilitation.

Charlotte Jacquemot, Département d'Etudes Cognitives, Ecole normale supérieure, PSL Research University; INSERM U955-E01, IMRB, Neuropsychologie Interventionnelle; and Université Paris-Est. Christophe Lalanne, Université Paris and Hôpital Hôtel-Dieu, Unité de recherche clinique en économie de la santé, Paris, France; Agnès Sliwinski and Page Piccinini, Département d'Etudes Cognitives, Ecole normale supérieure, PSL Research University; INSERM U955-E01, IMRB, Neuropsychologie Interventionnelle; and Université Paris-Est; Emmanuel Dupoux, Département d'Etudes Cognitives, Ecole normale supérieure, PSL Research Uni-

versity and Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS, CNRS, Paris, France; Anne-Catherine Bachoud-Lévi, Département d'Etudes Cognitives, Ecole normale supérieure, PSL Research University; INSERM U955-E01, IMRB, Neuropsychologie Interventionnelle; Université Paris-Est; and AP-HP, Service de neurologie, Hôpital Henri Mondor-Albert Chenevier, Créteil, France.

Correspondence concerning this article should be addressed to Charlotte Jacquemot, Neuropsychologie Interventionnelle, Département d'Etudes Cognitives, Ecole normale supérieure, 29, rue d'Ulm, 75005 Paris, France. E-mail: charlotte.jacquemot@ens.fr

AQ: 5 *Keywords:* aphasia, language disorders, scale, diagnosis, functional model of language, cognitive rehabilitation, validation

AQ: 6 *Supplemental materials:* <http://dx.doi.org/10.1037/pas0000683.supp>

Aphasia is one of the most devastating disorders in patients with brain impairments, regardless of etiology (stroke, traumatic brain injury, or neurodegenerative disease; Denier et al., 2015; Kirshner, 2014; Mesulam, 2001). Aphasia limits social interaction (Laska, Mårtensson, Kahan, von Arbin, & Murray, 2007), which hinders cognitive rehabilitation efforts (Duncan et al., 2005; Kauhanen et al., 2000), and delays the return to work, with a major impact on care pathways and health care costs (Ellis, Simpson, Bonilha, Mauldin, & Simpson, 2012).

Following the diagnosis of aphasia, rehabilitation is more effective when intensive and focused on the impaired component than when unspecific (Breitenstein et al., 2017; Hillis, 1998; Jacquemot, Dupoux, Robotham, & Bachoud-Lévi, 2012). In order to propose intensive deficit-based rehabilitation, identification of the patient-specific language deficits is critical. However, there is currently no validated scale for identifying the specific components impaired within a functional model of language processing. On the one hand, long scales, such as the Boston Diagnosis Aphasia Evaluation (BDAE; Goodglass & Kaplan, 1972)—considered as the gold standard for comprehensive language assessment—are used for assessing oral comprehension, production, repetition, and writing and reading abilities. They distinguish between the various types of aphasia (e.g., Broca, Wernicke, etc.). The BDAE requires several sessions and about two or three hours with a trained speech-language pathologist for its administration and rating, precluding its use in patients with fatigue. On the other hand, a very short scale such as the Language Screening Test (LAST) briefly assesses oral comprehension, production, and repetition in 2 min (Flamand-Roze et al., 2011). The LAST is used to detect language disorders in acute stroke and to orient toward thrombolysis or not (Denier et al., 2016). However, none of these scales assesses language components within a functional model of language processing, precluding their use to guide aphasia rehabilitation (Caramazza & Hillis, 1993).

To fill this gap, we developed and validated the French Core Assessment of Language Processing (CALAP), a short easy-to-use scale for the assessment of oral language. The CALAP is based on a functional model of language composed of several components reported in Figure 1A (Caramazza, 1997; Jacquemot, Dupoux, & Bachoud-Lévi, 2007; Jacquemot et al., 2012; Levelt, 1999; Miceli, Capasso, & Caramazza, 1994) and focused on the evaluation of oral language. The model, purposely simplified, makes the distinction between the comprehension, production, and repetition modalities. Each modality is composed of several components: speech sounds are processed through the phonological component, the combination of sounds to form the words is performed through the morphological component, the recognition or selection of words involves the lexical component, and the words are combined to form sentences through the syntactic component. Finally, semantic knowledge (which includes common knowledge of facts, events, concepts of the world, and personal experience) involves the concept component. The CALAP has two steps: (a) a CALAP-screening phase to detect

aphasia and to provide indications, for (b) a CALAP-detailed evaluation phase identifying the impaired language components (Figure 1B). The impairment of a particular component gives rise to a specific language deficit that the CALAP-detailed evaluation can identify through the different subtasks of the scale.

The aim of this study was to determine whether the screening part of the CALAP could be used to diagnose aphasia and whether the detailed evaluation could identify impaired language components, thereby making it possible to improve the effectiveness of rehabilitation.

Method

CALAP Material and Administration

The CALAP is a French-language scale. The material supplied for the CALAP includes the following: a booklet for CALAP administration, a booklet of pictures, and detailed instructions for the rater (see the CALAP administration booklet and CALAP pictures booklet in the online supplemental materials). Figure 1C displays an example of each CALAP task and the number of trials in each task and subtask.

The scale is composed of two different and independent phases: CALAP-screening and CALAP-detailed evaluation. The CALAP-detailed evaluation is administered only if the total CALAP-screening score is lower than the cutoff score of 22 (Figure 1B; see the Results section on external validity). The scale is self-explanatory. The response of the patient is either perfectly correct (otherwise incorrect), making it possible for clinicians inexperienced in language assessment to score and interpret the responses given. Perfectly correct refers only to the expected responses written in the administration booklet. In both screening and detailed evaluation for comprehension and repetition tasks, a score of 2 (perfectly correct), 1 (perfectly correct after one repetition of the trial by the experimenter), or 0 (incorrect) is assigned for each trial. The experimenter repeats the trial if, and only if, requested by the patient. For production and concept, because there is no trial to be repeated by the experimenter, the score of each trial is either 2 (perfectly correct) or 0. Since healthy participants always needed less than five seconds to respond in each trial, beyond this time delay of 5 s, the answer is considered incorrect (0 point). For the CALAP-screening, a total score is computed (0–24 pts) whereas for the CALAP-detailed evaluation, because the aim is to evaluate each component separately for identifying the impaired ones, the score of each subtask is not summed. The maximum scores for the screening and for each subtask of the detailed evaluation are reported Figure 1B.

CALAP—Screening

The screening phase assesses language comprehension, production, and repetition. The trials are purposely set so as to be difficult, to increase the screening sensitivity.

CORE ASSESSMENT OF LANGUAGE PROCESSING

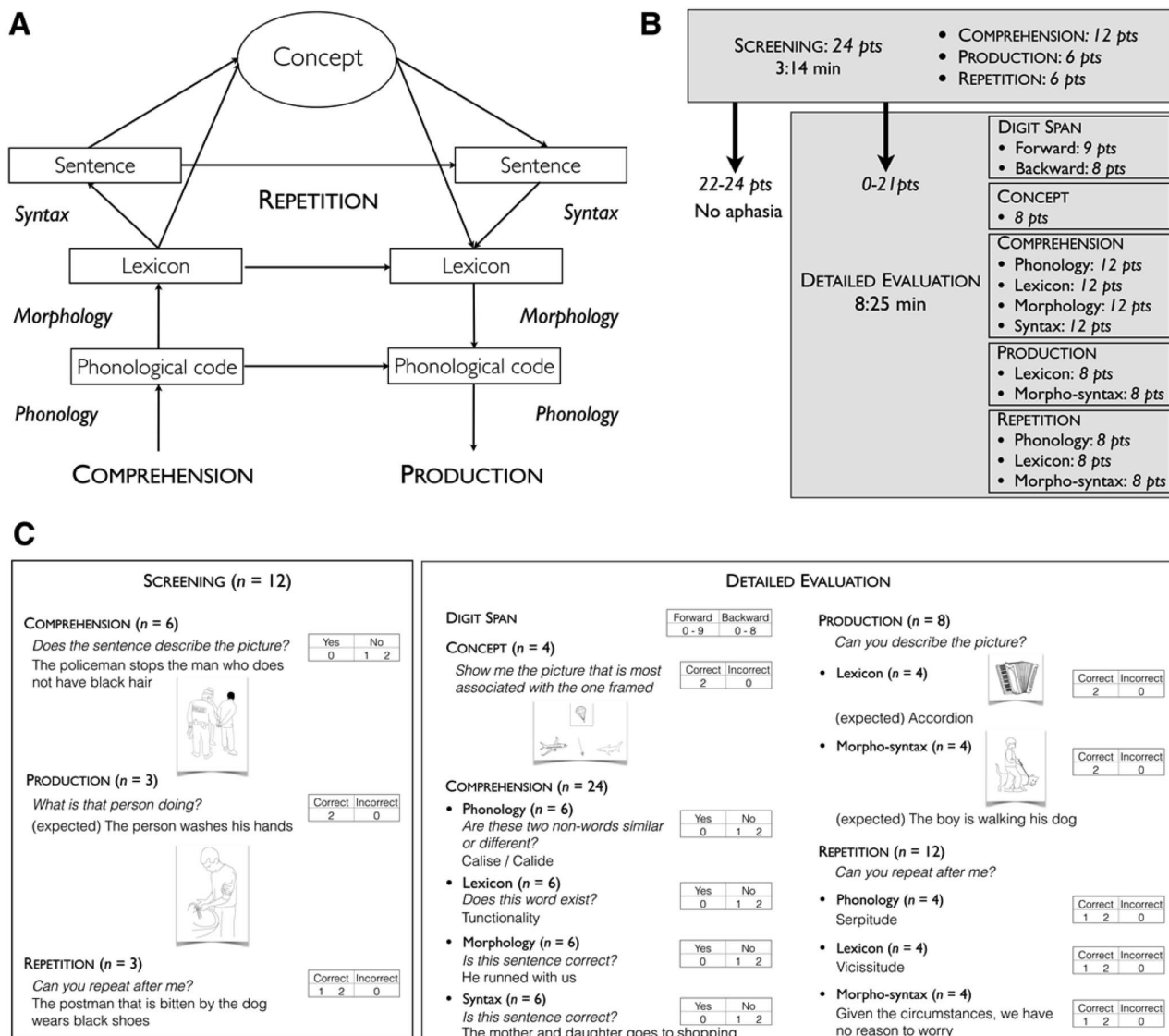


Figure 1. CALAP material and underlying rationale. (A) Simplified model of oral language processing and distinguishing between the comprehension, production, and repetition modality, and, within each modality, the different components: phonology (set of speech sounds), morphology (how the sounds combine to form words), lexicon (words), syntax (how words combine to form sentences) and concept (semantic knowledge). From “Specificity in Rehabilitation of Word Production: A Meta-Analysis and a Case Study,” by C. Jacquemot, E. Dupoux, L. Robotham, and A. C. Bachoud-Lévi, 2012, *Behavioural Neurology*, 25, p. ●●●. Copyright 2012 by ●●●. Adapted with permission. (B) CALAP structure, including the screening and detailed evaluation of each component of the functional model of language processing. (C) CALAP overview: a single trial, together with its scoring grid, is reported for each task, for both screening and detailed evaluation. The instructions given to the participant are shown in italics. The number of trials in each task and subtask is reported in parentheses.

AQ: 8

Comprehension is assessed in a sentence-picture matching task including six complex sentences (mean word number: 9.16 ± 1.47 ; mean syllable number: 10.6 ± 1.14) with a noncanonical grammatical construction: relative clause, negative form and passive construction. The participant is asked to decide whether the sentence heard matches the picture. A “yes” response is expected in half of the trials.

Production capacity is assessed in a picture description task including three pictures that the participant is asked to describe in one sentence. The expected sentences are highly constrained and have noncanonical constructions: a reflexive verbal construction, a reciprocal verbal construction, and a passive form.

Repetition capacity is assessed with a sentence repetition task including three long complex sentences with relative clauses

(mean word number: 10.33 ± 1.52 ; mean syllable number: 14 ± 1.73).

CALAP—Detailed Evaluation

Digit span. Forward and backward digit spans are used to assess verbal short-term memory (STM) and working memory, respectively. Verbal STM can store a limited amount of verbal information for a brief period of time, whereas working memory is used to manipulate the information stored in STM (Cowan, 2008; but see also Aben, Stapert, & Blokland, 2012; Richardson, 2007). In the digit span task, participants are presented with a series of two digits and must immediately repeat them (Wechsler, 1981). The length of the series is increased each time the participant succeeds. In case of failure to repeat the series, another series of the same length is provided. If the participant successfully repeats this series, the test continues, if not, it stops. The digit span is defined as the longest series of digits that the participant can successfully repeat. In the forward version, the participant repeats the digits in the order in which they are presented. In the backward version, the participant repeats the digits in reverse order. Distinguishing between verbal STM and working memory is not relevant in the context of the CALAP scoring, so a digit span score is determined by adding the forward and backward scores independently from the CALAP scores in language subtasks.

Concept. Conceptual knowledge or semantic memory refers to a person's general knowledge of the world and to his personal experience. It is not dependent on language capacity, but may affect language performance if impaired. Conceptual knowledge is assessed through a nonverbal picture-matching task including four trials (Caramazza & Mahon, 2003). In each trial, a target picture is presented with three other pictures. The target picture must be associated with the one among the three pictures which is functionally matched (Jacquemot et al., 2012; concept-depicted mean familiarity¹ value: 2.45 ± 0.9 ; Alario & Ferrand, 1999; mean imaginability² value: 4.39 ± 0.23 ; Bonin et al., 2003; mean emotional³ value: 3.12 ± 0.72 ; Bonin et al., 2003).

Comprehension. For the comprehension task, four subtests assess the phonological, lexical, morphological, and syntactic components of the language-processing model (see Figure 1; Baddecker, Nathan, & Caramazza, 1991; Jacquemot, Dupoux, Decouche, & Bachoud-Lévi, 2006; Jacquemot, Dupoux, Pallier, & Bachoud-Lévi, 2002; Martin, 2006; Miceli, 1994). In these subtests, a yes response is expected in half of the trials. Each subtest consists of six trials and begins with a training trial. The phonological component is assessed with a task in which the participant is asked to discriminate between two-syllable pseudowords (nonsense words phonologically legal in French). When different, the two pseudowords differ by a single phonological feature. The lexical component is assessed with a lexical-decision task in which the participant must determine whether or not the item is a French word (mean item length: 4 ± 0.89). The items presented are words with a very low frequency (mean frequency: 0.47 ± 0.73 data from www.lexique.org), and pseudowords differing from a word by a single phoneme. The morphological component is assessed through an anomaly detection task including six sentences (mean words per sentence: 5.33 ± 1.032 ; mean syllables: 8.33 ± 1.50) that may or may not contain a morphological error. The participant must decide whether or not the sentence is correct. Morphological errors

are incorrect conjugation forms that, although incorrect, are not aberrant in French. The syntactic component is assessed with an anomaly detection task including six sentences (mean words per sentence: 5.16 ± 1.47 ; mean syllables: 8.33 ± 1.50) that may or may not contain a syntactic error. Syntactic errors are incorrect gender and number agreements and an incorrect reflexive verbal construction not aberrant in French.

Production. In the production task, two subtests are used to assess the lexical and morpho-syntactic components (Bock & Levelt, 1994). Each subtest consists of four trials and begins with a training trial. The lexical component is assessed through a picture-naming task including long low-frequency words (mean number of syllables: 3.75 ± 0.5 , mean frequency: 7.16 ± 10.17). There are two nouns, a verb, and a number. The morpho-syntactic component is assessed with a picture description task. The sentences to be produced are of the subject-verb-object type with low frequency verbs and objects (mean frequency: 48.9 ± 60.64) and contain transitive and reflexive verbal constructions.

Repetition. The repetition task consists of three subtests for assessing the phonological, lexical, and morpho-syntactic components (Jacquemot et al., 2007). The phonological component is assessed with a pseudoword repetition task including four pseudowords (mean number of syllables: 3.5 ± 1.29) that the participant is asked to repeat. The lexical component is assessed with a repetition task including four long low-frequency words (mean number of syllables: 4 ± 0.81 ; mean frequency: 0.17 ± 0.28). The morpho-syntactic component is assessed with a repetition task including four long sentences (mean number of words: 10.75 ± 1.26 ; mean number of syllables: 15.25 ± 2.36) with noncanonical constructions: relative clause, subject inversion, interrogative or passive construction, and a nonplausible sentence.

Participants

Consecutive sampling of patients with neurological disorders ($n = 101$) and healthy volunteers with no known neurological disorders ($n = 88$) was conducted, respectively, in the Neurology Department of Créteil Hospital (France) and in Paris, between October 2011 and June 2015. All patients and healthy participants were tested with the CALAP ($n = 189$, 82 men; $M_{\text{age}} = 49.1 \pm 18.4$ years; Table 1). For inclusion, the patients had to be French speakers over the age of 18 years. Severe dementia was an exclusion criterion. The causes of patients' neurological disorders are reported Figure 2. The CALAP was either administrated by a speech-language pathologist ($n = 4$), a neurologist ($n = 1$), or a neuropsychologist ($n = 2$). None of them received training for CALAP administration. All patients attended a consultation with a speech-language pathologist at hospital. They were classified as patients unequivocally without aphasia ($n = 29$) or as patients with possible aphasia ($n = 72$). The latter underwent the French version of the BDAE (Mazaux & Orgogozo, 2007), used as the gold

¹ Familiarity value refers to the familiarity of the concept depicted, rated on a 5-point scale (1 = very unfamiliar object, 5 = very familiar object).

² Imageability value refers to the mental image of the concept depicted, rated on a 5-point scale (1 = very difficult and slow construction of the mental image, 5 = very easy and quick construction of the mental image).

³ Emotional value refers to the emotional valence of the concept depicted, rated on a 5-point scale (1 = very unfamiliar object, 5 = very familiar object).

Table 1
Demographic Data for the Participants

Participants	Status	N	Age	Education level	Laterality (right-handed)
Neurology patients	Aphasia	44	64.8 (15.3)	11 (3.6)	38
	No aphasia	57	53.1 (14.8)	12.6 (3.2)	51
Healthy participants	No aphasia	88	38.7 (15.2)	14.3 (2.3)	81

standard for aphasia assessment in this study. These patients were then classified as patients with aphasia ($n = 44$) or not ($n = 28$) on the basis of their BDAE results in the evaluation of oral comprehension, production, and repetition (Flamand-Roze et al., 2011; see Figure 2). The local ethics committee approved the study and participants gave written informed consent.

Statistical Methodology

In this study, all participants performed both the CALAP-screening and CALAP-detailed evaluation. The scores of participants with aphasia ($n = 44$) and without aphasia ($n = 145$) were compared in Kruskal-Wallis rank-sum tests. The effect of age and educational level on CALAP-screening and CALAP-detailed evaluation scores was assessed in healthy participants, with Pearson’s correlation analyses (two-tailed, $n = 88$). The validation process is

displayed in Figure 2. All analyses were performed with R.3.3.2 software.

External Validity

Aphasia/no aphasia status from BDAE outcome was used to determine the cutoff score for CALAP-screening by receiver operating characteristic (ROC) analysis ($n = 189$); sensitivity and specificity were calculated as proportions, with 95% bootstrap (2,000 resamples) confidence intervals (CIs).

Concurrent Validity

We assessed the concordance between CALAP-detailed evaluation and BDAE scores, by calculating Pearson correlation coefficients between the CALAP and BDAE scores, separately for the

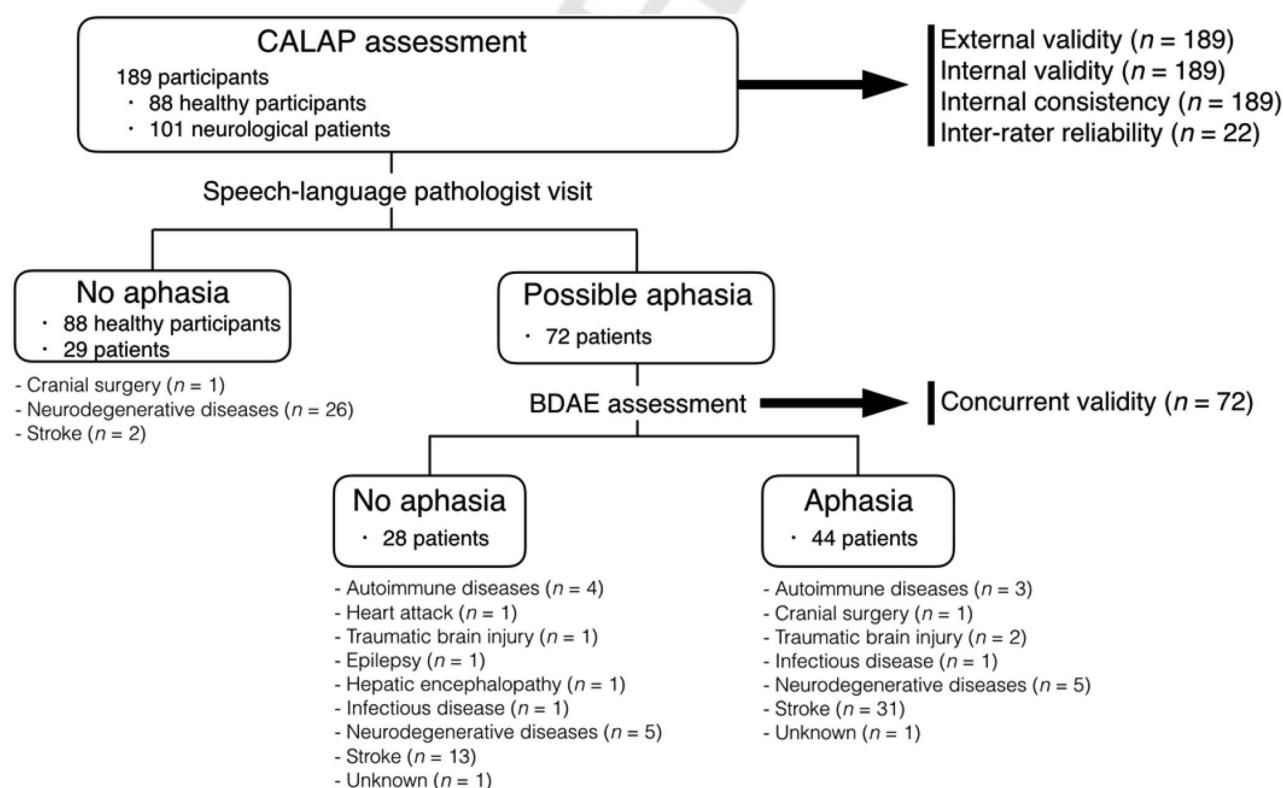


Figure 2. Design of the validation study and cause of the neurological disorder in patients with and without aphasia ($n = 101$).

comprehension, production, and repetition tasks on the subgroup of patients who were tested both with the CALAP and the BDAE (two-tailed, $n = 72$ patients, including 44 patients with aphasia).

Internal Validity

We assessed the correlation between the CALAP-screening and CALAP-detailed evaluation scores, by calculating Pearson correlation coefficients (two-tailed, $n = 189$) between the scores for corresponding tasks in the screening and detailed evaluation (e.g., between the comprehension score in screening and the comprehension score in the detailed evaluation, etc.).

Internal Consistency

Principal component analysis (PCA) was used to explore the relationship between the CALAP-screening tasks with no a priori hypothesis and data reduction possibility. We assessed the weight of each task (comprehension, production, repetition) on scaled and centered data using singular value decomposition. The internal consistency of CALAP-screening, a measurement of reliability based on the internal correlation of screening items, was assessed by calculating Cronbach's alpha (pairwise, $n = 189$). Internal consistency was also used to measure the correlation between the different items of the CALAP-detailed evaluation. It was assessed by calculating Cronbach's alpha for all tasks of the detailed evaluation, together and separately: digit span, concept, comprehension, production, and repetition.

Interrater Reliability

Two experimenters rated the responses of a subgroup of participants simultaneously and independently: one experimenter administered the CALAP to the participant, and both rated the responses obtained ($n = 22$ participants, including seven patients with aphasia). The aphasic status of the participants was masked to the raters. The intraclass correlation coefficient (ICC) was calculated to evaluate interrater reliability.

Results

The mean duration of the CALAP-screening was 3.14 ± 1.23 min and the total duration of the CALAP (screening and detailed evaluation) was 11.39 ± 4.1 min (Figure 1B). All scores for both the CALAP-screening and CALAP-detailed evaluation (digit span, concept, comprehension, production, and repetition) were lower in participants with aphasia than in participants without aphasia (Table 2). Age and education had no significant effect on CALAP-screening scores (Pearson correlation coefficient for age: -0.04 , 95% CI $[-0.25, 0.17]$; $p = .8$, and for education: 0.09 , 95% CI $[-0.12, 0.3]$; $p = .4$), nor on CALAP-detailed evaluation scores (Pearson correlation coefficient for age: -0.14 , 95% CI $[-0.2, -0.17]$; $p = .1$, and for education: 0.09 , 95% CI $[-0.12, 0.3]$; $p = .2$).

External Validity

The ROC analysis identified a cutoff score of 22 (Figure 3) as the best compromise between sensitivity (the ability to identify patients with aphasia correctly; 1, 95% CI [1, 1]) and specificity

(the ability to identify patients without aphasia correctly, 0.72, 95% CI [0.65, 0.79]).

Concurrent Validity

The CALAP-detailed evaluation and BDAE scores were highly correlated. Pearson's correlation coefficient⁴ was 0.62, 95% CI [0.46, 0.75], for comprehension; 0.83, 95% CI [0.75, 0.89], for production; and 0.74, 95% CI [0.61, 0.83], for production (all p values <0.001).

Internal Validity

The CALAP-screening and CALAP-detailed evaluation scores were highly correlated. Pearson's correlation coefficient⁵ was 0.56, 95% CI [0.45, 0.65], for comprehension; 0.77, 95% CI [0.71, 0.82], for production; and 0.74, 95% CI [0.61, 0.83], for repetition (all p values <0.001).

Internal Consistency

For CALAP-screening, the first two components of the PCA (eigenvalues, 2.26 and 0.52), accounted for 93% of the total variance. Each task contributed equally (comprehension -0.53 , production: -0.59 , repetition: -0.61) to the first principal component (accounting for 76% of the variance). The second component only accounted for 17% of the variance with a larger contribution of the comprehension component (comprehension -0.84 , production: 0.47, repetition: 0.27).

Internal consistency was good for the CALAP-screening score (Cronbach's alpha: 0.88, 95% CI [0.8, 0.86]), and excellent for the CALAP-detailed evaluation score (Cronbach's alpha: 0.94, 95% CI [0.93, 0.95]; Kline, 2000; Taber, 2017). Calculated separately for the various tasks, Cronbach's alpha was 0.85, 95% CI [0.80, 0.89], for digit span; 0.45, 95% CI [0.34, 0.56], for concept; 0.85, 95% CI [0.83, 0.88], for comprehension; 0.80, 95% CI [0.76, 0.85], for production; and 0.94, 95% CI [0.92, 0.95], for repetition.

Interrater Reliability

Interrater reliability was excellent for CALAP-screening score (ICC: 0.99, 95% CI [0.98, 0.99], $p < .001$) and CALAP-detailed evaluation⁶ scores (ICC of 0.99, 95% CI [0.99, 0.99] for digit span; 0.95, 95% CI [0.88, 0.98] for comprehension; 0.95, 95% CI [0.88, 0.98] for production; and 0.99, 95% CI [0.99, 0.99] for repetition; all p values <0.001).

Discussion

Efficient and effective functional diagnosis of aphasia remains an unsolved issue. We designed the CALAP, a new short and easy-to-use scale, usable by clinicians nonexpert in language as-

⁴ Spearman correlation coefficients are reported in Table 1 in the online supplemental materials.

⁵ Spearman correlation coefficients are reported in Table 1 in the online supplemental materials.

⁶ Since concept was equally rated by the two experimenters for all the participants and the standard deviation equal to zero, the ICC could not be calculated.

Table 2
Mean (SD) Scores for CALAP Screening and Detailed Evaluation by Aphasia Status (No Aphasia and Aphasia)

CALAP	No aphasia (n = 145)	Aphasia (n = 44)	Kruskal-Wallis χ^2	p value
Screening total (24 pts)	23.0 (1.6) [22.7, 23.3]	15 (5)	96.2	$p < .001$
Digit span total (17 pts)	10.0 (2.4) [9.6, 10.4]	6.2 (2.3)	63.5	$p < .001$
Concept total (8 pts)	8 (.0) [8, 8]	7.5 (1.2)	22.9	$p < .001$
Comprehension total (48 pts)	47.5 (1.1) [47.3, 47.7]	41.2 (6.6)	80.3	$p < .001$
<i>Lexicon</i> (12 pts)	11.8 (.7) [11.7, 11.9]	10.4 (1.9)	44.1	$p < .001^a$
<i>Morphology</i> (12 pts)	11.9 (.4) [11.8, 12.0]	10.4 (1.9)	54.9	$p < .001^a$
<i>Phonology</i> (12 pts)	11.8 (.5) [11.7, 11.9]	10.1 (2.5)	41.6	$p < .001^a$
<i>Syntax</i> (12 pts)	12.0 (.2) [12, 12]	10.3 (2.1)	85.0	$p < .001^a$
Production total (16 pts)	15.2 (1.3) [15.0, 15.4]	9.7 (4.7)	74.3	$p < .001$
<i>Lexicon</i> (8 pts)	7.5 (1.0) [7.3, 7.7]	4.5 (2.5)	67.1	$p < .001^a$
<i>Morpho-syntax</i> (8 pts)	7.8 (.8) [7.7, 7.9]	5.2 (2.9)	64.7	$p < .001^a$
Repetition total (24 pts)	23.1 (1.6) [22.8, 23.4]	12.5 (8.1)	86.9	$p < .001$
<i>Lexicon</i> (8 pts)	7.9 (.5) [7.8, 8.0]	5.2 (3.0)	61.6	$p < .001^a$
<i>Phonology</i> (8 pts)	7.7 (.6) [7.6, 7.8]	4.5 (3.0)	65.5	$p < .001^a$
<i>Morpho-syntax</i> (8 pts)	7.5 (1.0) [7.3, 7.7]	2.8 (2.7)	93.1	$p < .001^a$

Note. The maximum score is reported in parentheses for each task: Screening, Digit span, Concept, Comprehension, Production, and Repetition and for each subtask (in italics). For participants without aphasia, the 95% confidence interval is reported in brackets. All the components assessed in the detailed evaluation are reported. The p values presented are those for Kruskal-Wallis tests for an effect of group (aphasia/no aphasia status) for the various components of the Core Assessment of Language Processing (CALAP).

^a Corrected for multiple comparison.

assessment to detect language disorders, to identify the impaired component within a functional model of language processing in less than 12 min. We showed that the CALAP-screening detects 100% of patients with aphasia in the sample we studied. The psychometric properties of CALAP-screening and CALAP-detailed evaluation scores (external and concurrent validity with the BDAE, internal validity and consistency for screening and detailed evaluation scores, and reliability across raters) were found to be very good.

The CALAP is a new tool that can be used by neurologists, neuropsychologists, and speech-language pathologists to guide clinical

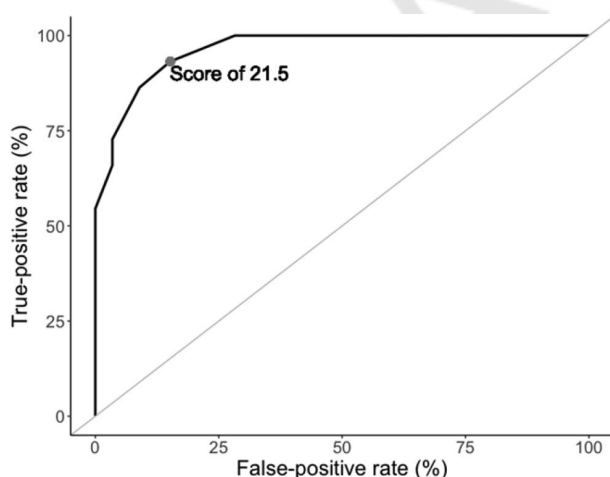


Figure 3. Receiver operating characteristic (ROC) curve showing the sensitivity (true-positive rate) and 1-specificity (false-positive rate) for CALAP screening. The optimal cutoff point was 21.5. The area under the curve was estimated at 96.4 (95% CI [94.2, 98.7]).

practice in diverse contexts. The CALAP-screening is easy to use for the rapid diagnosis of aphasia in clinical settings. The CALAP-detailed evaluation, with its theoretical basis, is useful for both clinical care and follow up. The CALAP-detailed evaluation scores at each subtask (phonology, morphology, lexicon, syntax in comprehension, production, and repetition modalities) have the potential to determine the relative strengths and weaknesses of patients for these various language components and may guide further potential extensive assessment by a speech-language pathologist. For instance, if a patient has a low score in the CALAP subtask “syntax in comprehension,” it will be worth further exploring the syntactic deficit through standardized tests focused on syntax comprehension. The CALAP-detailed evaluation can also be used by speech-language pathologists to guide the design of deficit-based rehabilitation. Indeed, the CALAP, which is based on a functional model of language processing, goes beyond the aphasia syndromes of the Broca–Wernicke classification. This previous classification provided a starting point for understanding aphasia. Although it helped for developing therapeutic interventions, as the practice of rehabilitation has proceeded for many years on the basis of this classification, it was not the most effective way (Marshall, 2010; Nickels, 2002; Nickels & Howard, 1995). For instance, in a patient with a Wernicke profile, several components can be impaired (lexicon in comprehension, syntax in comprehension, lexicon in production . . .) but will not be specifically identified with the classical scales. In contrast, rather than classifying patients into Broca-Wernicke syndromes, the CALAP allows identification of language components which are impaired. As rehabilitation of patients with aphasia is more effective when focused specifically on the impairment (Jacquemot et al., 2012), the CALAP outcome might help in developing a more effective rehabilitation program. For example, if syntax in comprehension is impaired, or at least more impaired than the other components assessed by the CALAP, rehabilitation will be more effective if it

focuses on tasks of sentence comprehension which modulate syntactic complexity (active/passive forms, affirmative/interrogative/negative forms, etc.). By contrast, if “phonology in comprehension” is impaired, rehabilitation will be more effective if it focuses on tasks of increasing difficulty involving pseudoword perception (number of syllables, number of phonemes, phonological complexity, etc.).

The CALAP, which can be administered in one session by clinicians not previously trained and nonexpert in language assessment, is more user-friendly than other scales, such as the BDAE, the Montreal Toulouse 86 (Nespoulous et al., 1992), or the Comprehensive Aphasia Test (Swinburn, Porter, & Howard, 2005), all of which must be administered by experts in language assessment over several sessions. In addition, in contrast to other scales, the CALAP focuses on the evaluation of oral language without requiring any written response. It can therefore be used for patients with writing difficulties, as frequently observed in patients with left hemisphere lesions (Denier et al., 2015). The short duration of the evaluation also makes it possible to test patients with fatigue and concentration difficulties unable to complete longer scales such as the BDAE.

In addition to aphasia, other disorders such as working memory or attention disorders may affect language performance (Jacquemot et al., 2006; Murray, 2000) and most forms of dementia eventually affect language (Patterson, Nestor, & Rogers, 2007). However, as the CALAP assesses conceptual knowledge, and working memory independently of language components, it can distinguish between different diseases in patients with brain impairments. For example, patients with aphasia would not be expected to perform poorly in the concept task, whereas semantic dementia patients will have lower scores for the concept task. Patients with working memory deficits are expected to have a lower score in the digit span task than patients with no working memory deficit, and patients with attention disorders would display a rather homogeneous pattern of impairment in all subtasks instead of an impairment focused on a specific subtask. Thus, these different patterns of impairment may be used for detecting disorders other than language deficit.

The CALAP is a valuable tool not only for research allowing clustering patients on their functional impairment rather than on the neurological usual classification, but also for the diagnosis and rehabilitation of language disorders. The CALAP can be administered during patient hospitalization, home visits, or in medical centers. The rapid diagnosis of aphasia and the identification of patient-specific language deficits should have a major impact on clinical practice and on the orientation of deficit-based rehabilitation. The short duration of the CALAP and its self-explanatory nature should make it possible to implement patient care more rapidly. This aspect is particularly important for patients with neurological disorders, in whom interventions must be initiated as soon as possible (Dobkin, 2005; Laska, Kahan, Hellblom, Murray, & von Arbin, 2011). Thus, it can greatly improve routine patient care and positively impact the burden placed upon caregivers. Although the CALAP is an appropriate tool to use in French-speaking environments, the fact that this scale is based on a functional model of language processing, independent from any language, makes it readily adaptable for universal use.

References

- Aben, B., Stapert, S., & Blokland, A. (2012). About the distinction between working memory and short-term memory. *Frontiers in Psychology, 3*, 301. <http://dx.doi.org/10.3389/fpsyg.2012.00301>
- Alario, F. X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments & Computers, 31*, 531–552. <http://dx.doi.org/10.3758/BF03200732>
- Badecker, W., Nathan, P., & Caramazza, A. (1991). Varieties of sentence comprehension deficits: A case study. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 27*, 311–321. [http://dx.doi.org/10.1016/S0010-9452\(13\)80136-0](http://dx.doi.org/10.1016/S0010-9452(13)80136-0)
- Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). London, UK: Academic Press.
- Bonin, P., Méot, A., Aubert, L., Malardier, N., Niedenthal, P., & Capelle-Toczek, M. C. (2003). Normes de concrétude de valeur d’imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots. *L’Année Psychologique, 103*, 655–694. <http://dx.doi.org/10.3406/psy.2003.29658>
- Breitenstein, C., Grewe, T., Flöel, A., Ziegler, W., Springer, L., Martus, P., . . . the FCET2EC study group. (2017). Intensive speech and language therapy in patients with chronic aphasia after stroke: A randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. *The Lancet, 389*, 1528–1538. [http://dx.doi.org/10.1016/S0140-6736\(17\)30067-3](http://dx.doi.org/10.1016/S0140-6736(17)30067-3)
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*, 177–208. <http://dx.doi.org/10.1080/026432997381664>
- Caramazza, A., & Hillis, A. E. (1993). For a theory of remediation of cognitive deficits. *Neuropsychological Rehabilitation, 3*, 217–234. <http://dx.doi.org/10.1080/09602019308401437>
- Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge: The evidence from category-specific semantic deficits. *Trends in Cognitive Sciences, 7*, 354–361. [http://dx.doi.org/10.1016/S1364-6613\(03\)00159-1](http://dx.doi.org/10.1016/S1364-6613(03)00159-1)
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169*, 323–338. [http://dx.doi.org/10.1016/S0079-6123\(07\)00020-9](http://dx.doi.org/10.1016/S0079-6123(07)00020-9)
- Denier, C., Chassin, O., Vandendries, C., Bayon de la Tour, L., Cauquil, C., Sarov, M., . . . Flamand-Roze, C. (2016). Thrombolysis in stroke patients with isolated aphasia. *Cerebrovascular Diseases (Basel, Switzerland), 41*, 163–169. <http://dx.doi.org/10.1159/000442303>
- Denier, C., Flamand-Roze, C., Dib, F., Yeung, J., Solignac, M., de la Tour, L. B., . . . Pico, F. (2015). Aphasia in stroke patients: Early outcome following thrombolysis. *Aphasiology, 29*, 442–456.
- Dobkin, B. H. (2005). Rehabilitation after stroke. *The New England Journal of Medicine, 352*, 1677–1684. <http://dx.doi.org/10.1056/NEJMc043511>
- Duncan, P. W., Zorowitz, R., Bates, B., Choi, J. Y., Glasberg, J. J., Graham, G. D., . . . Reker, D. (2005). Management of adult stroke rehabilitation care. *Stroke, 36*, e100–e143. <http://dx.doi.org/10.1161/01.STR.0000180861.54180.FF>
- Ellis, C., Simpson, A. N., Bonilha, H., Mauldin, P. D., & Simpson, K. N. (2012). The one-year attributable cost of poststroke aphasia. *Stroke, 43*, 1429–1431. <http://dx.doi.org/10.1161/STROKEAHA.111.647339>
- Flamand-Roze, C., Falissard, B., Roze, E., Maintigneux, L., Beziz, J., Chacon, A., . . . Denier, C. (2011). Validation of a new language screening tool for patients with acute stroke: The Language Screening Test (LAST). *Stroke, 42*, 1224–1229. <http://dx.doi.org/10.1161/STROKEAHA.110.609503>
- Goodglass, H., & Kaplan, E. (1972). *The assesment of aphasia and related disorders*. Philadelphia, PA: Lea & Febiger.
- Hillis, A. E. (1998). Treatment of naming disorders: New issues regarding old therapies. *Journal of the International Neuropsychological Society, 4*, 648–660. <http://dx.doi.org/10.1017/S135561779846613X>

- Jacquemot, C., Dupoux, E., & Bachoud-Lévi, A. C. (2007). Breaking the mirror: Asymmetrical disconnection between the phonological input and output codes. *Cognitive Neuropsychology*, *24*, 3–22. <http://dx.doi.org/10.1080/02643290600683342>
- Jacquemot, C., Dupoux, E., Decouche, O., & Bachoud-Levi, A. C. (2006). Misperception in sentences but not in words: Speech perception and the phonological buffer. *Cognitive Neuropsychology*, *23*, 949–971. <http://dx.doi.org/10.1080/02643290600625749>
- Jacquemot, C., Dupoux, E., Pallier, C., & Bachoud-Lévi, A.-C. (2002). Comprehending spoken words without hearing phonemes: A case study. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *38*, 869–873. [http://dx.doi.org/10.1016/S0010-9452\(08\)70059-5](http://dx.doi.org/10.1016/S0010-9452(08)70059-5)
- Jacquemot, C., Dupoux, E., Robotham, L., & Bachoud-Lévi, A. C. (2012). Specificity in rehabilitation of word production: A meta-analysis and a case study. *Behavioural Neurology*, *25*, 73–101. <http://dx.doi.org/10.1155/2012/418920>
- Kauhanen, M. L., Korpelainen, J. T., Hiltunen, P., Maatta, R., Mononen, H., Brusin, E., . . . Myllyla, V. V. (2000). Aphasia, depression, and non-verbal cognitive impairment in ischaemic stroke. *Cerebrovascular Disease*, *10*, 455–461. doi:16107
- Kirshner, H. S. (2014). Frontotemporal dementia and primary progressive aphasia, a review. *Neuropsychiatric Disease and Treatment*, *10*, 1045–1055. <http://dx.doi.org/10.2147/NDT.S38821>
- Kline, P. (2000). *The handbook of psychological testing*. Florence, KY: Taylor & Francis/Routledge.
- Laska, A. C., Kahan, T., Hellblom, A., Murray, V., & von Arbin, M. (2011). A randomized controlled trial on very early speech and language therapy in acute stroke patients with aphasia. *Cerebrovascular Diseases Extra*, *1*, 66–74. <http://dx.doi.org/10.1159/000329835>
- Laska, A. C., Mårtensson, B., Kahan, T., von Arbin, M., & Murray, V. (2007). Recognition of depression in aphasic stroke patients. *Cerebrovascular Diseases (Basel, Switzerland)*, *24*, 74–79. <http://dx.doi.org/10.1159/000103119>
- Levelt, W. J. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*, 223–232. [http://dx.doi.org/10.1016/S1364-6613\(99\)01319-4](http://dx.doi.org/10.1016/S1364-6613(99)01319-4)
- Marshall, J. (2010). Classification of aphasia: Are there benefits for practice? *Aphasiology*, *24*, 408–412. <http://dx.doi.org/10.1080/02687030802553688>
- Martin, R. C. (2006). The neuropsychology of sentence processing: Where do we stand? *Cognitive Neuropsychology*, *23*, 74–95. <http://dx.doi.org/10.1080/02643290500179987>
- Mazaux, J. M., & Orgogozo, J. (2007). *HDAE, échelle d'évaluation de l'aphasie adaptation française du Boston diagnostic aphasia examination de H. Goodglass et E. Kaplan*. Paris, France: Les Éd. du Centre de psychologie appliquée.
- Mesulam, M. M. (2001). Primary progressive aphasia. *Annals of Neurology*, *49*, 425–432. <http://dx.doi.org/10.1002/ana.91>
- Miceli, G. (1994). Morphological errors and the representation of morphology in the lexical-semantic system. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, *346*, 79–87. <http://dx.doi.org/10.1098/rstb.1994.0131>
- Miceli, G., Capasso, R., & Caramazza, A. (1994). The interaction of lexical and sublexical processes in reading, writing and repetition. *Neuropsychologia*, *32*, 317–333. [http://dx.doi.org/10.1016/0028-3932\(94\)90134-1](http://dx.doi.org/10.1016/0028-3932(94)90134-1)
- Murray, L. L. (2000). The effects of varying attentional demands on the word retrieval skills of adults with aphasia, right hemisphere brain damage, or no brain damage. *Brain and Language*, *72*, 40–72. <http://dx.doi.org/10.1006/brln.1999.2281>
- Nespoulous, J. L., Lecours, A. R., Lafond, D., Lemay, M. A., Puel, M., Joannette, Y., & Rascol, A. (1992). *Protocole Montréal-Toulouse d'examen linguistique de l'aphasie: MT-86 module standard initial, M1b*. Isbergues, France: Ortho Edition.
- Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising and reviewing. *Aphasiology*, *16*, 935–979. <http://dx.doi.org/10.1080/02687030244000563>
- Nickels, L., & Howard, D. (1995). Aphasic naming: What matters? *Neuropsychologia*, *33*, 1281–1303. [http://dx.doi.org/10.1016/0028-3932\(95\)00102-9](http://dx.doi.org/10.1016/0028-3932(95)00102-9)
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–987. <http://dx.doi.org/10.1038/nrn2277>
- Richardson, J. T. (2007). Measures of short-term memory: A historical review. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *43*, 635–650. [http://dx.doi.org/10.1016/S0010-9452\(08\)70493-3](http://dx.doi.org/10.1016/S0010-9452(08)70493-3)
- Swinburn, K., Porter, G., & Howard, D. (2005). *The Comprehensive Aphasia Test*. Hove, UK: Psychology Press.
- Taber, K. (2017, June). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*. Advance online publication.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised manual*. New York, NY: Psychological Corporation.

Received May 4, 2018

Revision received October 9, 2018

Accepted October 31, 2018 ■