

## BEPS: Development, validation, and normative data of a sentence production test in French

Valérie Coulombe<sup>a,b</sup>, Marion Fossard<sup>c</sup>, and Laura Monetta<sup>a,b</sup>

<sup>a</sup>Department of Rehabilitation, Faculty of Medicine, Laval University, Quebec City, Quebec, Canada;

<sup>b</sup>CERVO Brain Research Centre, Centre intégré universitaire de santé et de services sociaux de la Capitale-Nationale, Quebec City, Québec, Canada;

<sup>c</sup>Institute of Language Sciences and Communication, Université de Neuchâtel, Neuchâtel, Switzerland

### ABSTRACT

Assessing one's sentence production ability can help to identify difficulties experienced by individuals with aphasia and enables targeted intervention. Speech-language pathologists working with the francophone population must contend with the current lack of a sentence production evaluation tool. The aim of this article is to present the development, validation, and normative data of a French-language test used to evaluate the production of sentences: the BEPS (Batterie d'Évaluation de la Production Syntaxique). The BEPS includes four tasks based on Bock and Levelt's model of sentence production, namely (1) Verb Naming Task, (2) Verb Inflection Task, (3) Constituent Assembly Task, and (4) Thematic Role Assignment Task. For each task, various psycholinguistic parameters were controlled or manipulated. The BEPS had good content and construct validity, excellent test-retest reliability and good internal consistency. This battery fulfills the current needs for clinical tool in the franco-phone population when assessing sentence production capacities. Normative data are pre-sented for individuals 19–79 years of age ( $n = 191$ ). Raw scores were converted to percentiles for each task. The BEPS represents significant assistance to help clinicians identify the origins of the syntactic production deficits of patients with poststroke and degenerative diseases.

**KEYWORDS** Agrammatism; aphasia; assessment; syntactic production; verb retrieval

### Introduction

Aphasia is an acquired language disorder caused by a brain injury, primarily a stroke or degenerative disease that usually affects the left hemisphere of the brain. The different types of aphasia are classified according to the pathology and the manifestations of language difficulties that result from it (Verstichel & Cambier, 2005). There are various types of language disorders associated with aphasia; one of them is a sentence production deficit (SPD). SPD is often observed in nonfluent aphasia profiles (Caramazza & Zurif, 1976; Mitchum & Berndt, 2008).

SPDs are commonly referred to as agrammatism and generally imply an inability to formulate complex sentences. Some examples are the omission of grammatical morphemes, verb retrieval difficulties, exclusive use of verbs bearing few grammatical roles, and the inability to adequately allocate the thematic role of a verb (Kemmerer, 2014; Pillon, 2014). Depending on the individual, these syntactic difficulties can be

combined in diverse ways and gives rise to distinct profiles, irrespective of the type of aphasia. These heterogeneous origins of SPDs make the conclusion of agrammatism insufficient in a clinical setting, and it highlights the importance of being able to determine the specific underlying deficit of individuals when producing sentences. The ability to identify the specific underlying problem will help to better target the intervention for each individual with aphasia (IWA) (Bastiaanse, Edwards, Mass, & Rispens, 2003; Pillon, 2014).

Bock and Levelt's (1994) theoretical model of sentence production provides a sequential point of view that allows for the evaluation of each step of sentence production. The four levels of sentence processing, defined as follows, can be explored independently:

1. Conceptual level: This level corresponds to a pre-linguistic level, in which a mental representation of the message to deliver is conceived.

2. Functional level: This level is subdivided into two sublevels. First, it involves the lexical selection of the concepts previously activated in the conceptual level (i.e., the lemma selection from the lexicon), including various verbal information, such as combinatorial (phrasal and subcategorization) and featural information (e.g., diacritic markers for tense, aspect). Second, the assignment of thematic roles (agent, theme; *who* did what to *whom*?) to lemmas and the encoding in terms of their grammatical structure (subject, verb, object) is performed.
3. Positional level: This level has two sublevels of processing: constituent assembly and inflectional processes. First, the constituents (i.e., lemmas) are organized into phrasal constituents and placed in a specific order respecting a syntactic organizational frame. Second, the grammatical morphemes (e.g., number and tense) and function words (prepositions, determiners, and auxiliary verbs) are added to provide the grammatical structure of the sentences.
4. Phonological level: This level allows the establishment of the phonological encoding, where the forms of the words on the phonological level are recovered, and the prosody and the rhythm are added to the statement.

An SPD conclusion should then specify the origin as a deficit in lexical access, thematic role processing, verb inflection, or grammatical ordering or a combination of the various items.

A recent systematic review of all existing tests for the assessment of sentence production in English has resulted in 20 different clinical and research tests (Mehri & Jalaie, 2014). Among them, some are validated and easy to use in clinics (e.g., Wilshire, Lukkien, & Burmester, 2014). Others, much rarer, are constructed in such a way that they allow the assessment of the underlying deficits in sentence production, such as the Verb and Sentence Test (VAST; Bastiaanse, Edwards, & Rispen, 2002).

The lack of assessment tools targeting the underlying deficits of SPD appears in French, too. Thus, it is difficult for francophone speech-language pathologists to evaluate the abilities of an IWA to produce a sentence. For reference, the Fine Morphosyntactic Expression Test (TEMF; Bernaert-Paul & Simonin, 2011) does exist to detect the deficient structures of European French speakers. This test forces the patient to produce various syntactic structures, but it does not target the underlying nature of the SPD and has

no known theoretical foundations. Currently, in clinical settings, the production of sentences is generally evaluated through spontaneous conversation and through the narrative discourse of the Montreal-Toulouse Battery (MT-86) bank robbery scenario (Nespoulous et al., 1992), which consists in describing an image. Both spontaneous production and image description pose several challenges in the evaluation of sentence production of IWAs. It is difficult to impose a specific syntactic structure on these types of tasks, and speakers sometimes avoid more complex structures. When a speaker does make an error, it is difficult to determine with certainty the initial intention. Spontaneous conversation is rich in information, but it is difficult to control the topic and naturally obtain the desired information. Adding to the complexity, the duration of the exchange is not controlled, which can result in a corpus that is not optimal for analysis. Be that as it may, clinicians must deal with the lack of standardized and validated clinical assessment tools. Currently, no known instrument in French can highlight the source of the difficulties of syntactic production and explain them theoretically.

The primary objective of this article is to present the development, validation and normative data of a new French-language assessment tool based on a theoretical model that allows clinicians to detect sentence production deficits in IWAs with poststroke or neurodegenerative lesions. The interventions can be adjusted to the targeted problem, the management of patients can be improved, and the patients' chance of recovery may be optimized. Three studies are presented in this article: Study I: development of the Sentence Production Assessment Battery (BEPS); Study II: validation and reliability; and Study III: normative data establishment. The Ethical Committee of CERVO Research Center approved the study (project # 2016-148).

## Study I. Development of the BEPS

The purpose of Study I was to create the tasks and to choose the stimuli of the BEPS.

### Materials and methods

#### Content validity

The BEPS is a French-language tool for evaluating syntactic production in IWAs. The assessment tasks were selected for the assessment battery according to the theoretical model of sentence production by Bock and Levelt (1994). Two professors in speech-language




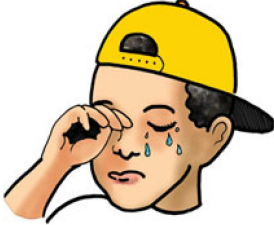
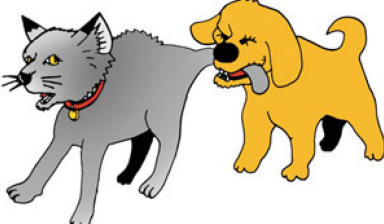

<p><b>Qu'est-ce que le policier fait?</b></p>  <p>Task 1. Verb Naming</p>	<p>Hier, le garçon _____.</p> <p>(pleurer)</p>  <p>Task 2. Verb Inflection</p>							
<table border="1" data-bbox="162 607 788 656"> <tr> <td>CHAT</td> <td>EST</td> <td>LE</td> <td>CHIEN</td> <td>PAR</td> <td>LE</td> <td>MORDU</td> </tr> </table>  <p>Task 3. Constituent Assembly</p>	CHAT	EST	LE	CHIEN	PAR	LE	MORDU	<p>femme gâteau couper</p>  <p>Task 4. Thematic Role Assignment</p>
CHAT	EST	LE	CHIEN	PAR	LE	MORDU		

Figure 1. Examples of the four tasks of the BEPS.

pathology (LM and MF) and a speech-language pathology graduate student created the tasks of the BEPS.

The BEPS was based on the same view of syntactic production as the VAST (Bastiaanse et al., 2002), but each task aimed to target a specific stage of the grammatical encoding (functional and positional levels) as described by Bock and Levelt (1994). The BEPS was explicitly developed for the French-language populations, and it contained tasks relevant for the syntactic production in this language.

The following four language assessment tasks were selected for the BEPS: A Verb Naming Task (Task 1) that evaluates the lexical selection; a Verb Inflection task (Task 2) that evaluates the conjugation of verbs; a Constituent Assembly Task (anagram task, Task 3) that evaluates the constituent assembly and a Thematic Role Assignment Task (Task 4) that evaluates the function assignment. Examples of the tasks are presented in Figure 1.

The functional level was evaluated by the Verb Naming Task, which involves the identification of the lexical concepts and the selection of lemmas in naming the action presented in the picture. The stimuli were all verbs because of their importance in sentence production. The second stage of the functional level, the function assignment, was evaluated by the Thematic Role Assignment Task. This task

implemented the evaluation of the assignment of thematic roles and syntactic functions to lemmas (given in the task). Participants had to produce a sentence with the given words, and they had to start it with the targeted word (indicated by an arrow). This task involved the production of different sentence structures that was either an active or passive voice. The sentences had different verbs requiring two or three arguments. The Thematic Role Assignment Task offered the opportunity to assess sentence construction without the influence of the retrieval of lexical items and allowed the evaluation of function assignment when the structures of sentences are manipulated. This ability of the BEPS was relevant for the targeted assessment of SPDs.

The positional level included two tasks: The Constituent Assembly Task and the Verb Inflection Task. The Constituent Assembly Task was adapted from the Sentence Order Task of Saffran, Schwartz, and Marin (1980). Participants had to assemble words in order to produce a grammatical sentence that described a picture. The major difference between this task and the initial Sentence Order Task was that participants had to construct sentences from isolated words rather than phrases. This adaptation added complexity in grammatical terms to the task. For the Verb Inflection Task, participants had to produce

verbs with the correct tense inflection (past, present, or future), based on a time adverb presented in a carrier sentence. In French, the ability to investigate various tense and persons is essential. The Verb Inflection Task had the advantage over existing tasks to assess essential inflectional abilities in the French-language, including the conjugations of regular and irregular verbs.

The stimuli for the tasks were picked based on their psycholinguistic characteristics. Verbs had to present a high frequency according to a score of 20 or higher in a linguistic database (Lexique.org) and to be subjectively imageable. Once selected, the stimuli created for the BEPS were all drawn by the same artist. All the pictures were original artworks.

### **Face validity**

The face validity of the BEPS was assessed by six professionals (four speech-language pathologists and two neuropsychologists from Western Switzerland and Quebec, Canada). They were invited to review the preliminary version of the BEPS by means of a questionnaire specifically developed to evaluate its appropriateness, usefulness, ease-of-use, and clarity (administration sheet, instructions). Likert scales ranging from 1 (e.g., inappropriate) to 5 (e.g., absolutely appropriate) and open questions were used to collect information about the face validity of the overall battery and the four specific tasks. The questionnaires were analyzed qualitatively and quantitatively. Adjustments were made to all the elements where a lack of clarity in the instructions or the procedure was identified. The cutoff for face validity was 4 points on the 5-point Likert scale.

### **Pilot study**

Once the tasks and items were chosen, a pilot study was conducted to develop the final version of the battery. This pilot test with a preliminary version of the BEPS comprised the administration protocol for each of the four tasks including the stimuli, the psycholinguistic variables controlled and manipulated, and the instructions for standardized administration and scoring. This preliminary version of the BEPS was administered to nine healthy participants. They were tested by an experimenter to estimate the ability of participants without SPD to name the desired verb illustrated in the Verb Naming Task (Task 1), and to produce inflections (Task 2) or sentences expected (Tasks 3 and 4).

## **Results**

Professionals who reviewed the battery indicated that the BEPS was appropriate to evaluate syntactic production in poststroke and degenerative aphasia, that it was easy to use and useful for clinical settings. Adjustments were made to the instructions of the Verb Naming Task and the Verb Inflection Task because professionals identified a lack of clarity (an average evaluation of 3.88 on the 5-point Likert scale). As a result of the pilot study, five pictures were removed from the Verb Naming Task and the entire battery. In addition, 12 items from the Verb Inflection Task, three from the Constituent Assembly Task and four from the Thematic Role Assignment Task were removed to respect the balance between psycholinguistic parameters in each task.

The final version of the BEPS had 18 stimuli for the Verb Naming Task, 24 for the Verb Inflection Task, 12 for the Constituent Assembly Task and 8 for the Thematic Role Assignment Task. The tasks and characteristics of the items chosen as a result of the development process are presented in Table 1.

### **Summary**

The BEPS was developed based on the scientific literature and with a comprehensive test development approach. It is the first assessment battery in French to be able to identify the nature of the SPDs. The tasks and items of the BEPS were considered representative of the sentences production processes in French and useful for clinicians.

## **Study II. Validity and reliability of the BEPS**

The purpose of Study II was to provide data on the BEPS's construct validity and reliability and to provide preliminary data on the criterion validity.

### **Materials and methods**

#### **Participants**

The BEPS was designed to assess syntactic impairments in poststroke and neurodegenerative aphasia. Validity was established by including participants presenting poststroke aphasia or primary progressive aphasia (PPA), and a control group. Six participants with poststroke aphasia and six participants with PPA were recruited from an aphasia association (APIA-AVC, Quebec, Canada) and from a list of participants who had already participated in studies of the contributing research centers and had agreed to be contacted again.

**Table 1.** Cognitive domains, tasks, and item characteristics of the BEPS.

Cognitive domain	Task	Stimulus type <sup>a</sup>	Psycholinguistic variables (number of items)
Lexical access	1: Verb Naming	18 verb pictures	
Inflection	2: Verb Inflection	24 written and spoken sentences beginning with a time adverb and missing the targeted verb	Tenses: past (8), present (8), future (8) Persons: third person singular (12), plural (12) Regularity: regular (12), irregular (12)
Constituent assembly	3: Constituent Assembly	12 pictures with written words (anagram) Verbs with 2 arguments Reversible sentences	Syntactic structures: passive (4), pronominal (4), cleft (4)
Function Assignment	4: Thematic Role Assignment	8 pictures with written words (lemmas) and a narrow indicating the subject of the sentence	Number of arguments: 2 (6), 3 (2) Voice of verbs: active (4), passive (4)

Note. BEPS: Batterie d'Évaluation de la Production Syntaxique.

<sup>a</sup>All verbs have a high lexical frequency and high imaginability.

Inclusion criteria for IWAs ( $n = 12$ ) taking part in the validation study were (a) diagnosis of chronic (more than 1-year poststroke) aphasia (mild to moderate) or neurodegenerative aphasia (nonfluent/agrammatic variant of PPA) (b) French as a mother tongue that was used on a daily basis and (c) aged 19 or more. We excluded IWAs with moderate to severe speech comprehension disorders, with a history of or current psychiatric illness, with significant uncorrected vision or audition problems, and with a current or past follow-up (less than three years) in speech-language pathology. Information regarding inclusion and exclusion criteria were obtained from self-reports.

As a control group, 14 healthy control (HC) participants (all from Quebec, Canada) who had agreed to take part in the validation study, and 3 participants with primary progressive apraxia of speech (PPAOS) were included. PPAOS is associated with a progressive loss of speech production without aphasia (no SPD) (Duffy, 2006).

### Procedure

All IWAs ( $n = 12$ ) were assessed with the BEPS, and four of them (with poststroke aphasia) agreed to be assessed twice. Five participants with poststroke aphasia were also assessed with the Dénomination de verbes lexicaux en images Test (DVL 38; Hammerlath, 2001) verb-naming task; the auditory lexical decision task of the Batterie d'Évaluation Cognitive du Langage (BECLA; Macoir, Jean, & Gauthier, 2015); and the image description task of the MT-86 (Nespoulous et al., 1992).

Eleven participants of the 14 healthy controls were available to be assessed a second time for the test-retest reliability and all of the 14 HC participants were assessed with the DVL 38 verb-naming task (Hammerlath, 2001). Participants' characteristics and scores for the validation study are summarized in Table 2.

Participants were tested individually in a quiet room at their home or at the research center. All participants provided informed consent. Research assistants collected written protocols. The BEPS's stimuli were presented on a laptop or tablet computer. Other tasks (DVL 38, BECLA, and MT-86) were administered according to the standardized instructions of the tests. The order of administration of tests was systematically randomized. There was no time limit to complete the tests. The participants received no financial compensation.

### Construct validity

**Convergent validity.** The lexical selection of verbs was assessed with two tasks: The Verb Naming Task (BEPS) and the DVL 38, for 19 participants (14 HC and 5 IWAs). Performances to those tasks were studied with correlation coefficient to establish the convergent validity of the Verb Naming Task of the BEPS, as the two tests aim to assess the same construct. However, no test measures constructs similar to the three other tasks of the BEPS. Thus, convergent validity was only studied for the first task.

**Divergent validity.** To ensure the validity of the BEPS to assess SPD, scores should not be related to constructs not implicated in sentence production, like the acoustic analysis of spoken words. Performances of five IWAs to the Thematic Role Assignment Task and the auditory lexical decision task of the BECLA were studied with correlation coefficients.

**Discriminant validity.** To test whether the BEPS scores distinguished between the performances of participants with and without SPD, comparisons were made between two groups matched for their age and education level. Participants were classified into the SPD group ( $n = 9$ ; 3 IWAs poststroke and 6 PPAs) or the control group with no deficit of sentence production ( $n = 9$ ; 6 HC and 3 PPAOS) as a function of

**Table 2.** Descriptive statistics (mean, standard deviation, median, and quartiles) of the four groups of participants for the validation study.

Group Characteristic	Post-stroke ( <i>n</i> = 6)			PPA ( <i>n</i> = 6)			PPAOS ( <i>n</i> = 3)			HC ( <i>n</i> = 14)			Total <i>n</i>
	<i>M</i> ( <i>SD</i> )	<i>Mdn</i> ( <i>q1–q3</i> )		<i>M</i> ( <i>SD</i> )	<i>Mdn</i> ( <i>q1–q3</i> )		<i>M</i> ( <i>SD</i> )	<i>Mdn</i> ( <i>q1–q3</i> )		<i>M</i> ( <i>SD</i> )	<i>Mdn</i> ( <i>q1–q3</i> )		
Age <sup>a</sup>	64.5 (11.3)	68.5 (51.8–74.3)		77.3 (7.61)	76.5 (71.5–83.0)		72.7 (2.082)	72.0 (71.0–72.0)		51.5 (15.53)	51.5 (47.3–60.8)		29
Education	15.8 (1.94)	16.5 (13.8–17.3)		8.00 (2.37)	8.50 (6.25–9.50)		14.0 (2.65)	13.0 (12.0–13.0)		14.6 (3.16)	13.8 (12.9–17.3)		29
Measure (number of items)													
1: Verb Naming <sup>a</sup> (18)	15.0 (2.53)	15.0 (13.3–17.3)	14	14	14					17.1 (0.95)	17.5 (16.0–18.0)		21
2: Verb Inflection <sup>a</sup> (24)	18.8 (4.49)	21.0 (14.8–22.0)	21	21	21					23.9 (0.27)	24.0 (24.0–24.0)		21
3: Constituent Assembly <sup>a</sup> (12)	9.67 (1.37)	10.0 (9.25–10.3)	5.50 (3.99)	5.50 (3.99)	5.00 (2.25–10.0)	11.0 (10.0–11.0)	11.0 (1.00)	8.00 (8.00–8.00)		11.6 (0.51)	12.0 (11.0–12.0)		29
4: Thematic Role Assignment <sup>a</sup> (8)	6.17 (1.60)	6.00 (4.75–8.00)	4.17 (1.72)	4.17 (1.72)	4.00 (2.75–5.50)	8.00 (8.00–8.00)	8.00 (0.00)			7.8 (0.43)	8.00 (7.75–8.00)		29
BEPS <sup>a</sup> (62)	49.7 (6.35)	51.0 (43.8–54.3)	41	41	41					60.4 (1.40)	61.0 (59.0–62.0)		21
DVL 38 <sup>a</sup> (114)	84.6 (11.2)	82.0 (75.0–95.5)								102.4 (4.67)	102.5 (99.3–106.0)		19
Lexical Decision <sup>a</sup> (20)	18.6 (0.89)	19.0 (18.0–19.0)											5

Note. PPA: Primary progressive aphasia; PPAOS: Primary progressive apraxia of speech; HC: Healthy controls; q1: First quartile; q3: Third quartile; BEPS: Batterie d'Évaluation de la Production Syntaxique.

<sup>a</sup>The data are not normally distributed ( $p < .05$ ).

clinical diagnosis of SPD (MT-86 for the poststroke's group and all PPA participants).

### Reliability

**Test-retest reliability.** The BEPS was administered twice (T1 and T2) to 15 participants (4 IWAs post-stroke and 11 HC) with a minimum of 4-month intervals between administrations.

**Internal consistency.** Reliability was measured based on the degree of interrelatedness among the items of the Verb Naming Task (Task 1) and the Verb Inflection Task (Task 2). The scores for the 18 and 24 items of each task were expected to be correlated because they measured a single ability. The two other tasks of the BEPS led to the expectation that the internal consistency was to be lower because of the higher item-specific variance as a result of the manipulated psycholinguistic variables. The sample studied included 20 participants (6 IWAs poststroke and 14 HC).

### Criterion validity

**Concurrent validity.** To address the concurrent validity, participants in the group of poststroke aphasia ( $n = 5$ ) were identified as having SPD or not, according to their sentence production in the bank robbery description task (MT-86; Nespoulous et al., 1992). Normative data of the BEPS (presented in Table 3) was then used to classify participants' performance as  $\leq 5$ th percentile or within the normal range ( $> 5$ th percentile). The sensitivity and the specificity of the BEPS could not be assessed with great accuracy because of the small sample size. Therefore, we described and analyzed the inconsistencies between the BEPS and the MT-86 (Nespoulous et al., 1992).

### Statistical analysis

For statistical analysis, the software Statistical Package for the Social Sciences (SPSS) v. 25 for Mac (SPSS, Chicago, IL) was used. The alpha risk was set to 5% ( $p = .05$ ). Data demonstrated non-normal distributions (age and BEPS, DVL 38, and BECLA scores) according to the Shapiro-Wilk test of normality.

The association between scores on the BEPS and other assessment instruments was measured using Spearman's correlation coefficients. For the convergent validity, scores were expected to show a significant and positive relation, and for the divergent validity, they were expected to show no significant relation.

**Table 3.** Percentiles stratified by age for the BEPS.

Task (number of items)	Age	<i>M</i>	<i>SD</i>	Percentiles							
				1	2	5	10	15	25	50	95
1: Verb Naming (18)	–	17.2	1.08	13	14	15	16	16	16	18	18
2: Verb Inflection (24)	–	23.8	.759	20	21	22	23	23	24	24	24
3: Constituent Assembly (12)	≤49	11.9	.399	10	10	11	11	12	12	12	12
	50–64	11.6	.816	7	8	10	11	11	11	12	12
	≥65	11.4	.866	8	8	9	10	11	11	12	12
4: Thematic Role Assignment (8)	≤49	7.86	.423	6	6	7	7	8	8	8	8
	50–64	7.62	.799	4	4	6	7	7	7	8	8
	≥65	7.67	.555	6	6	6	7	7	7	8	8
BEPS: Syntactic Production (62)	≤49	60.9	1.37	56	56	57	59	60	60	61	62
	50–64	60.3	1.94	54	54	56	57	58	59	61	62
	≥65	59.6	1.90	54	54	55	56	58	59	60	62

Note. BEPS: Batterie d'Évaluation de la Production Syntaxique. Percentiles stratified by age for the BEPS. Age was not a significant correlate to tasks 1 and 2.

Comparisons between the scores of the SPD and control groups were carried out with unpaired two-sample Wilcoxon rank-sum Tests ( $W_s$ ). The group with SPD was expected to have lower scores than the control group on the four BEPS's tasks. Approximate effects sizes were then calculated for each task from the z-scores (Rosenthal, 1994).

The relationship between BEPS scores at T1 and T2 was measured with an intraclass correlation coefficient (ICC; Two-way mixed effects, absolute agreement). ICC was expected to be  $>.75$ . Internal consistency was measured using Cronbach's alpha coefficient ( $\alpha$ ).

## Results

### Convergent validity

The Spearman's correlation coefficients showed that scores to the verb naming tasks of the BEPS (Task 1;  $Mdn = 17/18$ ,  $IQR = 2$ ) and the DVL 38 ( $Mdn = 101/114$ ,  $IQR = 11$ ) were significantly and positively correlated. This indicates that the same construct was measured in both tasks,  $r_s = .54$ ,  $p = .017$ .

### Divergent validity

The Spearman's correlation coefficient showed that scores to the auditory lexical decision task (BECLA) and the sentence production tasks did not correlate significantly. This indicates that distinct concepts of language were assessed by the tasks (Verb Naming:  $r_s = .71$ ,  $p = 0.182$ ; Verb Inflection:  $r_s = .73$ ,  $p = .165$ ; Constituent Assembly:  $r_s = -.25$ ,  $p = 0.685$ ; and Thematic Role Assignment:  $r_s = -.54$ ,  $p = .343$ ).

### Discriminant validity

Age and education levels of participants had normal distributions in each group included in the discriminant validity analysis. There was no significant difference for age or education level between the SPD

group ( $n = 9$ ; Age:  $M = 75.1$ ,  $SD = 7.1$ ; Education:  $M = 10.1$ ,  $SD = 3.8$ ), and the control group with no deficit ( $n = 9$ ; Age:  $M = 74.6$ ,  $SD = 3.1$ ; Education:  $M = 11.1$ ,  $SD = 3.0$ ) according to t-tests for independent samples, Age:  $t(16) = 0.214$ ,  $p = .833$ ; Education:  $t(16) = -0.623$ ,  $p = .542$ . The SPD group had significantly lower scores than participants with no syntactic deficit for the Verb Inflection, Constituent Assembly, and Thematic Role Assignment Tasks, and the total battery. The effects sizes were large for those four significant differences. Groups did not differ significantly on scores to the Verb Naming Task (Table 4).

### Test-retest reliability

Scores to the battery between first assessment ( $M = 57.1$ ,  $SD = 6.70$ ) and retest ( $M = 57.8$ ,  $SD = 5.45$ ) strongly resemble,  $ICC = .98$ ,  $F(59, 59) = 53.6$ ,  $p < .001$ . These results suggest that BEPS's test-retest reliability is excellent (Koo & Li, 2016).

### Internal consistency

Cronbach's  $\alpha$  coefficients were .74 for the Verb Naming Task (Task 1; 18 items) and .92 for the Verb Inflection Task (Task 2; 24 items). Tasks 1 and 2 had good internal consistency (Cortina, 1993). The Constituent Assembly Task (Task 3) and the Thematic Role Assignment Task (Task 4) had relatively low and moderate reliability with Cronbach's  $\alpha$  coefficients of .51 (12 items) and .66 (8 items). Because of the small number of items and the psycholinguistic variables manipulated for those two tasks (syntactic structures, and number of arguments and voice of verbs), the subscales implied high item-specific variance.

### Concurrent validity

Three participants from the five participants in the group with poststroke aphasia were identified as having SPD ( $n = 3$ ) via the image description task of the

**Table 4.** Wilcoxon rank-sum test examining scores to the BEPS in groups with and without syntactic production deficit.

Group Task (number of items)	SPD ( <i>n</i> = 9)		No deficit ( <i>n</i> = 9)		<i>W<sub>S</sub></i>	<i>p</i> value	<i>r</i>
	<i>M</i>	<i>Mdn</i>	<i>M</i>	<i>Mdn</i>			
1: Verb Naming (18) <sup>a</sup>	14.5	15.0	16.3	17.0	16.0	.185	−0.42
2: Verb Inflection (24) <sup>a</sup>	17.0	18.0	24.0	24.0	10.0	.004*	−0.91
3: Constituent Assembly (12)	6.67	7.00	11.2	10.7	47.0	.001*	−0.82
4: Thematic Role Assignment (8)	4.67	4.00	7.78	8.00	51.5	.002*	−0.74
BEPS (62) <sup>a</sup>	44.0	43.0	59.3	60.0	50.0	.010*	−0.82

Note. BEPS: Batterie d'Évaluation de la Production Syntaxique; *W<sub>S</sub>*: Wilcoxon W statistic for independent samples; *r*: Rosenthal's approximate effect size for nonparametric statistics.

<sup>a</sup>The SPD group had 6 participants and the control without deficit had 4 participants for this statistical analysis.

\**p* < .05.

MT-86 (Nespoulous et al., 1992), used as a criterion measure. Two poststroke participants had normal sentence production (*n* = 2) in this task. Using as reference the BEPS, four IWAs had SPD, and one had not.

The BEPS correctly identified the three participants who have SPD. Their total scores to the battery were 40, 45, and 50 out of 62. The BEPS identified one participant out of two with normal performance in sentences production. Their total scores were 52 and 58 out of 62. This supposes that three of the four positive results for SPD were correct (Loong, 2003). However, the number of cases presented is not enough to draw sound conclusions about the concurrent validity of the BEPS.

One participant with poststroke aphasia (74 years of age and 18 years of education) was identified with SPD by the BEPS, but not by the standard task of the MT-86, which consists of an oral picture description task. The performances of this participant were further analyzed. The participant showed weak scores in the Verb Naming Task (14/18; 2d percentile), the Verb Inflection Task (22/24; 5th percentile), and the Thematic Role Assignment Task (6/8; 1–5 percentiles). Even though the sentence production through the oral description was syntactically correct, the participant exhibited word-finding difficulties (superordinate verbal paraphasias, vague pronoun reference, exclusive use of high-frequency verbs carrying few grammatical roles, and verbalizations of his difficulties to retrieve the word). The picture description corpus consists of six sentences: two active sentences with the canonical (subject-verb-object) pattern, and four repetitive structures with a subordinate clause introduced by a relative pronoun (e.g., There's one who runs). In most sentences, verbs were in the present tense and the third person singular. In the inflection task, his mistakes were on items targeting the past and future tenses and the inflection in the third person plural. Because the MT-86 task cannot impose a specific syntactic structure, his difficulties with noncanonical sentences have not been identified with this test, although he failed to produce two out of four passive

sentences in the Thematic Role Assignment Task (BEPS). Moreover, he was not able to complete two passive sentences in the Constituent Assembly Task, but his score remained at the 10th percentile (within the normal range).

### Summary

Data shows that the BEPS has good construct validity. Sentence production can be assessed with the BEPS with good accuracy. The subtest scores can differentiate between patients with and without SPD in post-stroke and degenerative aphasia, in PPAOS, and the healthy-adult population. The BEPS can help clinicians in the differential diagnosis for agrammatic PPA and PPAOS. Each task can be used independently, but the entire battery may identify specific deficits and preserved functions in syntactic production stages described by Bock and Levelt (1994). As demonstrated with the good reliability of the battery, the BEPS is stable over time and has good internal consistency. The number of cases is not sufficient to draw a sound conclusion about the criterion validity of the BEPS. However, the case described in this study seemed to be correctly identified by the battery as presenting sentence production difficulties, but not by the image description task of the MT-86 (Nespoulous et al., 1992).

### Study III. Normative data establishment

The purpose of Study III was to provide normative data for the assessment of acquired syntactic deficits, adapted to clinical adult and aged populations from two French-speaking countries (Western Switzerland and Quebec, Canada).

### Materials and methods

#### Participants

From December 2015 to April 2017, 196 French-speaking adults were recruited for this study by research assistants through public advertisements.

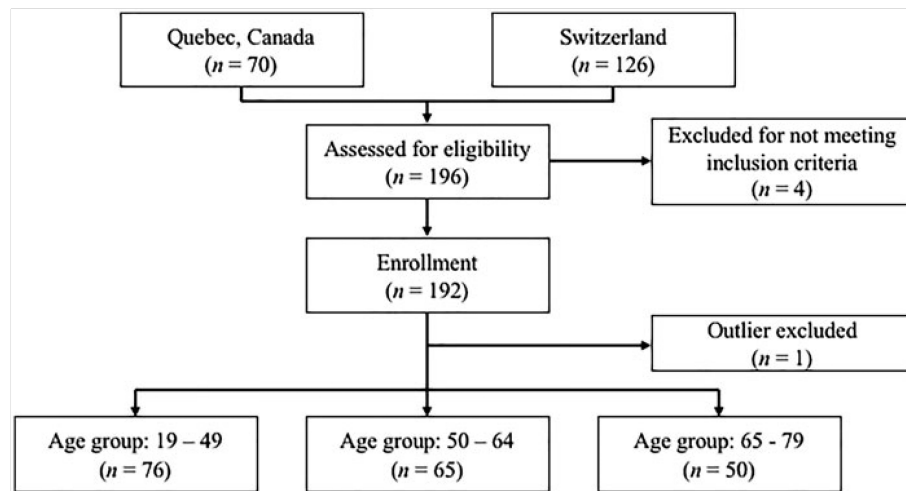
**Table 5.** Participants' characteristics and scores (mean and standard deviation) of the normative sample.

	Canada ( <i>n</i> = 66 <sup>a</sup> )		Switzerland ( <i>n</i> = 125)		<i>t</i>	df	<i>p</i> value
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Characteristic							
Age (years)	47.2	18.1	51.1	17.2	1.47	189	.143
Education Level (years)	14.5	2.77	14.4	3.65	−0.257 <sup>b</sup>	168.4	.798
Measure							
1: Verb Naming	17.4	.922	17.1	1.15	−1.62	189	.169
2: Verb Inflection	23.7	1.04	23.8	.558	0.954	189	.341
3: Constituent Assembly	11.7	.581	11.7	.781	−0.228	189	.820
4: Thematic Role Assignment	7.67	.664	7.76	.588	−0.998	189	.320
BEPS: Syntactic Production	60.2	2.32	60.3	1.76	0.272	189	.803

Note. No significant difference between groups ( $p > .05$ ). BEPS: Batterie d'Évaluation de la Production Syntaxique.

<sup>a</sup>One participant was removed from the analyses because he exhibited extreme performance and, thus, he was not included in the normative sample.

<sup>b</sup>Unequal variances across groups ( $p = .028$ ), thus Welch's *t*-Test was used.

**Figure 2.** Sampling procedures for Study III.

Eligibility criteria for participants included in the normative sample were (a) French as a mother tongue and daily used language, and (b) aged between 19 and 79. Exclusion criteria were (a) medical history of brain, head, or neck cancer; (b) history or current neurological or psychiatric illness self-reported from participants, and cognitive impairment according to the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005); (c) significant uncorrected vision or audition problems; and (d) current or past follow-up (less than 3 years) in speech-language pathology. This information regarding exclusion criteria was obtained from participants' self-reports.

Four participants scored below the normal range of the MoCA and were rejected from the study. The sample was composed of 192 healthy participants coming from two country groups (Canada:  $n = 67$ ; Switzerland:  $n = 125$ ). All included participants had normal education-adjusted MoCA scores ( $\text{MoCA} \geq 26$ ;  $M = 27.9$ ;  $SD = 1.45$ ), indicating absence of cognitive impairment (Nasreddine et al., 2005). The groups (Canada and Switzerland) did not significantly differ in age (aged between 19 and 79) or education

level (varying between 7 and 27 years). Thus, all participants constitute one sample ( $n = 192$ ) without country distinction (age:  $M = 49.5$ ;  $SD = 17.5$ ; education level:  $M = 14.5$ ;  $SD = 3.33$ ). Participants' characteristics and scores are summarized in Table 5, and the sampling procedure is shown in Figure 2.

### Procedure

All participants provided informed consent. Participants were tested individually in a quiet room at their home or at the research center. The MoCA (Nasreddine et al., 2005) was administered immediately before performing the four tasks of the BEPS. The BEPS's stimuli were presented on a laptop or tablet computer. There was no time limit to complete the tests. Research assistants collected written protocols. The participants received no financial compensation.

### Statistical analysis

SPSS v. 25 (SPSS, Chicago, IL) was used for statistical analyses. The alpha risk was set to 5% ( $p = .05$ ). All data were examined for normality, skewness and extreme values. All scores of the BEPS were found to

have non-normal distributions (Kolmogorov-Smirnov Tests,  $p < .05$ ). The inspection of the histograms showed negatively skewed distributions. In order to carry out statistical analysis, the data from participants with a total BEPS score less than or equal to 3.0 inter-quartile range (IQR) below the 25th percentile were considered outliers and removed. Data of one participant had to be removed from the statistical analyses because he showed an extreme value for the total score on the BEPS (6.0 IQR under the median of the group for the total score). This was the only outlier according to the boxplots and the histograms. The descriptive and statistical analyses were then carried out with a normative sample of 191 participants.

**Effects of sociodemographic variables.** To ensure that participants in Canada and Switzerland had similar performances on the BEPS, t-tests were carried out. Partial Spearman's correlation coefficients were calculated to identify the relationships between age (years), education level (years) and scores from the tasks. Partial correlations were used to describe the relationship between each sociodemographic variable (age and education level) and scores while controlling for the effects of the other sociodemographic variable on this relationship.

**Effects of psycholinguistic variables.** The interactions between significant effects of sociodemographic variables and effects of psycholinguistic variables on performances to the BEPS's tasks were investigated with robust two-way mixed ANOVAs (Mair & Wilcox, 2015; Wilcox, 2017). The within-subject factors were psycholinguistic variables (see Table 1), and the between-subjects factors were the age or education groups (19 to 49; 50 to 64; 65 to 79 years of age and 0 to 11; 12 and more years of education). These analyses compared the effects of psycholinguistic variables between sociodemographic groups.

**Normative data.** The 1st, 2nd, 5th, 15th, 25th, 50th, and 95th stratified percentiles were calculated for each task according to significant sociodemographic variables.

## Results

### Effects of sociodemographic variables

The groups (Canada and Switzerland) did not significantly differ in performance to the BEPS (Table 5).

No stratification was necessary for Naming of Verbs (Task 1) and Verb Inflection (Task 2) because

age and education were not significant correlates to these tasks, indicating that they had very little influence on performance. Partial Spearman's correlations showed statistically significant correlations between age and the Constituent Assembly Task (Task 3, Partial  $r_s(188) = -.24$ ,  $p = .001$ ); age and the Thematic Role Assignment Task (Task 4, Partial  $r_s(188) = -.16$ ,  $p = .029$ ); age and total score (Partial  $r_s(188) = -.24$ ,  $p = .001$ ), while controlling for education level. These results suggest that performance deteriorates with age for the Constituent Assembly and Thematic Role Assignment Tasks. Education level was not significantly correlated with any of the tasks. Table 6 shows the effects of sociodemographic variables on the BEPS's scores.

### Effects of psycholinguistic variables

The results of the robust ANOVAs confirmed the significant main effects of age on constituent assembly (Task 3:  $F_t(2, 75.40) = 8.55$ ,  $p < .001$ ) and on function assignment (Task 4:  $F_t(2, 55.49) = 4.33$ ,  $p < .018$ ). There were also significant main effects of the syntactic structure ( $F_t(2, 77.71) = 5.39$ ,  $p < .001$ ) and the verb's voice on performances ( $F_t(1, 78.64) = 8.33$ ,  $p = .005$ ). The number of arguments was not a significant factor for the performances in the function assignment task,  $F_t(1, 79.64) = 1.80$ ,  $p = .183$ . The interaction effects were significant between the age group and the syntactic structures in the constituent assembly Task ( $F_t(4, 73.36) = 5.39$ ,  $p < .001$ ) and between the age group and the voice of verbs in the function assignment Task ( $F_t(2, 55.49) = 4.33$ ,  $p = .018$ ) were both significant.

In the constituent assembly (Task 3), the elderly group (65–79 years of age) had lower scores in the passive sentences,  $M = 3.31$  [3.04, 3.56] compared to the 19–49 group,  $M = 3.77$  [3.63, 3.90] and 50–64 group,  $M = 3.72$  [3.57, 3.88]. Participants in the 65–79 age group also had lower scores,  $M = 2.71$  [2.33, 3.10] than the 19–49 group,  $M = 3.49$  [3.27, 3.72] in the cleft syntactic structure.

In the function assignment (Task 4), the passive voice of verbs seemed to have a more deleterious effect on the performances of the 65–79 group than the other groups. Their mean scores went from 3.98 [3.94, 4.00] in the active sentences to 3.55 [3.36, 3.75] in the passive sentences. As for the other groups, their mean scores were 3.95 [3.90, 4.00] for the 19–49 group and 3.91 [3.84, 3.98] for the 50–64 group for the active sentences and 3.81 [3.71, 3.90] and 3.61 [3.44, 3.79] for the passive sentences.



**Table 6.** Effects of age and education level on BEPS scores.

Measure	Age		Education level	
	Partial $r_s$	$p$ value	Partial $r_s$	$p$ value
1: Verb Naming	-.138	.057	.005	.948
2: Verb Inflection	-.072	.323	-.080	.273
3: Constituent Assembly	-.240	.001*	.116	.112
4: Thematic Role Assignment	-.159	.029*	.074	.308
BEPS: Syntax Production	-.236	.001*	.081	.266

Note. BEPS: Batterie d'Évaluation de la Production Syntaxique; Partial  $r_s$ : Partial Spearman's correlation coefficients.

\*  $p < .05$ .

### Normative data

Table 3 shows normative data (means, standard deviations, and percentiles). The normative sample was stratified according to the significant sociodemographic variable, with a current distribution across categories in tests for French-speaking populations (19 to 49; 50 to 64; 65 to 79 years of age). The difference between group means were statistically significant [Task 3: Welch's  $F(2, 95.61) = 6.11$ ,  $p = .003$ ; Task 4: Welch's  $F(2, 100.76) = 3.35$ ,  $p = .039$ ; total score: Welch's  $F(2, 106.12) = 8.28$ ,  $p < .001$ ].

After visual exploration and according to the usual criteria (Crawford & Garthwaite, 2009), the 5th percentile, which approximately corresponds to one and a half standard deviations below the mean, was chosen as the cutoff score for all tasks.

### Summary

Study III provides normative data for a new clinical battery for the assessment of acquired syntactic production deficits in French adults and the elderly (19–79). These norms were established with populations from two French-speaking countries (Western Switzerland and Quebec, Canada).

### Discussion

These studies presented the development, the validation and the normative data of the BEPS, a new French-language assessment tool allowing clinicians to detect and target sentence production deficits in IWAs with post-stroke or neurodegenerative lesions.

The development of a sentence production test in French was necessary for clinicians to evaluate SPD with a better specificity regarding the source of the deficit. This was possible due to the sequential model of the normal sentence production processes from Bock and Levelt (1994). The BEPS evaluates SPDs and targets the source of the deficits with four tasks based on Bock and Levelt (1994) Model of sentence production, namely (1) Verb Naming Task; (2) Verb Inflection Task; (3) Constituent Assembly Task; and

(4) Thematic Role Assignment Task. The BEPS had good face validity, content validity, construct validity, and reliability. This assessment battery fulfills the current needs for clinical tool in the francophone population when assessing sentence production capacities. Therefore, this new assessment tool can help clinicians to identify syntactic production deficits in IWAs with post-stroke or degenerative lesions. Normative data are presented for individuals 19–79 years of age with percentile scores for each task. The BEPS is freely available on Research Gate [https://www.researchgate.net/publication/329590702\\_Batterie\\_d'Evaluation\\_de\\_la\\_Production\\_Syntaxique\\_BEPS\\_-\\_Monetta\\_Perron\\_Coulombe\\_et\\_Fossard\\_2018](https://www.researchgate.net/publication/329590702_Batterie_d'Evaluation_de_la_Production_Syntaxique_BEPS_-_Monetta_Perron_Coulombe_et_Fossard_2018)

The validation study showed that the total score on the BEPS and also the subtest scores are sufficiently accurate to distinguish between aphasic and control groups. Scores to the subtests are relevant to identify the nature of the SPD, notably in consideration to the thematic role processing, inflection and grammatical ordering abilities. Because SPD in IWAs can manifest in a wide range of symptoms, the BEPS tasks capture the sources of the deficit and give information about the underlying mechanisms of sentence production.

Results of the normative data showed effects of age on Constituent Assembly (3) and Thematic Role Assignment (4) Tasks. It suggests that performances on the Constituent Assembly (3) and Thematic Role Assignment (4) Tasks deteriorate with age. There were also significant interaction effects between the age group and the syntactic structures. These effects indicated that the older group had weaker performances in sentences where they had to apply several rules than in sentences that could be resolved with no syntactic movement rules.

These findings could be related to the decline in executive functions and working memory associated with aging (Sander, Lindenberger, & Werkle-Bergner, 2012). It has been suggested that sentences that require thematic roles to be assigned in an atypical order mandate more syntactical operations and then, more executive resources (Burchert, Meissner, & De Bleser, 2008). This hypothesis has also been proposed in sentence production of participants with traumatic brain injury. In the previous work of Peach (2013), the sentence planning impairments were associated with measures of working memory and executive function. Thus, the two tasks of the BEPS (3 and 4) that need participants to manipulate the sentence's constituents seemed to be more demanding on executive functions and working memory and that could

explain the observed decline in performance with advancing age.

Another suggestion is that difficulties in the construction of grammatical structures could result from a deficit in the procedural memory, which is not the same process as the declarative memory system that underlies verb naming and verb inflection (Macoir, Fossard, Nespoulous, Demonet, & Bachoud-Levi, 2010). Our results are also in agreement with dissociation between these cognitive functions. Effects of age were observed in the Constituent Assembly and Thematic Role Assignment Tasks, but not in the Verb Naming and Verb Inflection Tasks. According to Macoir et al. (2010), a procedural deficit in combining several syntactic rules is evidenced by the effects of the syntactic structure, in the absence of a length effect in sentence productions tasks. In our study, the older group was less successful than younger groups on the execution of syntactic procedures. However, the number of arguments, which is associated with the length of sentences, had no significant effect on performances.

The large group of participants in the normative study and the desire to represent the French-speaking population of Canada and Switzerland are major strengths of this study. One limitation of the study is that norms are presented for persons of 19 to 79 years old; further studies are needed to extend the normative data to people aged 80 and above. Another limitation of the present study is the nonrandom sample that could bring sampling bias in the results. The low number of IWAs with SPD in the validation study is a significant shortcoming, particularly concerning the measures of concurrent validity.

Unfortunately, the normative data of the BEPS cannot be used with IWAs presenting dyslexia or visual agnosia. In contrast, the anagram task (Constituent Assembly Task) allows assessing sentence production without the impact of motor speech deficits.

## Conclusion

To conclude, the BEPS is a sentence production assessment battery that can help to identify syntactic difficulties experienced by aphasic patients to enable targeted intervention on the origin of the SPD. The development, the validation and the normative data have been presented in the present study. The BEPS can help speech-language pathologists to identify SPDs in patients with poststroke and degenerative disorders. This test battery fulfills the current needs for

clinical tool in the francophone population when assessing sentence production capacities.

## Acknowledgments

We wish to express sincere appreciation to the practicing clinicians who have contributed to the validation process of the BEPS.

## Disclosure statement

The authors have no conflict of interest to declare.

## Funding

This research did not receive any specific research funding.

## ORCID

Valérie Coulombe <http://orcid.org/0000-0002-4660-4876>

## References

- Bastiaanse, R., Edwards, S., Mass, E., & Rispens, J. E. (2003). Assessing comprehension and production of verbs and sentences: The Verb and Sentence Test (VAST). *Aphasiology*, 17(1), 49–73. doi:10.1080/729254890
- Bastiaanse, R., Edwards, S., & Rispens, J. E. (2002). *The verb and sentence test (VAST)*. Bury St. Edmunds, United Kingdom: Thames Valley Test Company.
- Bernaert-Paul, B., & Simonin, M. (2011). *Test d'expression morpho-syntaxique fine T.E.M.F.* Marseille, France: Solal Editeurs.
- Bock, K., & Levelt, W. J. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.
- Burchert, F., Meissner, N., & De Bleser, R. (2008). Production of non-canonical sentences in agrammatic aphasia: Limits in representation or rule application? *Brain and Language*, 104(2), 170–179. doi:10.1016/j.bandl.2007.06.004
- Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, 3(4), 572–582. doi:10.1016/0093-934X(76)90048-1
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104. doi:10.1037/0021-9010.78.1.98
- Crawford, J. R., & Garthwaite, P. H. (2009). Percentiles please: The case for expressing neuropsychological test scores and accompanying confidence limits as percentile ranks. *Archives of Clinical Neuropsychology*, 23(2), 193–204. doi:10.1080/13854040801968450
- Duffy, J. R. (2006). Apraxia of speech in degenerative neurologic disease. *Aphasiology*, 20(6), 511–527. doi:10.1080/02687030600597358

- Hammerlath, C. (2001). *Test de dénomination de verbes lexicaux en images (DVL 38)*. Isbergues, France: Ortho Édition.
- Kemmerer, D. (2014). Sentence production. In D. Kemmerer (Ed.), *Cognitive neuroscience of language* (pp. 391–417). New York, NY: Psychology Press.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. doi:10.1016/j.jcm.2016.02.012
- Loong, T. W. (2003). Understanding sensitivity and specificity with the right side of the brain. *British Medical Journal*, 327(7417), 716–719. doi:10.1136/bmj.327.7417.716
- Macoir, J., Fossard, M., Nespoulous, J. L., Demonet, J. F., & Bachoud-Levi, A. C. (2010). The application of rules in morphology, syntax and number processing: A case of selective deficit of procedural or executive mechanisms? *Neurocase*, 16(4), 358–376. doi:10.1080/13554791003620272
- Macoir, J., Jean, C., & Gauthier, C. (2015). *Batterie d'évaluation cognitive du langage (BECLA)*. Quebec, Canada: Laval University.
- Mair, P., & Wilcox, R. (2019). WRS2: A collection of robust statistical methods. R package version 1.0-0. Retrieved from <https://r-forge.r-project.org/projects/psychor/>
- Mehri, A., & Jalaie, S. (2014). A systematic review on methods of evaluate sentence production deficits in agrammatic aphasia patients: Validity and reliability issues. *Journal of Research in Medical Sciences*, 19(9), 885–898. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4268199/#idm139747349493104title>
- Mitchum, C. C., & Berndt, R. S. (2008). Comprehension and production of sentences. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders* (pp. 632–653). Baltimore, MD: Lippincott Williams & Wilkins.
- Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53, 695–699. doi:10.1111/j.1532-5415.2005.53221.x
- Nespoulous, J. L., Lecours, A. R., Lafond, D., Lemay, M., Puel, M., Joannette, Y. ... Rascol, A. (1992). *Protocole Montréal-Toulouse d'examen linguistique de l'aphasie MT-86 module standard initial, M1 β*. Isbergues, France: Ortho Édition.
- Peach, R. K. (2013). The cognitive basis for sentence planning difficulties in discourse after traumatic brain injury. *American Journal of Speech-Language Pathology*, 22, 285–297. doi:10.1044/1058-0360(2013/12-0081)
- Pillon, A. (2014). L'évaluation des troubles de la production et de la compréhension des phrases. In X. Seron & Van der Linden (Eds.), *Traité de neuropsychologie clinique de l'adulte. Tome 1 - Evaluation*. (pp. 321–330). Paris, France: Solal Editeurs.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York, NY: Russell Sage Foundation.
- Saffran, E. M., Schwartz, M. F., & Marin, O. S. (1980). The word order problem in agrammatism. II. *Production of Brain and Language*, 10(2), 263–280. doi:10.1016/00934X(80)90056-5
- Sander, M. C., Lindenberger, U., & Werkle-Bergner, M. (2012). Lifespan age differences in working memory within a two-component framework. *Neuroscience & Biobehavioral Reviews*, 36(9), 2007–2033. doi:10.1016/j.neubiorev.2012.06.004
- Verstichel, P., & Cambier, J. (2005). Les aphasies. In Botez-Marquard & F. Boller (Eds.), *Neuropsychologie clinique et neurologie du comportement* (3rd ed., pp. 471–500). Montréal, Canada: Les Presses Universitaires de Montréal.
- Wilcox, R. (2017). *Introduction to robust estimation and hypothesis testing* (4th ed.). Burlington, MA: Elsevier.
- Wilshire, C. E., Lukkien, C. C., & Burmester, B. R. (2014). The sentence production test for aphasia. *Aphasia*, 28(6), 658–691. doi:10.1080/02687038.2014.893555