



**HAL**  
open science

## Norms of vocabulary, reading, and spelling tests in French university students

Emilie Dujardin, Gaël Jobard, Théodora Vahine, Stéphanie Mathey

► **To cite this version:**

Emilie Dujardin, Gaël Jobard, Théodora Vahine, Stéphanie Mathey. Norms of vocabulary, reading, and spelling tests in French university students. *Behavior Research Methods*, 2021, 54 (4), pp.1611-1625. 10.3758/s13428-021-01684-5 . hal-03901816

**HAL Id: hal-03901816**

**<https://hal.science/hal-03901816>**

Submitted on 10 Nov 2023

**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.



# Norms of vocabulary, reading, and spelling tests in French university students

Emilie Dujardin<sup>1,2</sup> · Gaël Jobard<sup>3</sup> · Théodora Vahine<sup>1</sup> · Stéphanie Mathey<sup>1</sup>

Accepted: 9 August 2021  
© The Psychonomic Society, Inc. 2021

## Abstract

The purpose of the present paper is to provide the norms of widely used tests of vocabulary, reading, and spelling skills for a French population of young adults. Data were collected from 18- to 26-year-old university students during individual sessions. We assessed 771 participants with the Mill Hill part B vocabulary test, 410 with the LexTale-Fr test, 1231 with the *Alouette-R* test, and 361 with the *Pollueur* word/pseudoword dictation and text dictation. Stepwise regression analyses showed the need to stratify the reference population according to the level of education and gender for some tests. The *Alouette-R* and Mill Hill scores increased with educational level. Moreover, for the tests whose performance differed according to gender, women generally performed better than men. The present normative data concerning vocabulary, reading, and spelling skills should provide useful tools for researchers and practitioners alike to rate young individuals within their reference population.

**Keywords** Young adults · Vocabulary · Reading · Spelling · Lexical skills

## Introduction

While it is well acknowledged that language proficiency is important in society, whether to communicate with or to understand individuals, the issue of how to assess language skills and how they might influence cognitive functioning is still under investigation. Recently, Andrews (2012, 2015) pointed to the influence of individual differences in lexical skills (e.g., vocabulary, reading, and spelling skills) in English adult speakers in cognitive psychology research. She argued that most of the studies on skilled readers have been conducted on samples ranging from 20 to 30 university students, relying on the implicit assumption that they all process words in the same way, while only a few studies took into account the role of

differences in lexical skills amongst the student population in the effects under investigation. In France, according to the National Institute of Statistics and Economic Studies (INSEE, 2018), the student population is very large ( $N = 2,680,400$ ), and the proportion of young adults with a *baccalaureate* (French high school diploma) has also increased considerably in the last two decades (62.6% in 1998; 78.7% in 2017). Moreover, individual differences in reading, spelling, and vocabulary skills are important as they may be at the root of the disparities in outcomes observed in psychological studies and clinical practice in young adults. Researchers and practitioners need normed tests to investigate or control the possible influence of vocabulary, reading, and spelling levels and characterize and assess individuals' normal and abnormal functioning (i.e., memory, language, executive functions). We therefore believe it would be useful to propose an updated norm of vocabulary, reading, and writing tests adapted to a sample of young French-speaking adults at university. In this article, we selected available tests that are easy to administer (i.e., paper-and-pencil tests, short duration) and are commonly used to assess French vocabulary, reading, and spelling skills. We provide norms that are not readily available for a large population of young post-baccalaureate individuals.

---

✉ Emilie Dujardin  
emilie.dujardin@univ-lyon2.fr

<sup>1</sup> University Bordeaux, EA 4139 Labpsy, France

<sup>2</sup> Laboratoire EMC, Université Lyon 2, Bron, France

<sup>3</sup> Groupe d'Imagerie Neurofonctionnelle, CEA, University Bordeaux, CNRS, IMN, UMR 5293, Labpsy, France

## Vocabulary skills

Vocabulary skills can be examined accordingly to two complementary dimensions, the number of lexical representations stored by a given individual (i.e., the size or breadth of vocabulary) and the extent and accuracy of the semantic knowledge stored by an individual in relation to a specific word (i.e., the depth of vocabulary; see e.g., Ouellette, 2006). More precisely, vocabulary breadth, also referred to as the mental lexicon size, corresponds to all the words known by a given individual. Vocabulary breadth increases across one's lifespan (see Ben-David, Erel, Goy & Schneider, 2015, for a recent study of English-speaking adults), depending on print exposure (i.e., time spent reading). The increase in vocabulary size implies greater proficiency at discriminating and identifying words (e.g., Cohen-Shikora & Balota, 2016; Davies, Arnell, Birchenough, Grimmond & Houlson, 2017). Indeed, the more words individuals know, the more efficient the lexical identification process needs to be (Perfetti, 2007). To measure the breadth of knowledge, word recognition tasks are widely used. These tasks do not require knowledge of the definition of the words, but knowledge of the spelling form of the words stored in the lexicon (Ouellette, 2006).

Beyond vocabulary size, the depth of vocabulary knowledge can also be considered to assess vocabulary skills. Quality of word knowledge depends on the ability to link a word with a multitude of other words (i.e., vocabulary breadth) in a coherent semantic organization (i.e., vocabulary depth), such as synonyms or antonyms (e.g., Schwartz & Katzir, 2012). In psychological assessment, the choice of the vocabulary test depends on the dimension(s) under consideration (i.e., breadth, depth). To measure depth of vocabulary knowledge, using production tests involving giving the definition of a word would be most effective (see Mill Hill, Part A, Deltour, 1998). However, this makes scoring difficult and time consuming. A possible and widely used alternative is to use a synonym task in a receptive format (e.g., see Mill Hill, Part B, Deltour, 1998), involving more semantic and definition knowledge (i.e., depth) than breadth knowledge about words (Schwartz & Katzir, 2012). By using such a synonym test in English (Shiple, 1940), variation in young adults' vocabulary levels in visual word recognition performance was shown (Yap, Balota, Sibley, & Ratcliff, 2012). Individuals with high vocabulary skills were more accurate and efficient in visual word recognition tasks (e.g., pronunciation, lexical decisions, and semantic classification) than those with low vocabulary skills.

In the French language, Part B of the Mill Hill test (Deltour, 1998, a French adaptation of Raven, 1965) is frequently used to test vocabulary in cognitive research on young adults (e.g., Nelis, Quoidbach, Hansenne, & Mikolajczak, 2011; see also Dujardin & Mathey, 2019 for use in a composite score of lexical skills). Normative data on children and adolescents

have also been provided (Vigneau, 2007). In the field of cognitive aging, the Mill Hill part B is often used to take into account and/or to control the role of vocabulary in age-related effects in studies comparing memory or language performance of young and older adults (see e.g., Bertrand, Moulin, & Souchay, 2017; Dupart, Auzou, & Mathey, 2018; Robert & Mathey, 2007). In this multiple-choice test, participants have to select the synonym of each target word among six possible choices. Although reading and spelling skills as well as vocabulary size are implied in this test, meaning knowledge is involved to a greater extent in choosing the synonym of the stimulus word among the proposed set of words. The Mill Hill Part B test has 44 items arranged in an increasing difficulty order. The first 10 items are very easy and are only presented in the junior assessment (for 11–14-year-olds; see also Vigneau, 2007, for data on 9–11-year-olds), while the corresponding points are automatically attributed (without presenting these items) in the senior assessment (over 14 years). The Mill Hill test was initially normed on a population of 2104 individuals aged from 20 to 89 years old, taking into account age, sex, profession, and education (Deltour, 1998). The first age category corresponded to young adults aged from 20 to 29 years old ( $n = 291$ ). Among the first age category, only 59 individuals had a university degree, which is very few given the current increase in the number of individuals studying at university. As previously noted, the number of French students with a baccalaureate has increased substantially since 1998 (INSEE, 2018). Updating the Mill Hill norms for a university student population could therefore be useful for current research and clinical practice.

Another test of vocabulary that is becoming more widespread in cognitive psychology in several languages is LexTale (proposed by Lemhöfer and Broersma, 2012 in English; see Brysbaert, 2013 for the French version; Izura, Cuetos, & Brysbaert, 2014, for the Spanish version). The test consists of deciding whether visually presented stimuli correspond to real words (among a set of words and pseudowords). In the French version (Brysbaert, 2013), 84 items are presented in columns with a ratio of one in two words/pseudowords. Thus, the written stimuli need to be compared to the orthographic representations of known words stored in the mental lexicon. These orthographic representations must also be accurate enough to avoid the incorrect recognition of pseudowords sounding like real words (i.e., pseudo-homophones). For example, to differentiate the stimulus pseudo-homophone “*agire*” from the real word “*agir*”, the individual must know the spelling of the word. LexTale was initially proposed to measure the breadth of vocabulary in the second language of bilingual individuals. For instance, in a study of multilingual Switzerland adults, Willemin et al. (2016) examined the extent to which gender, manual preference, multilingualism, and vocabulary level could influence hemispheric lateralization. A lateralized lexical decision task (right vs. left

visual field presentation) was presented to participants in three languages of Switzerland (French, German, and Italian), and also in English and Dutch. All participants were assessed with the French version of LexTale. The results showed no influence of gender, manual preference, or multilingualism, but an advantage for presentation in the right rather than the left visual field. Furthermore, in early bilinguals, a higher LexTale-FR score was related to better performance on the lexical decision task when words were presented in the left visual field, whereas in late bilinguals, higher LexTale-FR scores were related to better performance when words were presented in the right visual field. Although LexTale-FR was initially used to assess vocabulary in multilingual individuals, Brysbaert (2013) noted the potential to use this test to measure the vocabulary level of monolingual individuals and to quantify individual differences in lexical skills (see e.g., Dujardin & Mathey, 2019), as well as in the evaluation of French-speaking patients by practitioners. Although it appears to be a useful tool for the French language, LexTale-FR has never been normed.

## Reading skills

While skilled adult readers do not typically experience any particular difficulties and are able to read quickly and without error, variability nonetheless exists among skilled readers (e.g., Andrews, 2015). Studies of individual differences in reading have generally focused on children's reading development to identify factors contributing to reading literacy (e.g., Ecalle & Magnan, 2015, in French; Pollatsek & Treiman, 2015, in English). However, reading difficulties encountered during childhood may persist into adulthood (Perfetti, 2007). Moreover, in psychological studies, variations in results may also emerge from individual differences in reading efficiency. Practitioners also need to identify difficulties in reading in order to help and diagnose patients.

One of the most widely used French reading tests is *Alouette* (Lefavrais, 1967, 2005 for the revised version offering new indexes of efficiency). The *Alouette* test involves reading aloud a text with little meaning. The sentences are grammatically simple and make sense individually, but not in relation with each other. The text is unusual and highly sensitive to readers' difficulties. The text is surrounded by drawings chosen on purpose to detect potential reading difficulties. When an individual has reading difficulties, he or she can implement reading strategies by anticipating or making inferences (Lefavrais, 2005). If the reader has these strategies, he or she will use the drawings to try to make sense of the text in order to read it more easily, which will lead to mispronunciations of words. In addition, some rare words are present in the text such as "hirondeau" [baby swallow] instead of "hirondelle" [swallow] which is more frequent. The beginning

of the word "hirondeau" may activate the word "hirondelle" and lead to a pronunciation error (Lefavrais, 2005). The *Alouette* test is generally used to detect dyslexia in children and adolescents (e.g., Ecalle & Magnan, 2008; Maionchi-Pino, Magnan & Ecalle, 2010) and in adults (e.g., Cavalli et al., 2017, for screening for dyslexia in university students). However, researchers have also used it to evaluate reading levels in children (e.g., Chetail & Mathey, 2012) and in adults (e.g., Gola-Asmussen, Lequette, Pouget, Rouyer & Zorman, 2011; Siéroff & Haehnel-Benoliel, 2015). The *Alouette* test was normed by Gola-Asmussen et al. (2011) on a population essentially composed of French schoolchildren (272), with only 14 university students. The norms were provided both for the number of words correctly read in 1 minute and the number of errors, but not for the new index proposed by Lefavrais (2005) that simultaneously takes speed and accuracy into account. Recently, among a population of young adults at university, including 164 typical readers and 83 dyslexic participants, Cavalli et al. (2017) proposed cut-offs to detect dyslexia using three indexes: accuracy, reading time, and a combined speed-accuracy score, but they did not provide norms for this test. Our study seeks to contribute to the norms of this reading test on a larger population composed of undergraduate students at university and including the new efficiency index (combined speed-accuracy score) and reading times.

Another test of reading aloud, this time based on a text conveying meaning, was devised by Gola-Asmussen et al. (2011) to evaluate reading skills with materials that correspond more closely to the kind of text usually encountered by readers. The *Pollueur* text, which is on the topic of pollution, does not present particular difficulties. It is used to assess reading fluency, reporting the efficiency of automated reading processes. It also includes some proper names, requiring the intervention of the phonological reading channel. It has recently been used in studies on dyslexia among young adult participants to characterize their reading performances (Bürki, Besana, Degiorgi, Gilbert & Alario, 2018; Mahé, Pont, Zesiger & Laganaro, 2018; Pattamadilok, Nelis & Kolinsky, 2014). The *Pollueur* test provides norms based on the same population as the *Alouette* test, comprising only 14 university students (Gola-Asmussen et al., 2011).

## Spelling skills

Writing and reading processes allow individuals to form, store, and access orthographic representations in their mental lexicon (see e.g., Andrews, 2015; McClung et al., 2012). Writing words implies knowing their pronunciation and, consequently, knowing the graphophonemic correspondence rules (e.g., Pollatsek & Treiman, 2015). Variations in orthographic knowledge (i.e., lexical spelling corresponding to how

to write words; grammatical spelling corresponding to morphology) could indicate a variation in the orthographic process (e.g., Andrews, 2015). Indeed, Burt and Tate (2002) showed that in a spelling task, written words with errors compared to words spelled correctly by young adult English speakers corresponded to words that were also recognized more slowly and less accurately in a lexical decision task. Furthermore, Andrews and colleagues (Andrews & Hersch, 2010; Andrews & Lo, 2012) showed that spelling skills could be a predictor of the orthographic neighborhood priming effect in English. In French, Pattamadilok et al. (2014) showed the importance of orthographic knowledge in an auditory lexical decision task in adult readers (Experiment 1). They reported an effect of letter-sound consistency (i.e., orthographic knowledge), with longer reaction times for inconsistent words (i.e., whose phonological rime had several possible spellings) than for consistent words (i.e., when phonological rime had only one possible spelling). Moreover, they showed that the size of this effect was correlated with actual language skills assessed through the subtests from ECLA-16+ (Gola-Asmussen et al., 2011): the better the actual language skills, the greater the effect of letter-sound consistency. This result suggests that orthographic code knowledge may modify speech processes only when individuals have reached a certain reading level (Pattamadilok et al., 2014).

Ensuring that individuals have good knowledge of orthography involves developing measures that account for spelling processing. The need for such measures in French adult speakers led Gola-Asmussen et al. (2011) to devise dictation tasks (i.e., words/pseudowords, and text) in ECLA-16+. Writing regular and irregular words and pseudowords leads to the assessment of both lexical and phonological processes. Writing an irregular word is more difficult than writing a regular word or a pseudoword (Gola-Asmussen et al., 2011). This is because an irregular word does not follow the phono-graphemic correspondence rules to be written. For example, the word /solanɛl/ is written "solennel" [solemn] (while a regular spelling would be "solanne"). Since pseudowords by definition do not exist, they are written through the phonological pathway (e.g., /ribyl/ is written "ribule"). Finally, regular words follow the phono-graphemic correspondence rules (/vi/ is written "vigne"). The text dictation partially involves using verbal memory and the attention span required to retain part of the sentences to be written. Moreover, in the latter dictation task, individuals need to retrieve the lexical spelling and the grammatical spelling in order to write the text correctly. Indeed, they must maintain in memory a part of the text in order to retrieve the lexical spelling (e.g., "souterrain" [underground]) and the grammatical spelling (e.g., "petits", plural form of [small]). For grammatical spelling, individuals must know the syntactic rules of writing. The more these rules are

automated, the easier it will be for the individuals to write. Thus, one must be attentive to the slowness of writing. All these dictations were reference-normed with the same sample of participants as the one for the *Pollueur* and the *Alouette* tests (Gola-Asmussen et al., 2011), in which only 14 university students were assessed.

## The present study

The aim of this research is to provide normative data of lexical skills tests currently used in French studies and clinical practice on young adult populations due to the need and demand of practitioners and researchers to be able to rate young adults' performance in a reference population. Vocabulary, reading, and spelling skills were assessed in samples of several hundred students per test. For vocabulary skills, we provide norms for the Mill Hill part B test (Deltour, 1998) and the LexTale-FR test (Brybaert, 2013) as two complementary vocabulary tests. The first one involves retrieving the exact and precise meaning of words (i.e., depth of vocabulary), and the latter involves retrieving the correct spelling of known words (i.e., breadth of vocabulary). For reading level, we standardized two complementary tests: *Alouette-R* (Lefavrais, 2005), a text with little meaning, and *Pollueur* (Gola-Asmussen et al., 2011), a more standard and meaningful text for readers with no particular difficulties. Finally, we present normative data for the dictation tasks from ECLA-16+ (Gola-Asmussen et al., 2011) to assess spelling skills (i.e., the phonological and lexical process of writing, lexical and syntactic spelling).

## General methods

### Participants

All the participants were volunteer native French speakers aged 18–26 years. They were recruited during their breaks on the Humanities and Social Sciences campus of Bordeaux University, and most of them were psychology students. They had an education level spanning from the baccalaureate (high school diploma in North America) to the second year of a master's degree (5 years of university studies in France). They all signed an informed consent form prior to their participation. All were tested individually in a quiet room in the presence of an experimenter. They took part in various reading experiments conducted in our laboratory from 2013 to 2018. All the participants had normal or corrected-to-normal vision, and we excluded participants reporting a

history of reading or oral language difficulty from the current sample. Tests of vocabulary, reading, and/or writing were used as part of the protocols of former experiments as a means to control or investigate interindividual characteristics (e.g., Dujardin & Mathey, 2019). More specifically, each participant took all or some of the vocabulary skills tests (Mill Hill part B, LextTale-Fr vocabulary tests), reading skills (*Alouette R*, *Pollueur* tests), or spelling skills (word, pseudoword, and text dictations). The number of participants per test ranged from 361 to 1231, depending on the test. The main characteristics of the participants for each test are summarized in Table 1 (see Appendix for the correlation coefficient matrix of each test).

## Materials and procedures

Six paper-and-pencil tests were used to assess vocabulary, reading, and spelling. Four of these tests are freely available and the full instructions, items, and scoring for the French language can be found on the following websites. For the vocabulary, LexTale-FR is available on <http://crr.ugent.be/archives/921>. The ECLA-16+ battery (Gola-Asmussen et al., 2011), available at <http://www.cognisciences.com/accueil/outils/article/ecla-16>, provides the *Pollueur* test for assessing reading skills, and the two dictation tests for assessing the spelling skills. The two remaining tests are commercially available.

### Vocabulary skills

**Mill Hill test part B** Part B of the Mill Hill assessment consists of a multiple-choice test in which participants are invited to circle or underline the adequate synonym for each target word from among six proposals (Deltour, 1998). The 44 target words are ordered by increasing difficulty. The answers are

presented in columns, in which there is the synonym, three randomly chosen words, and two words that can be selected (phonologically close). We used the senior version of the task, designed for individuals aged over 14 years old (the first 10 words from the junior section are not presented and are replaced by 10 more difficult words at the end of the list). Therefore, 34 target words were presented, along with their six response proposals each, including one example for which the answer is already underlined. In the score computation, 1 point is given per correct response, including the example. The 10 points from the junior items are also automatically added in the senior evaluation, so the adults' final score corresponds to the number of correct responses out of 44 (Deltour, 1998). There is no time limit, but completion of this test took from 5 to 8 minutes.

**LexTale-FR** The LexTale-FR test is a free word recognition test (Brysbaert, 2013), in which participants are asked to select the words they know among several targets by checking the corresponding boxes on a paper sheet. Participants are informed beforehand that some stimuli are not words. Fifty-six words and 28 pseudowords are presented in a fixed random order, in three columns on one page. These targets are either words (more or less frequent) or word-like stimuli comprising existing morphemes (e.g., “*joueux*”, equivalent to “*playly*” in English). Participants are also invited to rate their proficiency in the French language with a score ranging from 1 to 10 (with 10 corresponding to perfect fluency in the French language). There is no time limit. Completion of this test took from 3 to 5 minutes. Two scores were computed to assess the participants' performance.

The “correct responses” score represents the proportion of items correctly classified as words or pseudowords, ranging from 0 (none correct) to 1 (all correct). The score is obtained with the following formula:

**Table 1** Number of participants, mean age, and standard deviations for each test

	<i>N</i>	Mean age	SD age	Mean years of post-lycée studies	SD years of post-lycée studies
Vocabulary skills					
Mill Hill test	771 (77)	20.52	2.18	2.09	1.21
LexTale-FR test	410 (85)	20.3	2.06	2	1.16
Reading skills					
<i>Alouette-R</i> test	1231 (78)	20.58	2.19	2.1	1.19
<i>Pollueur</i> test	361 (85)	20.29	2.05	2	1.17
Spelling skills					
Word/pseudoword dictations	361 (85)	20.29	2.05	2	1.17
Text dictation	361 (85)	20.29	2.05	2	1.17

*Note.* SD = standard deviation. *N* = number of participants (percentage of women in brackets).

### Correct responses

$$= \frac{(N_{\text{correctly selected words}} + 2 * N_{\text{correctly unselected pseudowords}})}{112}$$

This score is the most widely used and easily interpreted score (Brysbaert, 2013; Izura et al., 2014; Lemhöfer & Broersma, 2012). It corrects the score of error by taking into account the word/pseudoword ratio: If participants select more pseudowords than words, the score is below .05. The score, easy to interpret, can be transformed into a percentage, the process we applied in this study.

The second score is the  $d'$ , a signal detection measure of sensitivity. This score helps to determine the discrimination rate of participants: the higher the  $d'$ , the better the discrimination between words and pseudowords. It takes into account both the guesses and the personal response style (e.g., a bias toward yes or no answers, Lemhöfer & Broersma, 2012).

This score is calculated (see Stanislaw & Todorov, 1999) with the formula:

$$d' = Z(\text{hit rate}) - Z(\text{false rate})$$

The hit rate corresponds to the ratio between the number of correctly selected words and the maximum number of words (i.e., 56). False rate is the ratio between the number of incorrectly selected pseudowords and the maximum number of pseudowords (i.e., 28). The  $Z$ -values of hit and false alarm rates correspond to the distance from the mean (expressed in signed numbers of standard deviations) that lead to these probabilities in a Gaussian distribution (e.g., a hit rate of 95% corresponds to a  $Z$  score of +1.645).

### Reading skills

**Alouette-R test** The *Alouette-R* test (Lefavrais, 2005) involves reading aloud a text of 265 words. This text is composed of grammatically correct and simple sentences, with relatively easy to read words (pronunciation, and frequency) and rarer words. Although the sentences convey some meaning individually, they convey no clear meaning in relation with each other. Participants have to read the text surrounded by drawings on a sheet of paper and are asked not to touch the sheet to follow with their finger. They are also requested to read the text aloud as quickly and as accurately as possible within a maximum of 3 minutes. The drawings presented around the text could lead to contextual errors if relied upon during the reading process. For instance, the text features the drawing of a squirrel (“*écureuil*” in French) close to the written word “*écueil*” ([*reefs*]), which could lead to pronunciation errors. The experimenter notes the errors on the

scoring page (a text without drawings and with line numbers) during the 3-minute reading time. The number of correctly read words is transformed into a percentage corresponding to the precision index. The reading times and the number of correctly read words are combined to calculate the index of efficiency (called CTL; Lefavrais, 2005). This index corresponds to the number of words that participants can read correctly in 3 minutes.

### Index of efficiency

$$= \frac{\text{Number of correctly read words} \times 180}{\text{Reading times(s)}}$$

We focused our attention on this latter index as it simultaneously takes into account the speed and the number of correct responses (e.g., Cavalli et al., 2017). We provide norms for both this index of efficiency and for the reading times as these two scores are considered the best measures to discriminate between dyslexic and normal adult readers (see Cavalli et al., 2017).

**The Pollueur test** The *Pollueur* test, a subtest from ECLA-16+ (Gola-Asmussen et al., 2011), is an extract on the theme of pollution and its consequences from a news review for 14-year-olds. Participants are invited to read the text aloud as quickly and as accurately as possible. The text is composed of 296 words with no particular difficulty. It is a highly cohesive text and conveys meaning without any ambiguity, thus allowing the reader to build a clear representation of its content. The number of words correctly read and the number of errors are noted by the experimenter during the 1-minute time limit. The score corresponds to the number of errors subtracted from the number of correctly read words, yielding a number of correct words from 296 within the 1-minute limit.

### Spelling skills

**Word and pseudoword dictations** The individual item dictations (Gola-Asmussen et al., 2011) are composed of three lists of 10 stimuli: regular words (e.g., “*vigne*” [*vine*]), irregular words (e.g., “*solennel*” [*solemn*]), and pseudowords (e.g., “*ribule*”). They are successively dictated once by the experimenter to participants, who have to write them down on a sheet of paper on which three columns with 10 boxes were printed. The participants are requested to write the pseudowords as they think they can write them. Participants do not have the opportunity to correct themselves. The writing times and the number of correct responses for each list are noted. The total duration of the word and pseudoword dictation is approximately 4 minutes.

**Text dictation** The text is taken from "*Traité de l'existence de Dieu*" [*Treaty of the existence of God*] (Fénélon, 1701–1712). This literacy text was, for example, proposed in 2005 to evaluate the progression of spelling in French students (see Gola-Asmussen et al., 2011). The dictation of the text is composed of 83 words in four sentences. Following the procedure from ECLA-16+ (Gola-Asmussen et al., 2011), the text is first read by the experimenter in full. Participants must not write during this time. Then, participants are invited to write the sentences under dictation. The experimenter has to be attentive to the slowness of writing. Ten lexical words (e.g., "*souerrain*" [*underground*]) and 10 grammatical words (e.g., "*petits*", plural form of [*small*]) from the text are used to calculate a score of correct responses out of 20. The duration of the dictation is around 5 minutes.

## Statistical analyses

To provide the normative data for each test, we applied the following procedure. First, we tested for and removed potential outliers lying 1.5 standard deviations below the first or above the third quartiles for each test at the whole sample level. Then, for each test, we first performed a stepwise regression analysis on the results with age, gender, education, and all interactions between these variables. The results of these regressions allowed us to select the best explanatory models for each test. Finally, post hoc analyses on the best model were used to determine relevant subgroups of participants for each test. All statistical analyses of this article were performed using *R*, version 3.5.1 (R Core Team, 2018).

The test measures corresponding to the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles for each subgroup were obtained using the percentiles' definition advocated by Crawford, Garthwaite and Slick (2009) to be used in neuropsychology: "The percentage of scores that fall below the score of interest, where half [the participants] obtaining the score of interest are included in the percentage". For each of these percentiles, a 95% confidence interval was calculated based on the binomial test indicating the lower and upper limits of the scores that would be obtained in 95% of the cases for participants from this percentile. This information is an important reminder that percentiles are just estimates of the original population of the participants and, as such, come with a certain degree of uncertainty that depends on the study sample (e.g., sample size, representativeness...).

## Results

### Vocabulary skills

**Mill Hill test part B** No outliers were found for this test. The stepwise regression analysis run on 771 participants indicated a significant model,  $F(8, 762) = 16.64, p < .001$ , with only the

educational level having a significant effect on the Mill Hill score ( $p < .001$ , partial  $R^2 = .14$ ): the higher the educational level, the higher the Mill Hill Part B scores. The post hoc analysis (Tukey's test with a  $p < .05$ ) indicated that the sample could be stratified into three categories: 12–13 years of education ( $M = 30.9; SD = 4.26$ ), 14 years of education ( $M = 32.19; SD = 3.92$ ), and 15–17 years of education ( $M = 35.21; SD = 3.6$ ). The normative data, means and standard deviations of age, and Mill Hill scores of the sample and each subgroup are presented in Table 2. The internal consistency was calculated with a split-half correlation between the odd and the even items, corrected with the Spearman-Brown formula; we also provided the interval confidence of correlation (see Oosterwijk van der Ark, & Sijtsma, 2019). The coefficient was .83 95% CI [.80 .85].<sup>1</sup>

**LexTale-FR** Six outliers were excluded for the percentage of correct responses, the most frequently used index of this test. A stepwise regression analysis run on 410 participants indicated a significant model,  $F(8, 401) = 4.15, p < .001$ , adjusted  $R^2 = .066$ , with a significant effect of educational level on the percentage of correct responses (partial  $R^2 = .07, p < .001$ ). Since the educational level effect size was rather low, we provide the normative data, means, and standard deviations for the percentage of correct responses and  $d'$  of the sample in Table 3.

The post hoc analysis (Tukey's test with a  $p < .05$ ) indicated that the sample could be stratified into three categories: 12–13 years of education ( $M = 87.42; SD = 4.19$ ), 14 years of education ( $M = 88.43; SD = 4.00$ ) and 15–17 years of education ( $M = 90.08; SD = 4.38$ ). Percentile estimates based on this stratification can be found in Table S1 of the supplemental material. The internal consistency split-half correlation corrected with the Spearman-Brown formula for the LexTale-FR test was .77 95% CI [.72 .80].<sup>2</sup>

### Reading skills

**Alouette-R test** Eight outliers were found for the index of efficiency. A stepwise regression analysis run on 1231 participants (956 women) indicated a significant model,  $F(23, 1207) = 3.41, p < .001$ , adjusted  $R^2 = .04$ , with a significant effect of educational level ( $p < .001$ , partial  $R^2 = .028$ ), a significant effect of gender ( $p < .001$ , partial  $R^2 = .018$ ) and a significant interaction effect between gender, age, and educational level ( $p = .029$ , partial  $R^2 = .01$ ). Regression analyses conducted separately for each gender group revealed a significant effect of educational level for women,  $F(5, 950) =$

<sup>1</sup> The internal consistency of the Mill Hill test was calculated on 680 participants because some protocols were no longer available.

<sup>2</sup> The internal consistency of the LexTale test was calculated on 382 participants because some protocols were no longer available.



**Table 2** Normative data for the Mill Hill Part B scores (out of 44) depending on participants' education levels ( $N = 771$ )

	Mean age (SD)	Score (SD)	Skewness (Kurtosis)	Centiles						
				5%	10%	25%	50%	75%	90%	95%
Whole sample ( $n = 771$ )	20.52 (2.18)	32.45 (4.36)	-0.08 (-0.36)	25 [24 26]	27 [26 27]	29 [29 30]	33 [32 33]	35 [35 36]	38 [38 39]	40 [39 41]
12–13 years of education ( $n = 286$ )	19.17 (1.52)	30.9 (4.26)	-0.07 (-0.42)	24 [22 24]	25 [24 26]	28 [27 29]	31 [30 32]	34 [33 35]	36 [35 38]	38 [37 39]
14 years of education ( $n = 280$ )	20.21 (1.42)	32.19 (3.92)	0.24 (-0.45)	26 [26 27]	27 [26 28]	29 [28 30]	32 [31 33]	35 [35]	38 [36 38]	39 [38 41]
15–17 years of education ( $n = 205$ )	22.8 (1.96)	35.21 (3.6)	-0.47 (0.47)	29 [25 30]	30 [29 31]	33 [32 33]	35 [35 36]	38 [37 39]	40 [39 41]	41 [40 42]

Note. SD = standard deviation. Score = mean of Mill Hill Scores out of 44. Italic values in square brackets indicate the scores corresponding to the lower and upper limits of the percentiles' 95% confidence intervals.

8.4,  $p < .001$ ,  $R^2 = .04$ , but not for men,  $F(5, 269) = .88$ ,  $p = .49$ ,  $R^2 = .01$ . A higher index of efficiency was associated with a higher educational level only for women. The post hoc analysis (Tukey's test with a  $p < 0.05$ ) indicated that the sample could be stratified into three categories for women: 12–13 years of education ( $M = 483.99$ ;  $SD = 78.76$ ), 14–15 years of education ( $M = 500.84$ ;  $SD = 80.6$ ), and 16–17 years of education ( $M = 532.44$ ;  $SD = 77.78$ ). Because of the modest effect sizes of these factors, we provide the normative data, means and standard deviations of the index of efficiency, and reading times for the whole sample in Table 4. Norms associated with a more precise stratification according to gender and level of education can be found in Table S2 of the supplemental material.

**Pollueur test** In our study, we used three tests (i.e., *Pollueur*, word/pseudoword dictation, text dictation) extracted from the ECLA-16+ battery (Gola-Asmussen et al., 2011). In order to establish a range of performances across these tests for a same patient or participant, we divided the subgroups for all these tests on the basis of the number of words correctly read in 1

minute in the *Pollueur* test. One outlier was removed. A stepwise regression analysis run on 361 participants revealed a significant model,  $F(8, 352) = 2.82$ ,  $p = .005$ , adjusted  $R^2 = .04$ , with significant effects of the educational level ( $p = .03$ , partial  $R^2 = .03$ ), gender ( $p = .02$ , partial  $R^2 = .005$ ), and interaction between age and gender ( $p = .03$ , partial  $R^2 = .01$ ). Regression analyses conducted independently for each gender group indicated a significant effect of educational level ( $p = .01$ , adj.  $R^2 = .02$ ) for women, but not for men ( $p = .90$ , adj.  $R^2 = -.02$ ). Tukey's post hoc test indicated that only the scores of women with 16 years of education ( $n = 43$ ,  $m = 191.07$ ) were significantly higher than those of women with 13 ( $n = 125$ ,  $m = 179.96$ ,  $p = .04$ ) and 15 years of education ( $n = 43$ ,  $m = 173.72$ ,  $p = 0.02$ ). Because these factors had a limited practical impact on the norms, especially when taking the percentile confidence intervals into account, we provide the normative data, means and standard deviations of the number of words correctly read in 1 minute, and the number of errors for the whole sample (see Table 5). For the sake of exhaustivity, norms for the small sample of men and for women stratified into three levels of education (12–13 years of

**Table 3** Normative data of the percentage of correct responses and  $d'$  of the LexTale-FR ( $N = 410$ )

Centiles of the percentage of correct responses										
	Mean age (SD)	Score (SD)	Skewness (Kurtosis)	5%	10%	25%	50%	75%	90%	95%
Whole sample ( $n = 410$ )	20.29 (2.05)	88.4 (4.29)	-0.35 (-0.19)	80 [79 82]	83 [82 84]	86 [85 87]	88 [88 89]	92 [91 92]	94 [93 95]	95 [95 96]
Centiles of $d'$										
Whole sample ( $n = 410$ )	20.29 (2.05)	2.65 (0.5)	0.1 (-0.27)	1.82 [1.71 1.92]	1.99 [1.92 2.06]	2.26 [2.22 2.32]	2.66 [2.59 2.72]	3.02 [2.95 3.09]	3.27 [3.25 3.34]	3.45 [3.34 3.57]

Note. SD = standard deviation. Score = mean of the percentage of correct responses and  $d'$ . Italic values in square brackets indicate the scores corresponding to the lower and upper limits of the percentiles' 95% confidence intervals.

**Table 4** Normative data of the index of efficiency and the reading time of Alouette-R Test depending on participants' gender and education ( $N = 1231$ )

		Centiles of the index of efficiency									
Years of education		Mean age ( <i>SD</i> )	Score ( <i>SD</i> )	Skewness (Kurtosis)	5%	10%	25%	50%	75%	90%	95%
Whole sample ( $n = 1231$ )	12–17	20.5 (2.19)	493.4 (82.62)	0.14 (-0.25)	368 [355 378]	391 [387 385]	433 [426 438]	493 [486 500]	548 [542 555]	604 [596 611]	635 [624 648]
	12–13 ( $n = 361$ )	18.99 (1.42)	483.99 (78.76)	0.24 (-0.07)	370 [334 383]	390 [380 393]	426 [415 439]	482 [470 496]	532 [524 538]	585 [578 605]	624 [601 653]
	14–15 ( $n = 477$ )	20.36 (1.58)	500.84 (80.6)	0.227 (-0.27)	382 [366 392]	402 [392 413]	441 [433 450]	494 [485 504]	553 [544 566]	610 [602 627]	647 [626 667]
	16–17 ( $n = 118$ )	23.4 (1.87)	532.44 (77.78)	-0.02 (-0.33)	412 [379 423]	424 [411 438]	481 [451 500]	534 [522 544]	590 [564 611]	637 [614 663]	663 [638 712]
Male ( $n = 275$ )	12–17	21.39 (2.39)	476.1 (86.35)	0.07 (-0.32)	347 [311 356]	371 [353 383]	406 [395 422]	481 [457 494]	540 [522 553]	585 [567 602]	607 [592 629]
	12–13 ( $n = 1231$ )	20.5 (2.19)	97.73 (16.95)	0.82 (1.11)	126 [124 132]	119 [118 121]	108 [107 110]	95 [94 97]	86 [85 87]	78 [77 79]	74 [72 75]
	12–13 ( $n = 361$ )	18.99 (1.42)	98.97 (16.35)	-1.10 (1.12)	126 [139 121]	119.53 [122 116]	109 [111 107]	97 [100 94]	88 [90 87]	80 [81 77]	75.4 [77 72]
	14–15 ( $n = 477$ )	20.36 (1.58)	96.15 (15.68)	-1.15 (1.58)	121 [128 118]	117 [118 115]	107 [109 104]	95 [97 93]	85 [87 83]	77 [78 75]	73 [75 71]
Female ( $n = 956$ )	12–17	21.39 (2.39)	102 (9.49)	-1.85 (1.12)	134 [150 132]	127.13 [132 123]	116.83 [118 111]	97 [103 94]	87 [90 85]	80 [82 78]	77.1 [80 74]
	12–13 ( $n = 275$ )	20.5 (2.19)	90.34 (14.04)	-1.85 (1.12)	115.50 [124 110]	109.83 [116 106]	97.08 [105 94]	88 [90 86]	80 [84 77]	74 [77 71]	71 [74 67]
	14–15 ( $n = 477$ )	20.36 (1.58)	90.34 (14.04)	-1.85 (1.12)	115.50 [124 110]	109.83 [116 106]	97.08 [105 94]	88 [90 86]	80 [84 77]	74 [77 71]	71 [74 67]
	16–17 ( $n = 118$ )	23.4 (1.87)	102 (9.49)	-1.50 (3.33)	134 [150 132]	127.13 [132 123]	116.83 [118 111]	97 [103 94]	87 [90 85]	80 [82 78]	77.1 [80 74]

Notes. *SD* = standard deviation. Score = mean of the index of efficiency. Time = mean of the reading time. Italic values in square brackets indicate the scores corresponding to the lower and upper limits of the percentiles' 95% confidence intervals.

**Table 5** Norms of number of correctly read words in 1 minute and number of errors for the “Pollueur” test from ECLA-16+ ( $N = 361$ )

Sample ( $n = 361$ )	Mean age ( <i>SD</i> )	Scores ( <i>SD</i> )	Skew (Kurtosis)	5%	10%	25%	50%	75%	90%	95%
Number of words correctly read in 1 minute	20.29 (2.05)	183.4 (23.18)	0 (-0.14)	145 [140 149]	151 [149 156]	168 [163 172]	184 [182 186]	200 [196 202]	214 [209 218]	221 [217 228]
Number of errors	20.29 (2.05)	5.26 (3.71)	0.55 (0.32)	12 [11 13]	10 [9 11]	8 [7 9]	5 [4 5]	2 [3 3]	1 [0 1]	0 [0 0]

Note. *SD* = standard deviation. Scores = number of words correctly read in 1 minute and number of errors. Errors = number of errors during reading. Italic values in square brackets indicate the scores corresponding to the lower and upper limits of the percentiles' 95% confidence intervals. Inf = infinity.

education [ $M = 179.95$ ;  $SD = 21.55$ ], 14–15 years of education [ $M = 182.12$ ;  $SD = 21.91$ ] and 16–17 years of education [ $M = 191.56$ ;  $SD = 24.65$ ] can be found in Table S3 of the supplemental material.

### Spelling skills

**Regular and irregular word and pseudoword dictations** The normative data, means and standard deviation of writing times, and the number of correct responses for regular words, irregular words, and pseudowords for the whole sample are presented in Table 5. For the sake of exhaustivity sake, norms for the small sample of men and for women stratified into three levels of education (12–13 years of education [ $M = 179.95$ ;  $SD = 21.55$ ], 14–15 years of education [ $M = 182.12$ ;  $SD = 21.91$ ], and 16–17 years of education [ $M =$

191.56;  $SD = 24.65$ ]) can be found in Table S4 of the supplemental material.

The internal consistency split-half correlation corrected with the Spearman-Brown formula was .20 95% CI [.09 .30] for word dictation, .60 95% CI [.53 .66] for irregular word dictation, and .12 95% CI [.02 .22] for pseudoword dictation.

**Text dictation** The same subgroups as those considered for the *Pollueur* test were also used for this test. Based on the gender dichotomies and the educational dichotomies in three population subgroups for women only, the norms of the number of correct responses were determined for lexical and grammatical words. The normative data, means, and standard deviations of the number of correct responses for lexical and grammatical words are given in Table 5. The internal consistency, as measured by split-half correlation corrected with Spearman Brown formula, was .59 95% CI [.52 .65] for the text

**Table 6** Norms of writing times. Number of correct responses of words, pseudowords, and text dictations from ECLA-16+ ( $N = 361$ )

Sample ( $n = 361$ )	Mean age ( <i>SD</i> )	Time ( <i>SD</i> )	Skew (Kurtosis)	5%	Centiles of writing times					
					10%	25%	50%	75%	90%	95%
Regular words	20.29 (2.05)	44.39 (7.4)	2.15 (12.86)	56 [55 60]	54 [52 55]	48 [47 49]	44 [43 44]	39 [38 40]	37 [36 37]	36 [34 36]
Irregular words	20.29 (2.05)	47.16 (10.08)	1.92 (6.15)	67 [63 73]	58 [56 64]	50 [49 52]	46 [44 47]	41 [40 42]	37 [36 39]	35 [33 37]
Pseudowords	20.29 (2.05)	52.48 (9.75)	1.55 (4)	72 [66 77]	64 [61 67]	57 [56 57]	51 [50 51]	47 [46 47]	42 [41 43]	40 [39 41]
					Centiles of correct responses					
Regular words	20.29 (2.05)	8.7 (1.23)	-1.36 (4.09)	7 [6 7]	7 [6 7]	8 [8 8]	9 [8 9]	9 [9 9]	10 [9 10]	10 [10 10]
Irregular words	20.29 (2.05)	6.75 (1.95)	-0.2 (-0.58)	4 [3 4]	4 [4 5]	5 [5 6]	7 [7 7]	8 [8 8]	9 [9 10]	10 [10 10]
Pseudowords	20.29 (2.05)	8.65 (1.26)	-1.64 (6.19)	7 [6 7]	7 [7 7]	8 [8 8]	9 [8 9]	10 [9 10]	10 [10 10]	10 [10 10]

Note. *SD* = standard deviation. Score = mean of number of correct responses. Italic values in square brackets indicate the scores corresponding to the lower and upper limits of the percentiles' 95% confidence intervals.

altogether ( $\alpha = .42$  95% CI [.33 .50] for lexical words,  $\alpha = .55$  95% CI [.47 .63] for grammatical words).

## Discussion

The aim of this paper was to update and contribute to several lexical skills test norms frequently used in French studies and clinical practice on a young adult population. The vocabulary, reading, and spelling tests we selected are paper-and-pencil tests that are easy to administer in different settings in which computers are not necessarily available. Their relatively short duration (from 3 to 8 minutes each, depending on the test) makes them effective tools for researchers and practitioners alike. The new normative data we collected and analyzed for six lexical skills tests allowed us to rate young adults' performances within a reference population, aged from 18 to 26 years old, with different educational levels. Two recent international (Givord & Schwabe, 2019) and national (DEPP, 2021) reports have shown that language proficiency performance increases with educational level and that girls perform better in reading than boys. Interestingly our test measures analyses reported significant effects of the academic level only for the vocabulary measures (i.e., LexTale-FR and Mill Hill) and of both gender and academic level on the reading measures (i.e., *Pollueur* and *Alouette-R*), with better outcomes associated with higher education and better reading performance for women than men. These effects were nevertheless relatively small (partial effect sizes comprised between .02 and .14) and did not always have practical implications in terms of the discriminative power of the norms. Among all tests, only two benefitted from stratification to improve the precisions of the norms: the Mill Hill test, where norms are provided for three categories of levels of education, and the *Alouette-R* test, where the sample was stratified according to gender and three categories of educational level (for women).

## Vocabulary skills

Vocabulary tests can assess two different facets of vocabulary (see Ouellette, 2006). One relates to the depth of vocabulary, concerning knowledge of the meaning of words; the other corresponds to the breadth of vocabulary, referring to the number of lexical representations of words stored in the mental lexicon (e.g., Ouellette, 2006). In this paper, we present normative data for two vocabulary tests assessing these complementary components: the Mill Hill test part B (Deltour, 1998) and the LexTale-FR (Brybaert, 2013). Indeed, part B of the Mill Hill (synonym selection) test mainly assesses knowledge of the

word meaning, while the LexTale-FR test mainly evaluates the correct spelling of words, revealing the quality and retrieval of the lexical representations stored in the mental lexicon (Perfetti, 2007). This supports the idea that the two tools could be complementary in assessing vocabulary in young adults and do not substitute one another. These two tests presented sufficient to good internal consistency coefficients, and their scores were positively related to educational level, suggesting that the meaning of words and the quality of the lexical representations is more precise with an increase in educational level. In more practical terms, we advise one to consider the educational level when using the present Mill Hill norms, since the percentiles estimates of the two higher education groups (14 and 15–17 years of education) tend to be above those (and the confidence intervals) of both the whole sample and the group with 13 years of education. For example, an individual obtaining a score of 30 can be considered as having a low score after 15 to 17 years of schooling (percentiles 5 to 10), while this performance will be considered just average after 12 to 13 years of education (percentile 50). The measures obtained in LexTale-FR did not vary as much, and educational groups' percentiles overlapped the confidence intervals of the whole sample, making the latter a sufficient index of performance. A possible exception would be the somewhat higher estimates observed for the  $d'$  measures of the group with 15–17 years of education compared to the whole sample (true for centiles 10 to 95, see Table S1 in supplemental material).

Furthermore, an interesting finding is that the mean of the Mill Hill score part B in our sample (18–26 years old; mean = 32) was lower than the first category of age (20–29 years old; mean = 36) of Deltour's norms (1998). This indicates either a possible decrease in the level of vocabulary between 1998 and 2018 or a cohort effect that led to a decreased familiarity with the words used in the test. Either way, the decrease in vocabulary scores calls for the provision of updated norms that include the level of education.

## Reading skills

The *Alouette-R* (LeFavrais, 2005) and the *Pollueur* (Gola-Asmussen et al., 2011) tests are two reading-aloud tests performed in a limited timeframe that differ in their format. The *Alouette-R* test contains disorienting illustrations scattered around the text and the inclusion of phonological neighbors to frequent word associations in French. These elements could lead readers that rely on strategies other than simple decoding to make errors, and the weak semantic coherence of the text prevents any efficient prediction of the words to come. At the other end of the

scale, the *Pollueur* test is an easy-to-read, meaningful text in which poor readers can find cohesive content and in which expectations will generally lead to correct guesses. These tests can thus be complementary since the former mainly assesses the efficiency of cognitive processes at the root of reading (such as decoding), while the latter measures more ecological reading efficiency by allowing the use of compensatory strategies.

In the *Alouette-R* test, our participants had a mean efficiency index of 493.94, with a mean reading time of 97.73 seconds. These results are inferior to those of Cavalli et al. (2017) for their skilled readers (i.e., index of efficiency 552.2,  $t(1455) = -8.17$ ,  $p < .001$ ; reading time 87.2,  $t(1455) = 7.72$ ,  $p < .001$ ). One possible explanation could be the fact that the two samples were not entirely equivalent: Cavalli et al.'s 164 university students were 3.22 years older and had one more year of education on average than our sample. Interestingly, we noted an influence of gender on the efficiency index in our sample that was not reported in their study, with women performing better than men. Moreover, we observed that the efficiency index also increases with the level of education, albeit only for women. These education- and gender-based distinctions led us to stratify our sample since they can have important implications when evaluating an individual's reading ability. Based on data from the *Alouette-R* test, Cavalli et al. (2017) proposed a cut-off value to detect dyslexic students in a sample of students with dyslexia ( $n = 83$ ) and without dyslexia ( $n = 164$ ) based on the efficiency index. This cutoff (efficiency score  $< 402.26$ ) enabled the authors to correctly detect 83.1% of dyslexic individuals with a 100% specificity, as no skilled readers obtained a score below this value. Applying Cavalli et al.'s proposed cut-offs to our own sample would lead to a similar specificity (96 %) only for women having completed three years of higher education (16–17 years of education in total). This specificity would decrease to 92.2% for women with 14–15 years of education and to 76.9% for female participants with 12–13 years of education, and to 77.3 % for our male participants (independent of educational level). The provision of the present norms may therefore help reduce the misclassification of normal readers as dyslexics, especially for females just beginning university and for males.

Level of education and gender have also been found to influence reading performance in the *Pollueur* test, but not in a way that impacts the discrimination of individuals based on the percentile estimates. We advise one to use whole-sample-based norms because the confidence intervals of percentiles for each subgroup categorized according to educational level largely overlap with those of the whole sample. For example, although a number of 155

correctly read words corresponds to the 5th percentile of the more educated women, the uncertainty around this score allows a more credible alternative classification in the 10th percentile, which would correspond to the one indicated for the whole sample. Gola-Asmussen et al. (2011) normed this test on a sample of 311 individuals that included only 14 university students. The inclusion of a larger student sample allowed us to better and more specifically characterize the reading performance of the students. At the level of the whole sample, our results are quite comparable to those of Gola-Asmussen et al. (2011) in terms of reading efficiency (183.4 words correctly read in 1 minute vs. 183.51, respectively), although we observed a much higher mean number of errors (5.26 errors vs. 1.58). This indicates that our globally more educated participants could read faster than Gola-Asmussen's (2011), but at the cost of a higher number of errors, yielding an equivalent reading efficiency.

### Spelling skills

Spelling skills imply different processes, including the use of grapho-phonemic correspondence rules, the retrieval of lexical representations stored in the mental lexicon, and the use of grammar rules. These processes have been assessed using pseudoword, word, and text dictation tasks (Gola-Asmussen et al., 2011). The pseudoword dictation task requires precise knowledge of grapho-phonemic correspondence rules for the participants to correctly write these stimuli. Such knowledge is more evidently put to use in this task than in the word dictation task since pseudowords are encountered by the participants for the first time and require them to explicitly rely on what they know about these rules. The word dictation difficulty varies according to word regularity, regular words being easier to spell than irregular ones. To correctly spell irregular words, readers have to know them and retrieve the corresponding lexical representations stored in their mental lexicon. Finally, the text dictation involves knowing the grammar rules and being able to retain the sentence content in working memory accurately enough to correctly transcribe the gender and number agreements of verbs, adjectives, and nouns. Since these spelling tests belong to the ECLA-16+ battery, the present norms were stratified according to those we proposed for the *Pollueur* test, in which the participants were separated by gender (female vs. male) and by educational level for women only.

As in Gola-Asmussen et al. (2011), regular word and pseudoword dictation were more easily performed than irregular word dictation. However, our results for writing times from each list of word and pseudoword dictations were generally longer than those of Gola-Asmussen et al.

(2011), and the number of correct responses was also higher. Given the high number of correct responses for our sample and the resulting ceiling effect on this measure, we strongly advise one to rely on writing time to assess participants rather than on the number of correct responses. Since the instructions were similar, such variations do not seem to be due to procedural differences between the two studies. The higher age and/or years of education of our university students compared to Gola-Asmussen's high-schoolers may have led them to prioritize accuracy over writing speed in this test. It is noteworthy that no speed-accuracy trade-off was observed in the individuals in our sample.

A superior performance to that of Gola-Asmussen et al. was found in the dictation text lexical (8.47 vs. 7.22) and grammatical word scores (8.45 vs. 6.39) and is likely due to our participants' higher age and/or years of education. This observation suggests that lexical representations of words and grammatical knowledge may become more precise as individuals further their level of education. While Gola-Asmussen et al. (2011) did not provide gender-specific standards, they did highlight the significantly better achievement of girls compared to boys in a complementary analysis. The gender impact was also found in the present study.

Finally, it is noteworthy that the internal consistency coefficients for the dictation and their confidence intervals are low, ranging from .12 to .60. We do not have this information for the Gola-Asmussen et al. (2011) sample test, and therefore do not know whether the low internal consistency measured in our study stems from the test itself or from our population. This calls for an update of the items that may render the test more efficient at discriminating individuals in a population of 18–26-year-old students.

### Limitations

Several limitations are present in this study. First, the number of women was substantially higher than that of men. The female population is therefore overrepresented and allows a better precision than for male students. The provision of confidence intervals (that depend on the size of the sample) for all norms is an indicator that we strongly encourage taking into account. Second, the definition of subgroups by gender and/or educational level has been computed in a post-hoc fashion on the basis of the variations observed in our specific sample. This may result in categories that sub-optimally represent the variations that truly occur in the general population of young adults. For example, the educational categories relevant for the Mill Hill performance were 12–13, 14, and 15–17 years of schooling, while the ones for the *Alouette-R* test were 12–13, 14–15, and 16–17. It is not possible at this stage

to determine whether this difference is linked to an oddity of the random sampling or rather pertains to differential maturational trajectories of vocabulary and reading skills. Third, it should be noted that the spelling tests exhibited low internal consistency, which raises questions regarding their reliability. While the standards proposed here still are an improvement over existing norms, interpreting the results from these specific tests should be done with caution. Finally, of the six tests used in this study, one was initially created to evaluate a second language (i.e., LexTale-FR), but we believe it could provide useful information for first language assessments as well. The norms we propose are based on our sample of native French speakers and thus constitute a reference to which further studies could be compared, depending on the proficiency of the participants (e.g., native speakers or not).

## Conclusion and implications

This article provides current lexical skill testing standards in the French language for young adults, with data collected from several hundred students for six tests that are widely used by researchers and practitioners. All of these paper-and-pencil tests are quick and easy to administer. They are also easy for the participants to understand. We believe that our normative data, collected from university students, may be useful for future research and practice since the number of individuals with a baccalaureate and/or at university has increased considerably in the last decade (INSEE, 2018). These norms complement and/or improve the existing norms of these tests, allowing practitioners and researchers to classify individuals aged from 18 to 26 years old, taking into account the participants' demographical characteristics such as gender and/or educational level that can impact their expected performance.

## Appendix

**Table 7** Correlation Coefficient Matrix of Scores of the Mill Hill and the LexTale-FR Tests

	1	2	3
1. Mill Hill score	–		
2. <i>d'</i> of LexTale-FR test	0.54***	–	
3. % correct responses of LexTale-FR test	0.53***	0.94***	–

Notes. Significance codes: '\*\*\*'  $p < .001$ , '\*\*'  $p < .01$ , '\*'  $p < .05$ , '.'  $p < .10$ , ' '  $p > .10$ . % = percentages

**Table 8** Correlation Coefficient Matrix of Scores of the ‘Alouette’ and ‘Pollueur’ Reading Tests

	1	2	3	4
1. Reading times of <i>Alouette-R</i> Test	–			
2. Index of efficiency of <i>Alouette-R</i> Test	–0.97***	–		
3. Number of errors of <i>Pollueur</i> test	0.11**	–0.13**	–	
4. Number of correctly read words in 1 minute of <i>Pollueur</i> test	–0.70***	0.72***	–0.08.	–

Notes. Significance codes: ‘\*\*\*’  $p < .001$ , ‘\*\*’  $p < .01$ , ‘\*’  $p < .05$ , ‘.’  $p < .10$ , ‘ ’  $p > .10$ .

**Table 9** Correlation Coefficient Matrix of Scores of the Words, Pseudowords and Text Dictations from ECLA-16+

	1	2	3	4	5	6	7	8
1. Regular word dictation writing times	–							
2. Irregular word dictation writing times	0.65***	–						
3. Pseudoword dictation writing times	0.60***	0.63***	–					
4. % correct responses of regular word dictation	–0.17***	–0.18***	–0.08	–				
5. % correct responses of irregular word dictation	–0.11*	–0.16***	0.002	0.38***	–			
6. % correct responses of pseudoword dictation	–0.04	–0.01	–0.10.	0.49***	0.35***	–		
7. Correct responses of lexical words of text dictation	–0.09.	–0.11*	–0.001	0.13**	0.40***	0.43***	–	
8. Correct responses of grammatical words of text dictation	–0.12*	–0.09.	–0.13**	0.30***	0.18***	0.09.	0.12*	–

Notes. Significance codes: ‘\*\*\*’  $p < .001$ , ‘\*\*’  $p < .01$ , ‘\*’  $p < .05$ , ‘.’  $p < .10$ , ‘ ’  $p > .10$ . % = percentages

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.3758/s13428-021-01684-5>.

## References

- Andrews, S. (2012). Individual differences in skilled visual word recognition and reading: The role of lexical quality. In James Adelman (Ed.), *Visual Word Recognition Volume 2: Meaning and context, individuals and development*, (pp. 151-172). : Psychology Press
- Andrews, S. (2015). Individual Differences Among Skilled Readers. In A. Pollatsek & R. Treiman (Eds.), *The Oxford Handbook of Reading*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199324576.013.8>
- Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, 139(2), 299–318. <https://doi.org/10.1037/a0018366>
- Andrews, S., & Lo, S. (2012). Not all skilled readers have cracked the code: Individual differences in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(1), 152–163. <https://doi.org/10.1037/a0024953>
- Ben-David, B. M., Erel, H., Goy, H., & Schneider, B. A. (2015). “Older is always better”: Age-related differences in vocabulary scores across 16 years. *Psychology and Aging*, 30(4), 856–862. <https://doi.org/10.1037/pag0000051>
- Bertrand, J. M., Moulin, C. J. A., & Souchay, C. (2017). Short-term memory predictions across the lifespan: monitoring span before and after conducting a task. *Memory*, 25(5), 607–618. <https://doi.org/10.1080/09658211.2016.1200625>
- Brybaert, M. (2013). Lextale\_FR A Fast, Free, and Efficient Test to Measure Language Proficiency in French. *Psychologica Belgica*, 53(1), 23. <https://doi.org/10.5334/pb-53-1-23>
- Bürki, A., Besana, T., Degiorgi, G., Gilbert, R., & Alario, F.-X. (2018). Representation and selection of determiners with phonological variants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0000643>
- Burt, J. S., & Tate, H. (2002). Does a Reading Lexicon Provide Orthographic Representations for Spelling? *Journal of Memory and Language*, 46(3), 518-543. <https://doi.org/10.1006/jmla.2001.2818>
- Cavalli, E., Colé, P., Leloup, G., Poracchia-George, F., Sprenger-Charolles, L., & El Ahmadi, A. (2018). Screening for Dyslexia in French-Speaking University Students: An Evaluation of the Detection Accuracy of the *Alouette* Test. *Journal of Learning Disabilities*, 51(3), 268–282. <https://doi.org/10.1177/0022219417704637>
- Chetail, F. & Mathey, S. (2012). Effect of syllable congruency in sixth graders in the lexical decision task with masked priming. *Scientific Studies of Reading*, 16, 537-549.
- Cohen-Shikora, E. R., & Balota, D. A. (2016). Visual word recognition across the adult lifespan. *Psychology and Aging*, 31(5), 488–502. <https://doi.org/10.1037/pag0000100>
- Davies, R. A. I., Arnell, R., Birchenough, J. M. H., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(8), 1298–1338. <https://doi.org/10.1037/xlm0000366>
- Deltour, J. J. (1998). Échelle de vocabulaire Mill Hill. Paris: Éditions et Applications Psychologiques [Trad. de J.C. Raven. 1943. Mill Hill Vocabulary Scale].
- DEPP (2021). Journée défense et citoyenneté 2020 : près d’un jeune Français sur dix en difficulté de lecture. Note d’information, N°

- 21.27, Ministère de l'Éducation Nationale, de la Jeunesse et des Sports. <https://www.education.gouv.fr/journee-defense-et-citoyennete-2020-pres-d-un-jeune-francais-sur-dix-en-difficulte-de-lecture-323603>
- Dujardin, E., & Mathey, S. (2019). Effects of deletion neighbourhood frequency and individual differences in lexical decision, progressive demasking, and naming. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*. <https://doi.org/10.1037/cep0000193>
- Dupart, M., Auzou, N., & Mathey, S. (2018). Emotional Valence Impacts Lexical Activation and Inhibition Differently in Aging: An Emotional Hayling Task Investigation. *Experimental Aging Research* 44(3):206-20. <https://doi.org/10.1080/0361073X.2018.1449587>
- Ecalte, J., & Magnan, A. (2015). *L'apprentissage de la lecture et ses difficultés*. Paris: Dunod.
- Ecalte, J., & Magnan, A. (2008). Relations between print exposure and literacy skills: New evidence from Grade 1-5. *British Journal of Developmental Psychology*, 26(4), 525–544. <https://doi.org/10.1348/026151007X267959>
- Givord, P., & Schwabe, M. (2019). Note par Pays, France. PISA 2018 Results (Volume I-III). OECD. [https://www.oecd.org/pisa/publications/PISA2018\\_CN\\_FRA\\_FRE.pdf](https://www.oecd.org/pisa/publications/PISA2018_CN_FRA_FRE.pdf)
- Gola-Asmussen, C., Lequette, C., Pouget, G., Rouyer, C., & Zorman, M. (2011). *Outil d'évaluation de compétences de lecture chez l'adulte de plus de 16 ans* [Assessment tool for reading competence in adults over 16]. Université de Provence Aix-Marseille et Cognisciences LES Université Pierre Mendès.
- Institut national de la statistique et des études économiques. (2018). *Effectif d'élèves et étudiants en 2017 : Données annuelles de 1980 à 2017*. Paris: INSEE, 2018
- Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicologica*, 35(1), 49-66.
- Lefavrais, P. (1967). *Test de l'alouette: manuel*. : Les éditions du centre de psychologie appliquée.
- Lefavrais, P. (2005). *Alouette-R*. : Les éditions du centre de psychologie appliquée.
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods*, 44(2), 325-343.
- Mahé, G., Pont, C., Zesiger, P., & Laganaro, M. (2018). The electrophysiological correlates of developmental dyslexia: New insights from lexical decision and reading aloud in adults. *Neuropsychologia*, 121, 19–27. <https://doi.org/10.1016/j.neuropsychologia.2018.10.025>
- Maïonchi-Pino, N., Magnan, A., & Écalte, J. (2010). The nature of the phonological processing in French dyslexic children: evidence for the phonological syllable and linguistic features' role in silent reading and speech discrimination. *Annals of Dyslexia*, 60(2), 123–150. <https://doi.org/10.1007/s11881-010-0036-7>
- McClung, N. A., O'Donnell, C. R., & Cunningham, A. E. (2012). Orthographic learning and the development of visual word recognition. In James Adelman (Ed.), *Visual Word Recognition Volume 2: Meaning and context, individuals and development*, (pp. 173-195). : Psychology Press.
- Nelis, D., Quoidbach, J., Hansenne, M., & Mikolajczak, M. (2011). Measuring Individual Differences in Emotion Regulation : The Emotion Regulation Profile-Revised (ERP-R). *Psychologica Belgica*, 51(1), 49. <https://doi.org/10.5334/pb-51-1-49>
- Oosterwijk, P. R., van der Ark, L. A., & Sijtsma, K. (2019). Using Confidence Intervals for Assessing Reliability of Real Tests. *Assessment* 26(7):1207-16. <https://doi.org/10.1177/1073191117737375>
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554-566. <https://doi.org/10.1037/0022-0663.98.3.554>
- Pattamadilok, C., Nelis, A., & Kolinsky, R. (2014). How does reading performance modulate the impact of orthographic knowledge on speech processing? A comparison of normal readers and dyslexic adults. *Annals of Dyslexia*, 64(1), 57-76. <https://doi.org/10.1007/s11881-013-0086-8>
- Pollatsek, A., & Treiman, R. (2015). The Oxford Handbook of Reading. In A. Pollatsek & R. Treiman (Eds.), *The Oxford Handbook of Reading*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199324576.013.32>
- Perfetti, C.A. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, 11(4), 357–383. <https://doi.org/10.1080/10888430701530730>
- Raven, J. C. (1965). *Guide to using the Mill Hill vocabulary test with progressive matrices*. : HK Lewis.
- Robert, C., et S. Mathey. 2007. Aging and Lexical Inhibition: The Effect of Orthographic Neighborhood Frequency in Young and Older Adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 62(6):P340-42. <https://doi.org/10.1093/geronb/62.6.P340>
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Austria. URL <https://www.R-project.org/>.
- Shiple, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *The Journal of Psychology: Interdisciplinary and Applied*, 9, 371–377. <https://doi.org/10.1080/00223980.1940.9917704>
- Schwartz, M., & Katzir, T. (2012). Depth of lexical knowledge among bilingual children : The impact of schooling. *Reading and Writing*, 25(8), 1947-1971. <https://doi.org/10.1007/s11145-011-9308-9>
- Siérouff, E., & Haehnel-Benoliel, N. (2015). Environmental script affects lateral asymmetry of word recognition: A study of French–Hebrew bilinguals tested in Israel and in France. *Laterality: Asymmetries of Body, Brain and Cognition*, 20(4), 389–417. <https://doi.org/10.1080/1357650X.2014.988220>
- Vigneau, F. (2007). Trois mesures du vocabulaire chez des élèves d'écoles françaises du Nouveau-Brunswick. *Revue de l'Université de Moncton*, 38(2), 181. <https://doi.org/10.7202/038495ar>
- Willemin, J., Hausmann, M., Brysbaert, M., Dael, N., Chmetz, F., Fioravera, A., Gieruc, K., & Mohr, C. (2016). Stability of Right Visual Field Advantage in an International Lateralized Lexical Decision Task Irrespective of Participants' Sex, Handedness or Bilingualism. *Laterality: Asymmetries of Body, Brain and Cognition* 21(4-6):502-24. <https://doi.org/10.1080/1357650X.2015.1130716>
- Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance*, 38(1), 53–79 <https://doi.org/10.1037/a0024177>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.