Developmental dyslexia is a persistent reading disorder characterized by inaccurate or slow and effortful reading but also by poor spelling (American Psychiatric Association, 2013). It is a lifelong impairment, as numerous symptoms of childhood dyslexia persist into adulthood, particularly deficits in phonological processing, decoding skills, word reading fluency, phonological short-term memory, and phonemic awareness (Cavalli et al., 2016; Gagliano et al., 2015; Martin et al., 2010; Milne, Nicholson, & Corballis, 2003; Swanson & Hsieh, 2009). Neuroimaging studies have confirmed the persistence of phonological deficits (for a review, see Richlan, Kronbichler, & Wimmer, 2011), and some results suggest that this cognitive deficit may arise from congenital dysfunction in certain cortical areas involved in phonology and reading (Pugh et al., 2000; Shaywitz et al., 1998; Shaywitz et al., 2003). In spite of the persistence of reading deficits, some adults with dyslexia are able to study successfully at the university level. Despite their well-documented impairments in basic reading skills, they seem to have coped with these deficits in such a way that some semantic skills and reading comprehension is not (or less) affected (Cavalli, Duncan, Elbro, El Ahmadi, & Colé, 2017; Deacon, Cook, & Parrila, 2012; Law, Wouters, & Ghesquière, 2015). For instance, in the United Kingdom, Warminent, Stothard, and Snowling’s (2013) study reported that approximately 3.2% of university students have dyslexia. In France, such data are missing due to the lack of validated tools for the screening and diagnosis of these readers.

Developmental dyslexia is a persistent reading disorder characterized by inaccurate or slow and effortful reading but also by poor spelling (American Psychiatric Association, 2013). It is a lifelong impairment, as numerous symptoms of childhood dyslexia persist into adulthood, particularly deficits in phonological processing, decoding skills, word reading fluency, phonological short-term memory, and phonemic awareness (Cavalli et al., 2016; Gagliano et al., 2015; Martin et al., 2010; Milne, Nicholson, & Corballis, 2003; Swanson & Hsieh, 2009). Neuroimaging studies have confirmed the persistence of phonological deficits (for a review, see Richlan, Kronbichler, & Wimmer, 2011), and some results suggest that this cognitive deficit may arise from congenital dysfunction in certain cortical areas involved in phonology and reading (Pugh et al., 2000; Shaywitz et al., 1998; Shaywitz et al., 2003). In spite of the persistence of reading deficits, some adults with dyslexia are able to study successfully at the university level. Despite their well-documented impairments in basic reading skills, they seem to have coped with these deficits in such a way that some semantic skills and reading comprehension is not (or less) affected (Cavalli, Duncan, Elbro, El Ahmadi, & Colé, 2017; Deacon, Cook, & Parrila, 2012; Law, Wouters, & Ghesquière, 2015). For instance, in the United Kingdom, Warminent, Stothard, and Snowling’s (2013) study reported that approximately 3.2% of university students have dyslexia. In France, such data are missing due to the lack of validated tools for the screening and diagnosis of these readers.

Keywords
adults with dyslexia, screening methodology, validation procedure, speed-accuracy trade-off, ROC curves analysis
Dyslexia is a cognitive disadvantage, which in adulthood can become a social disadvantage as well (Harrison, Rosenblum, & Currie, 2010; for a discussion, see Konur, 2002). In the employment context, adult dyslexia contributes to social inequalities and discrimination, even for adults who reach university, as it has been observed to have an impact on choices of university program and profession (Maughan et al., 2009). Students with dyslexia are underrepresented in the sciences, management, and finance sectors, which require mastery of written language and dealing with numbers and symbols and which often involve working under time pressure (Taylor & Walter, 2003). Moreover, the importance today of Internet technologies, which are based mainly on written language, may also reinforce professional and economic inequalities. The ability to screen for and diagnose dyslexia at the beginning of university is thus essential to ensuring that pedagogical and therapeutic measures can be implemented to counteract the inequalities that affect individuals with dyslexia (for United Kingdom, see Warmington et al., 2013). In addition, having a better understanding of the academic success of university students with dyslexia can help academic support services make effective recommendations and provide effective supports for incoming students (Chevalier, Parrila, Ritchie, & Deacon, 2017). In France, there is no standardized tool with adequate psychometric properties in adulthood (>18 years) to either screen for dyslexia (i.e., identify risk factors or predictive signs of reading disabilities) or diagnose it (i.e., confirm the presence of reading disabilities and specify their nature).

However, probably the most widely used reading test in French-speaking countries is the Alouette test (Lefavrais, 1967, 2005). The Alouette test is currently used by health professionals (e.g., psychologists, physicians, speech therapists) to screen for dyslexia among children and adolescents. It has even been used as a “gold standard” test (i.e., a reference test) to confirm the diagnosis of dyslexia among children, since the performances on the Alouette test are frequently used to evaluate the diagnostic abilities of different French reading tests (Bertrand, Fluss, Billard, & Ziegler, 2010). The main reason is that the Alouette test taps skills for which speed and accuracy are frequently deficient among dyslexic readers—such as reading words with grapheme-phoneme correspondences, reading irregular words, and achieving reading fluency (Sprenger-Charolles, Siegel, Jimenez, & Ziegler, 2011; Sprenger-Charolles, Colé, Béchennec, & Kipffer-Piquard, 2005)—and it prevents them from compensating for their decoding difficulties by using contextual expectations.

Crucially, however, the necessary statistical properties of the Alouette test to screen for dyslexia, including sensitivity, specificity, and cutoff value, are still lacking. This test, which yields a score expressed in terms of reading age, assesses possible reading delay and can thus serve to recommend a more thorough assessment. Moreover, although psychometric normative data for the Alouette-R (Lefavrais, 2005) for children and adolescents are available, this test is not standardized for adults, particularly those who want to pursue postsecondary studies. There is a recognized need to establish a criterion for rapid and valid screening for dyslexia in adults at the university level. The main aim of the current study is to provide the necessary data from a large normative sample of university students without dyslexia and a large validation sample of students with dyslexia. To provide evidence of test validity, we used a receiver operator characteristic (ROC) curves analysis to determine not only the sensitivity and specificity properties of the Alouette test but also the optimal cutoff values relative to the Alouette’s scores while taking speed-accuracy trade-offs and the prevalence of dyslexia into account.

### Screening for Dyslexia at the University Level

Identifying university students with dyslexia at the university entrance is a critical issue for clinical and research studies. In comparison to English, for which questionnaires, tests, and batteries are available for the screening and diagnosis of dyslexia (e.g., Fawcett & Nicolson, 1998), the tools available to assess French-speaking individuals are much more limited. Currently, we identified only three tools designed to screen for dyslexia in adolescent and young adults: The first two—ECLA-16+ (Gola-Asmussen, Lequette, Pouget, Rouyer, & Zorman, 2011) and Adult PHONOLEC (Plaza, Robert-Jahier, Gatignol, & Oudry, 2008), assess written word recognition skills and phonological skills, respectively; however, neither of these tools has been validated with a dyslexic sample, thereby restricting their use for clinical and research studies. The third, EVALAD (Pech-George & George, 2011), is currently the most comprehensive tool to screen for dyslexia, as it assesses written word recognition, spelling, reading fluency, and reading comprehension skills. However, this tool has been standardized with only a normative sample of adolescents up to 17 years old (not adults) and validated with a small sample of adolescents with dyslexia and dysorthographia (n = 30). Consequently, evidence is still lacking for the discriminant validity of these tools with statistical properties such as sensitivity, specificity, and the optimal cutoff values.

In France, dyslexia screening/diagnosis at universities is the responsibility of university medical services and referral centers for learning disabilities. It is generally based on an interview focused on the individual’s history of reading and spelling disabilities and his or her persistence and, if applicable, a formal diagnosis of dyslexia during previous schooling (up to 17 years). Contrary to English-speaking clinicians and researchers, French practitioners do not systematically use questionnaires as a screening test to detect a
history of reading difficulties indicative of dyslexia. Currently, the most comprehensive questionnaire is the Adult Reading History Questionnaire–Revisited (ARHQ-R), primarily validated for English-speaking countries (Lefly & Pennington, 2000). The ARHQ-R is a self-report measure including items on reading habits, reading and spelling abilities, and attitudes toward school and reading, separately for elementary school, secondary school, postsecondary education, and current life. This questionnaire takes little time to complete and has demonstrated strong reliability, with reported Cronbach’s α values of .92 (Cavalli et al., 2016), .93 (Deacon et al., 2012), and .96 (Kirby, Silvestri, Allingham, Parrila, & La Fave, 2008). Interestingly, Deacon et al. (2012) used the ARHQ-R to identify university students with self-reported difficulties in reading acquisition during elementary school. They compared performance on word and pseudoword reading, reading comprehension, and phonological tasks among the self-report group and two other groups of university students: one with a recent diagnosis of dyslexia and one with no self-reported reading problems. Results showed that while the self-report and diagnosed groups performed similarly on word reading, pseudoword reading, and phonological tasks, the two recruitment methods identified individuals with different adaptive strategies, mainly when examining untimed reading comprehension (better performance for diagnosed) and reading rate scores (better performance for self-report). This distinction may be useful in interpreting some of the reported differences in certain reading-related skills among university students with dyslexia studies. Moreover, the ARHQ-R has been shown to have high validity in Icelandic, which has a shallow orthography (Bjornsdottir et al., 2014), as well as in Swedish (Wolff & Lundberg, 2003), Portuguese (Alves & Castro, 2005), and German (Maurer, 2005). The replication of these findings among participants with dyslexia who are speakers of languages with varying orthographic depth suggests that difficulties with decoding, reading fluency for isolated words or words in context, written word recognition more widely, and spelling represent specific behavioral markers of dyslexia in higher education. Consequently, screening tests for dyslexia in higher education should measure each of these specific reading and spelling difficulties.

Alouette Test: Psychometric Issues

Although the Alouette test was created to screen for difficulties in learning to read, it relied on only a sample of unimpaired readers, and no validation procedure was conducted on a sample of impaired readers (but see Bertrand et al., 2010). Such a validation procedure would establish the test’s discriminatory power to distinguish dyslexic readers from nondyslexic readers (Callens, Tops, & Brysbaert, 2012; Warmington et al., 2013). The Alouette test has thus far been used to measure the reading age of dyslexic participants, which must be lower than that of a chronological-age control group (i.e., typically 2 SD, 1.65 SD, or 1 SD below the control mean) to confirm the classification (for instance, see Cavalli et al., 2016; Dole, Meunier, & Hoen, 2013; Lallier, Donnadieu, Berger, & Valdois, 2010; Mahé, Doignon-Camus, Dufour, & Bonnefond, 2014; Martin et al., 2010; Martin, Frauenfelder, & Colé, 2013; Martínez Perez, Majerus, & Poncelet, 2013; Szenkovits & Ramus, 2005; Vinckenbosch, Robichon, & Eliez, 2005).

The statistical properties of the Alouette screening test are still not available for children, adolescents, or adult readers (Lefavrais, 1967, 2005). Yet, such statistical properties are necessary for fine-grained dyslexia screening. Despite the advantages of the Alouette test, to date, a significant limitation of this test has been the absence of empirically derived cutoff scores to denote dyslexia. As such, those wishing to use the Alouette test in clinical or research studies to screen for students with dyslexia have not had guidelines regarding what scores might be indicative of dyslexia. Contrary to a diagnostic test, whose aim is to confirm or refute the presence of a learning disability and specify its nature, the aim of a screening test is to identify a subgroup that presents predictive signs of learning disabilities in relationship to control population (Inserm, 2007). A screening test must be valid, easily distributed, and quick to complete (Bimes-Arbus, Lazothes & Rouge, 2006; Malek, Mino, & Lacombe, 2000). Validity reflects the ability of a screening test to accurately identify disabled and nondisabled individuals. Measures used to evaluate the validity of a clinical test include sensitivity and specificity, which are independent of the population of interest that is subjected to the test. In general, sensitivity refers to the test’s ability to correctly identify disability in people who truly have the disability, whereas specificity refers to the test’s ability to correctly identify truly nondisabled people (Lalkhen & McCluskey, 2008). Therefore, a test with 100% sensitivity correctly identifies all people with disability (i.e., true positives), while a test with 100% specificity correctly identifies all people without disability (i.e., true negatives). An ideal screening test is exquisitely sensitive (high probability of detecting disability) and extremely specific (high probability that those without disability will screen negative; Landais & Jais, 2002).

The challenge when designing tests to assess the presence of a disability in general is to identify the cutoff value between positive and negative results. The cutoff value is chosen to optimize sensitivity and specificity or to maximize one of them. Generally, there is a trade-off between the two, and the decision must be based on their relative clinical importance. One way to determine this cutoff value is to use ROC curves. ROC curves plot the specificity of a test on the X-axis against its sensitivity on the Y-axis for all possible cutoff points (Lalkhen & McCluskey, 2008).
analysis is derived from statistical decision theory, which is particularly well suited for evaluating discriminatory performance and establishing prognostic indices (Hanley, 1989).

Present Study

The present study had three goals. The first was to precisely establish the scores obtained by a large sample of university students with and without dyslexia on the Alouette reading test to provide data from a normative sample (unimpaired readers) and a validation sample (impaired readers). The second goal was to determine two statistical properties (i.e., sensitivity and specificity) of the Alouette test to support it as a valid screening test for dyslexia in adulthood. These properties refer here to the test’s discriminatory power to identify individuals with dyslexia. In this context, the aim was to determine which of the Alouette’s measures (i.e., accuracy or speed) was the most sensitive and specific to screen for dyslexia and then to determine the cutoff values for identifying possible dyslexia, taking the speed-accuracy trade-off into account. Finally, the third goal was to provide a methodology to statistically determine the optimal cutoff value for each measure, which takes dyslexia prevalence and estimated costs between sensitivity and specificity into account (the estimated cost is computed from the positive predictive value and the negative predictive value, which depend on the prevalence of dyslexia; for a detailed review, see Metz, 1978). We took advantage of the excellent discriminatory outcomes provided by the ROC curves method, based on a set of indicators: sensitivity, specificity, and the area under the curve (AUC; i.e., an index of the overall accuracy of a test).

Material

We administered to all participants the nonverbal IQ task (Raven’s matrices; Raven et al., 1995), the phonological tasks (EVALEC; Sprenger-Charolles et al., 2005), and the Alouette test (Lefavrais, 1967, 2005).

Method

Participants

A total of 247 university students participated in the study, including 164 (97 women, 67 men) without dyslexia (hereafter, skilled readers [SR]) and 83 (50 women, 33 men) with dyslexia (hereafter, dyslexic readers [DYS]). All were recruited from Aix-Marseille University and different universities in Paris (France) to meet the requirements for representativeness of the target population. Participants were native (monolingual) French speakers with normal or corrected-to-normal hearing and vision and presented no known neurological/psychiatric disorders. All participants’ nonverbal IQ fell within the normal range (i.e., >25th percentile, Raven’s matrices; Raven, Court, & Raven, 1995; see Table 1). Within the sample of SR, 55.4% were enrolled in social science programs (e.g., psychology, law, economics, or archeology) and 44.5% in science programs (e.g., neurosciences, pharmacy, medicine, chemical physics, or mathematics). SR ranged in age from 18.6 to 39.1 years and in educational level from 0 to 9 years (where 0 corresponds to the French baccalauréat, or high school diploma), and all SR reported having any known reading disorder. Within the sample of DYS, 56.6% were enrolled in social science programs and 43.3% in science programs. DYS ranged in age from 18.2 to 38.4 years and in educational level from 0 to 9 years. All DYS were recruited via university medical services and French referral centers for learning disabilities. All DYS had a formal diagnosis of dyslexia in both elementary school and high school (up to age 17 years), and all had received previous remedial teaching for an average of 4.65 years (SD = 1.3) and reported having experienced major difficulties in reading from childhood to adulthood. Moreover, the persistence of phonological deficit in adulthood was confirmed by two speech therapists (the third and fourth authors) using a set of tasks, including a pseudoword reading task, a phonological (phonemic) awareness task, and a phonological short-term memory task (see Results section).

Table 1. Characteristics of Skilled Readers (n = 164) and Dyslexic Readers (n = 83).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Skilled</th>
<th>Dyslexic</th>
<th>t(245)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>23.7 ± 3.8</td>
<td>23.5 ± 3.9</td>
<td>0.34</td>
<td>.73</td>
</tr>
<tr>
<td>Nonverbal IQ (Raven Matrices)</td>
<td>49.6 ± 4.3</td>
<td>49.4 ± 4.6</td>
<td>0.28</td>
<td>.89</td>
</tr>
<tr>
<td>Years of higher education completed</td>
<td>3.0 ± 1.7</td>
<td>2.6 ± 1.6</td>
<td>1.5</td>
<td>.12</td>
</tr>
</tbody>
</table>
aged 5 to 14 and provides a standardized score expressed in term of reading age.

The text is composed of five sections and is accompanied by drawings that promote contextual errors (e.g., a drawing of a squirrel [écureuil] close to the word écueil [pitfall]). The text includes rare words and some spelling traps: items with silent letters (temps /tɛ̃/, nids /ni/), contextual graphemes (gai /ɛi/, geai /ɛi/), and items that are phonologically similar (Annie /a.ni/, amie /a.mi/). The test also tracks contextual anticipation, which is characteristic of the youngest and least SR (Perfetti, Goldman, & Hogaboam, 1979; Stanovich, 1984). The text contains fixed expressions that are modified (“au clair de lune” instead of the usual “au clair de la lune”). It also contains words that are similar to those suggested by the context (e.g., poison [poison] rather than poisson [fish] after lac [lake]). The test thus prevents dyslexic readers and poor readers from compensating for their written word recognition difficulties by using contextual information (Rack, Snowling, & Olson, 1992). The number of words correctly read (reading accuracy) and the time taken to read the text (reading time) were measured (see Analyses section).

**Phonological tasks.** For the phonological assessments, we administered pseudoword reading, phonological short-term memory (i.e., phonemic awareness), and phonological short-term memory tasks from EVALEC, a computerized battery of tests of reading and reading-related skills (Sprenger-Cha rolles et al., 2005). According to the phonological deficit theory (Ramus, 2003), adults with dyslexia are systemat ically impaired on these tasks per their accuracy and reaction time scores (for French studies, see Cavalli et al., 2016; Martin et al., 2010; Martin et al., 2013; Martinez Perez et al., 2013). Moreover, this set of tasks has been shown to be reliable, as demonstrated by the high values of Cronbach’s alphas in two studies—all reliability scores >.70 in Cavalli et al. (2016) and Martin et al. (2013). To further confirm the reliability of these tasks, in this study Cronbach’s alphas were calculated for accuracy and response time, and all reliability scores were >.72, suggesting good internal consistency (Kline, 2000).

**Pseudoword reading.** This task assesses the efficiency (reading time and accuracy) of the phonological decoding procedure with 20 pseudowords: 10 short and 10 long (mean pseudoword length: 4.4 and 7.8 letters, respectively). The Cronbach’s α was .86 for accuracy and .94 for response time.

**Phonemic awareness.** This task involved the deletion of the first phoneme of 12 pseudowords composed of three phonemes (consonant-consonant-vowel structure; e.g., spo). The subjects heard the items one by one through headphones and had to repeat each item without the first phoneme as accurately as possible. The time taken to complete each task (response time) and accuracy were measured. The Cronbach’s α was .72 for accuracy and .91 for response time.

**Phonological short-term memory.** This task consisted of repeating 24 pseudowords from three to six syllables (e.g., moukola). The subjects heard the items one by one through headphones and had to repeat each item. Pseudowords were presented in order of increasing syllable length, with six items for each length. The time taken to perform the whole task (response time) and accuracy (span) were measured. The Cronbach’s α was .76 for accuracy and .84 for response time.

**Analyses**

The original measures of the first version of the Alouette test (Lefavrais, 1967) are expressed in terms of accuracy (i.e., the number of words correctly read), reading time, and reading age (6–14 years). The second version/standardization of this test, Alouette-R (Lefavrais, 2005), takes both accuracy (hereafter, A) and speed (hereafter, S; defined as 1 / reading time) into account and includes an additional measure. This measure, called the CTL score, is a reading efficiency score computed by the following formula: CTL = A × 180 / RT, where A = the number of words correctly read (self-corrections included) and RT = reading time (maximum = 180 s). In a concise formula, efficiency is expressed as f(S, A) = S × A. Thus, high efficiency requires cognitive mechanisms to dynamically determine the appropriate balance between speed and accuracy in processing printed words. The CTL (efficiency) score is a good indicator of this speed-accuracy trade-off. Figure 1 presents isoquants—that is, a series of isoefficiency curves (each representing a constant efficiency), with f(S, A) = CTL = constant—each showing various combinations of speed and accuracy levels that yield a given efficiency level. Along an isocline, an increase in accuracy comes at the cost of a decrease in speed, and vice versa. In other words, readers can modulate the proportions in which they allocate resources to the mutually incompatible demands of speed and accuracy, exhibiting a certain strategic flexibility. An isocline (also known as an expansion path) shows the movement from one isocline to another in an isocline map. An expansion path is a curve connecting optimal input combinations of speed and accuracy as efficiency increases. A reader seeking to achieve optimal performance on a task attempts to increase his or her efficiency along an expansion path.

In this study, we therefore used three scores—accuracy, speed, and a combined speed-accuracy score (i.e., reading efficiency)—and we reported the mean, standard deviation, and range (i.e., minimum to maximum) for each group. This set of data should constitute a useful tool, allowing researchers and practitioners to quickly detect reading difficulties with the Alouette test. Moreover, it is crucial when establishing normative data to determine whether there is any
relationship among the participants’ characteristics. We therefore examined whether individual variables—such as gender, chronological age, educational level, and academic program—modulated the performances on the Alouette test, by computing correlation analyses between Alouette scores and individual variables.

In addition, it has been suggested that university students with dyslexia might present no deficit in word reading accuracy (Deacon et al., 2012) but that speed deficits remained present (see meta-analysis Swanson & Hsieh, 2009). That is why we focused on the efficiency measure, which takes both speed and accuracy into account, to investigate the main nature (speed and/or accuracy) of the dyslexic deficits. Interestingly, DYS may present a speed deficit, which would suggest a delay in accessing word representations, and/or an accuracy deficit, which would suggest a deficit in the quality of word representations (for a review, see Ramus, 2003). However, although the determination of two types of deficits is crucial in dyslexia research (Vellutino, Fletcher, Snowling, & Scanlon, 2004), it is relatively difficult to determine which of them is present with use of an isolated word reading task (e.g., the lexical decision task) because dyslexic readers can trade off speed for accuracy or vice versa while reading words in context (for a review of findings on speed and accuracy in dyslexic word recognition, see Kunert & Scheepers, 2014). We therefore plotted the relationship between speed and accuracy for each participant to find out whether the distribution clearly distinguished the two groups (dyslexic readers and nondyslexic readers) and to highlight potential differences in the speed-accuracy trade-off.

Finally, ROC curves analysis was then conducted on the Alouette’s measures (accuracy, speed, and reading efficiency) to determine their optimal cutoff values, which takes into account dyslexia prevalence and estimated costs between sensitivity and specificity. As recently explained by Read, Haus, Radomski, Wickham, and Borish (2015), ROC curves analysis is an application of signal detection theory that provides a graphical representation of the performance of a binary classifier as discriminant thresholds vary (Swets, 1988). ROC curves plot sensitivity versus 100 – specificity to create a curve comparing the discriminative value of an instrument relative to no discrimination, informing on how the proportions of true positives and false positives change for each possible cutoff value. The area under the ROC curve (AUC) provides a metric of the utility of the instrument, with AUC values ≥.80 are indicative of good discrimination, whereas those <.70 do not provide sufficient discrimination (Read et al., 2015). The AUC is an effective and combined measure of sensitivity and specificity that describes the inherent validity of a test, and in this case, it can be interpreted as the probability that a randomly chosen disabled subject is rated or ranked as more likely to be disabled than a randomly chosen nondisabled subject (for a review, see Hajian-Tilaki, 2013). AUC, specificity, sensibility, and dyslexia prevalence were used to determine optimal cutoff values for accuracy, speed, and efficiency scores. We then reported three statistics of the ROC curves, including 95% confidence interval and associated cost, to provide different cutoff values (i.e., one that is optimal, one with 100% specificity, and one with 100% sensitivity) and account for speed-accuracy trade-offs. Practitioners can then directly use these values to screen for dyslexia. As part of a clinical process, this first screening stage can then suggest a diagnosis of dyslexia and should be done in association with other tests to later confirm the diagnosis through a more thorough assessment.

Results

Phonological Tasks

The mean scores and their difference effect sizes (Cohen’s $d$; Cohen, 1988) of the DYS and SR groups on the phonological tasks are given in Table 2. On the set of phonological
Table 2. Phonological Skills Assessment of Skilled Readers (n = 164) and Dyslexic Readers (n = 83).

<table>
<thead>
<tr>
<th>Skill: Measure</th>
<th>Skilled</th>
<th></th>
<th>Dyslexic</th>
<th></th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Pseudowords reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>96.7</td>
<td>4.60</td>
<td>82.3</td>
<td>15.44</td>
<td>1.2</td>
</tr>
<tr>
<td>Response time, ms</td>
<td>754</td>
<td>231</td>
<td>1,317</td>
<td>568</td>
<td>1.2</td>
</tr>
<tr>
<td>Efficiency, correct/ms</td>
<td>0.13</td>
<td>0.03</td>
<td>0.07</td>
<td>0.02</td>
<td>2.3</td>
</tr>
<tr>
<td>Phonemic awareness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>95.3</td>
<td>10.20</td>
<td>86.1</td>
<td>10.16</td>
<td>0.9</td>
</tr>
<tr>
<td>Response time, s</td>
<td>20.2</td>
<td>7.1</td>
<td>31.8</td>
<td>12.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Efficiency, correct/s</td>
<td>5.27</td>
<td>1.77</td>
<td>2.98</td>
<td>0.90</td>
<td>1.6</td>
</tr>
<tr>
<td>Phonological memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (span)</td>
<td>5.4</td>
<td>0.69</td>
<td>4.3</td>
<td>0.80</td>
<td>1.5</td>
</tr>
<tr>
<td>Response time, s</td>
<td>50.6</td>
<td>9.2</td>
<td>63.0</td>
<td>21.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Efficiency, correct/s</td>
<td>0.11</td>
<td>0.02</td>
<td>0.07</td>
<td>0.02</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Note. Each test item, \( p < .001 \) (t test).

Table 3. Means and Standard Deviations for Dyslexic and Skilled Readers on Accuracy, Reading Time, Speed, and Efficiency on the Alouette Test.

<table>
<thead>
<tr>
<th>Test items</th>
<th>Cohen’s d</th>
<th>Skilled (n = 164)</th>
<th>Dyslexic (n = 83)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Accuracy, words</td>
<td>1.10</td>
<td>262.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Reading time, s</td>
<td>2.65</td>
<td>87.2</td>
<td>11.9</td>
</tr>
<tr>
<td>Speeda</td>
<td>2.77</td>
<td>11.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Efficiencyb</td>
<td>2.87</td>
<td>552.2</td>
<td>81.1</td>
</tr>
</tbody>
</table>

Note. Each test item, \( p < .001 \) (t test).

In the aim of finding out whether specific participant characteristics—gender, chronological age, educational level (i.e., number of years of university completed), and academic program (social science vs. science)—are related to the distribution of the different performance measures, we computed correlation analyses for each group separately (see Table 4). Holm-Bonferroni correction for multiple comparisons was used to set significance thresholds (Holm, 1979). The results showed no significant correlations between individual characteristics and Alouette performances. We thus concluded that these individual variables were not relevant in analyzing Alouette performances for either group. Consequently, we did not distinguish groups based on these individual variables in subsequent analyses.

Alouette Test

Table 3 gives effect sizes, means, standard deviations, and ranges for accuracy, reading time, speed, and reading efficiency scores on the Alouette test for both groups. Accuracy is measured as the number of words correctly read (maximum = 265); reading time refers to the time taken to read the text, with a 3-minute time limit (maximum = 180 s); and speed is defined as 1 / \( RT \) (which we have arbitrarily multiplied by 1,000 to increase the clarity of results). Finally, the reading efficiency score, called CTL, is the combined speed-accuracy score based on a speed-accuracy model. The results showed that DYS were outperformed by SR on accuracy, \( t(245) = 10.6, p < .001 \), reading time and speed, \( t(245) = 21.9, p < .001 \), and reading efficiency, \( t(245) = 20.6, p < .001 \). All effect sizes were large, as computed with Cohen’s \( d \) (all \( d > 1.10 \)).
Table 4. Correlations Between Individual Variables (Gender, Age, Educational Level, and Academic Program) and Alouette Performance (Accuracy, Speed, and Efficiency Score).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Gender</th>
<th>Age</th>
<th>Educational level</th>
<th>Academic program</th>
<th>Accuracy</th>
<th>Speed</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>—</td>
<td>.12</td>
<td>—</td>
<td>—</td>
<td>.16</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Age</td>
<td>—.01</td>
<td>—</td>
<td>.48**</td>
<td>—</td>
<td>.08</td>
<td>—.03</td>
<td>.19</td>
</tr>
<tr>
<td>Educational level</td>
<td>.16</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—.11</td>
<td>—</td>
<td>—.10</td>
</tr>
<tr>
<td>Academic program</td>
<td>.02</td>
<td>.13</td>
<td>.05</td>
<td>—</td>
<td>—.01</td>
<td>—.00</td>
<td>.02</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.02</td>
<td>.04</td>
<td>—.01</td>
<td>—</td>
<td>—.05</td>
<td>—</td>
<td>—.14</td>
</tr>
<tr>
<td>Speed</td>
<td>.07</td>
<td>.16</td>
<td>.18</td>
<td>—</td>
<td>—.01</td>
<td>—</td>
<td>—.97**</td>
</tr>
<tr>
<td>Efficiency</td>
<td>.07</td>
<td>.16</td>
<td>.18</td>
<td>—</td>
<td>.01</td>
<td>—</td>
<td>—.99***</td>
</tr>
</tbody>
</table>

Note. Correlations for the dyslexic group are presented above the diagonal line and those for the skilled reading group below the diagonal line. **p < .01, ***p < .001. Holm-Bonferroni correction.

...phonological short-term memory efficiency. Correlations were calculated across groups. The results showed significant correlations between reading efficiency scores and pseudoword reading efficiency ($r = .50, p < .01$), phonological awareness efficiency ($r = .47, p < .01$), and phonological short-term memory efficiency scores ($r = .45, p < .01$), therefore confirming the convergent validity of the Alouette test to screen for dyslexia.

Figure 3 presents a scatter plot of accuracy and speed scores for both groups, with the associated linear and nonlinear regression lines. The results suggest different speed-accuracy trade-off profiles for DYS and SR. The isocline line of the DYS (in bold on the left part of the graph) shows that as efficiency increases, accuracy and speed increase concurrently. However, this efficiency reaches a limited level, beyond which dyslexic readers start to show different strategies, with some focusing on accuracy and others on speed. In contrast, SR (see their isocline line, in bold on the right part of the graph) increase efficiency exclusively by increasing speed, because of the ceiling effect on accuracy in this task. Interestingly, for the DYS, the scatter plot shows evidence of an inflection point, below which accuracy decreases while efficiency remains constant, while some manage to reach scores equivalent of the less proficient SR. One way to find out this cutoff value (or these cutoff values) is to analyze ROC curves.

To determine cutoff values, results from dyslexic and non-dyslexic groups were examined with ROC curves analysis (for a description of the methodology, see Halpern, Albert, Krieger, Metz, & Maidment, 1996), with the AUC as the index of discriminant ability (see Lindstrom, Coleman, Thomassin, Southall, & Lindstrom, 2011). For the three scores (accuracy, speed, and reading efficiency), the ROC curves (shown at left in Figure 4) follow the left-hand side and then the top border, indicating that Alouette scores show high classification accuracy. The AUC for accuracy was $.91 \pm .01$, 95% CI [0.87, 0.94]; for speed, $97 \pm .005$, 95% CI [0.96, 0.99]; and for reading efficiency, $.99 \pm .002$, 95% CI [0.97, 1.0]. Pairwise comparison of ROC curves (i.e., difference between areas) showed that the AUC for speed was greater than the AUC for accuracy ($p < .001$) and that the AUC for reading efficiency was greater than the AUC for speed ($p < .05$) and accuracy ($p < .001$). Therefore, the reading efficiency (CTL) score seems to be the best discriminatory measure of the three for screening for dyslexia in adulthood.

On the basis of the ROC curves analysis, Figure 4 presents the optimal criterion cutoff values for accuracy, speed, and reading efficiency scores, as computed on the basis of dyslexia prevalence (i.e., here we have selected 5%; Swanson, 2012) and the estimated cost resulting in the best balance between sensitivity (i.e., the proportion of true positives correctly identified with the cutoff) and specificity (i.e., the proportion of true negatives correctly identified with the cutoff). Results showed that the best balance between sensitivity and 100 – specificity was obtained by classifying individuals as dyslexic with the following cutoff values: accuracy score <253 (53% sensitivity, 100% specificity), speed score <8.7 (81.9% sensitivity, 99.4% specificity), and reading efficiency score <402.26 (83.1% sensitivity, 100% specificity).

Table 5 presents the classification data of the ROC curves with 95% CI for three scores: the optimal criterion score, as determined with the ROC curves analysis (in bold), and two scores selected on the basis of their statistical properties—one with 100% sensitivity and one with 100% specificity. We illustrate how to interpret the table with two examples in the Discussion section.

Discussion

The main goal of the present study was to determine the statistical properties of the Alouette reading test in screening for dyslexia in French-speaking adults. Specifically, after gathering data from a large normative sample of 164 adults without dyslexia and a validation sample of 83 adults with dyslexia, we used ROC curves analysis to statistically determine optimal cutoff values, while taking into account the critical issue of speed-accuracy trade-offs. The study presents an original method for both research and practice, making it possible to classify...
individuals as potentially dyslexic on the basis of scores obtained on an important clinical test.

**Alouette Screening Test: Clinical and Theoretical Issues**

For the current study, we selected a reading test that is widely used in France for dyslexia screening because it taps skills that are always deficient in dyslexia (Bertrand et al., 2010; Sprenger-Charolles et al., 2005). Moreover, since adults with dyslexia are prone to exploit contextual cues within sentences (Bruck, 1990; Corkett & Parrila, 2008), this test is particularly appropriate, as it prevents the use of contextual expectations to compensate for deficits. The present results demonstrated that the discriminative power of the *Alouette* test is good, with dyslexic readers showing significant impairments, in comparison with nondyslexic readers, on measures of accuracy, reading speed, and combined speed-accuracy score (i.e., reading efficiency). As expected, the data showed that adults with dyslexia continue to have difficulties in phonological and orthographic word processing (Gagliano et al., 2015; Martin et al., 2010; Suárez-Coalla & Cuetos, 2015).

As explained by Sprenger-Charolles et al. (2005), the *Alouette* test assesses the development of both sublexical (i.e., phonological) and lexical (i.e., orthographic) reading procedures, since regression analyses on data from children of different ages showed that, regardless of grade, measured reading age was predicted mainly by irregular word reading accuracy and by pseudoword reading latency (for adolescent readers, see also Pourcin, Sprenger-Charolles, El Ahmadi, & Colé, 2016).
The data also highlight differences between groups in speed-accuracy trade-offs (see Sprenger-Charolles, Colé, Lacert, & Serniclaes, 2000), thus emphasizing the main nature (speed and/or accuracy) of the dyslexic deficits. The present findings have implications for theories of word recognition and dyslexia, supporting the view of speed deficits as a core impairment in dyslexia (see Swanson & Hsieh, 2009). However, crucially, our findings also highlight that speed-accuracy trade-offs are unlikely to take a single unified form across all cases of dyslexia. In this context, it is worth noting the potential methodological advantages of the Alouette test, which can be used to distinguish differences in processing speed from differences in processing accuracy. This differentiation between speed- and accuracy-related effects is difficult to achieve with a classical isolated word reading task, where the generative process behind the data

Figure 4. Receiver operator characteristic (ROC) curves with confidence intervals (left) and optimal criterion cutoff plots (right) for (A) accuracy, (B) speed, and (C) efficiency (CTL) scores on the Alouette test. On the right, 0 codes for skilled readers and 1 for dyslexic readers. The horizontal line represents the optimal cutoff value and the associated sensitivity and specificity indicators.
may remain ambiguous (Kunert & Scheepers, 2014). Here, we demonstrated that a given reading efficiency level may be associated with different scores on speed and accuracy, depending whether the dyslexic participant trades off speed for accuracy or vice versa (Figures 1 and 3). More precisely, the scatter plot shows that dyslexic readers can improve speed and accuracy concurrently (i.e., the left part of the isocline line), although their efficiency remains lower than that of SR. However, once a critical speed is reached, accuracy scores generally decline (i.e., the right part of the isocline line), suggesting that some DYS trade off accuracy for speed, although some manage to reach scores equivalent to those of less proficient SR. The aim of determining cutoff values for speed and accuracy is therefore of interest for dissociating distinct dyslexic deficits and for distinguishing dyslexic from nondyslexic populations. This is precisely why we used ROC curves analysis, which showed that the Alouette test has good discriminatory power overall, as indicated by its sensitivity and specificity on each measure. We found that the optimal cutoff for accuracy was 253, which yielded perfect 100% specificity, but a medium sensitivity of 53%. The latter is presumably due to a ceiling effect in the nondyslexic group, as observed in most studies with French-speaking participants, even children (for a review, see Sprenger-Charolles et al., 2011). The optimal cutoff on speed was 8.7, which yielded 81.9% sensitivity and 99.4% specificity. Pairwise comparison of ROC curves based on the AUC showed that the discriminative power of the Alouette test’s reading efficiency measure (CTL score) was significantly greater than that of either accuracy or speed taken separately. In Table 5, we report different cutoff scores whose discriminatory power varies within a range that meets the requirements for a test to screen for dyslexia (see Glascoe & Byrne, 1993; Singleton, Horne, & Simmons, 2009). In addition, the Alouette test constitutes an important test to screen for dyslexia, but it should be used in association with other standardized screening tests in the French language, assessing reading, spelling, and reading-related skills, to diagnose dyslexia.

### Clinical Application

One aim of this study was to enable clinical applications of the Alouette test in screening for dyslexia in adults. In this section, we illustrate how to interpret results from the test using two clinical single cases, merging the different analyses used in this study (see Table 5 and Figures 3 and 4).

Participant A reads the Alouette text in 2 min 30 s (i.e., 150 s), making 15 errors (i.e., 250 words correctly read). The participant thus has an accuracy score of 250, a speed score of 6.6, and a reading efficiency (CTL) score of 300. Thus, as indicated in Table 4, the participant scored below the optimal cutoff of 253 (53% sensitivity, 100% specificity) for accuracy, below the optimal cutoff score of 8.7 (81.9% sensitivity, 99.4% specificity) for speed, and below the optimal cutoff score of 402.2 (83.1% sensitivity, 100% specificity) for efficiency (CTL score).

Participant B reads the Alouette text in 1 min 44 s (i.e., 104 s), making 14 errors (i.e., 251 words correctly read). This participant thus has an accuracy score of 251, which is below the optimal cutoff of 253 (53% sensitivity, 100% specificity); a speed score of 9.6, above the optimal cutoff score of 8.7 (81.9% sensitivity, 99.4% specificity) for accuracy, below the optimal cutoff of 10.5 (100% sensitivity, 68.9% specificity) for speed, and below the optimal cutoff of 402.2 (83.1% sensitivity, 100% specificity) for efficiency (CTL score).

### Table 5. Classification Based on Receiver Operator Characteristic Analysis (Diagnosed Dyslexic and Nondyslexic Readers) for Accuracy, Speed, and Efficiency Cutoff Scores.

<table>
<thead>
<tr>
<th>Score</th>
<th>Sensitivity</th>
<th>95% CI</th>
<th>Specificity</th>
<th>95% CI</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>264</td>
<td>100</td>
<td>[95.7, 100]</td>
<td>18.9</td>
<td>[13.2, 25.7]</td>
<td>0.730</td>
</tr>
<tr>
<td>258</td>
<td>68.6</td>
<td>[57.6, 78.4]</td>
<td>93.2</td>
<td>[88.3, 96.6]</td>
<td>0.091</td>
</tr>
<tr>
<td>253</td>
<td>53.0</td>
<td>[41.7, 64.1]</td>
<td>100</td>
<td>[97.8, 100]</td>
<td>0.047</td>
</tr>
<tr>
<td><strong>Speed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.5</td>
<td>100</td>
<td>[95.7, 100]</td>
<td>68.9</td>
<td>[61.2, 75.9]</td>
<td>0.280</td>
</tr>
<tr>
<td>8.7</td>
<td>81.9</td>
<td>[72.0, 89.5]</td>
<td>99.4</td>
<td>[96.6, 100]</td>
<td>0.023</td>
</tr>
<tr>
<td>8.4</td>
<td>74.7</td>
<td>[64.0, 83.6]</td>
<td>100</td>
<td>[97.8, 100]</td>
<td>0.025</td>
</tr>
<tr>
<td><strong>Efficiency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>458.5</td>
<td>100</td>
<td>[95.7, 100]</td>
<td>92.0</td>
<td>[86.6, 95.7]</td>
<td>0.071</td>
</tr>
<tr>
<td>430.2</td>
<td>93.9</td>
<td>[86.5, 98.0]</td>
<td>98.1</td>
<td>[94.7, 99.6]</td>
<td>0.022</td>
</tr>
<tr>
<td>402.2</td>
<td>83.1</td>
<td>[73.3, 90.5]</td>
<td>100</td>
<td>[97.8, 100]</td>
<td>0.016</td>
</tr>
</tbody>
</table>

Note. Three coordinates of the receiver operator characteristic curves (sensitivity and specificity) are presented with 95% confidence intervals (CIs) and the associated cost. The bold row represents the optimal criterion associated with the cutoff value for each measure. The interpretation of this table is illustrated with examples in the Clinical Application section.
specificity); and an efficiency (CTL) score of 434.4, above the optimal cutoff score but below the cutoff score of 458.5 (100% sensitivity, 92% specificity).

These scores should be interpreted through Figure 3. Based on their respective speed and accuracy scores, Participant A’s proficiency can be plotted on the ascending part of the isocline line, at left, while Participant B’s proficiency can be plotted on the descending part of the same isocline line, at right. The two participants thus clearly present distinct profiles of reading difficulties. On the basis of these scores, Participant A can be considered to have major reading difficulties. The practitioner may hypothesize a diagnosis of dyslexia, which should be confirmed through a more thorough assessment. Participant B may also be considered to have a reading impairment, albeit with apparently fewer reading difficulties than Participant A. Nevertheless, both participants’ efficiency (CTL) scores are below the cutoff of 458.5. The discriminatory power of this measure is greater than those of the other two measures (100% sensitivity, 92% specificity; see Figure 4). Therefore, the clinician can and should confidently hypothesize a diagnosis of dyslexia in both cases, which should then be confirmed through a further assessment.

**Screening and Diagnosis of Dyslexia in Adulthood**

Dyslexia is considered one of the most prevalent developmental specific learning disabilities. Although the diagnosis of dyslexia among children and adolescents has considerably improved, less is currently known about the prevalence and cognitive characteristics of adults with dyslexia. Specifically, even though some “screening tests” are available, their screening properties are still not known (but see, for instance, Bjornsdottir et al., 2014, for Icelandic; Warnington et al., 2013, for English). Instead, reading disabilities are mainly assessed with a classical method whereby individuals with scores either >2 SD or >1.65 SD below the mean on a reading test (phonological awareness, word and pseudoword reading, spelling, etc.) are assumed to have dyslexia. However, this methodology does not specify any of the statistical properties required to use a test for screening or diagnosis. Only a concurrent validation procedure can provide appropriate measures of sensitivity and specificity to assess the discriminatory power of a test.

As stated in the introduction, while diagnostic tests are already available for the early identification of French-speaking children and adolescents with reading difficulties, we have identified only three diagnostic batteries for the assessment of reading and reading-related skills in French-speaking adults (ECLA-16+, EVALAD, Adult PHONOLEC). Typically, a diagnostic battery is based on a cognitive model, and the assessment process consists of assessing the skills that feature in each component and route of the model. One limitation of the available French diagnostic batteries is the lack of a theoretical framework for dyslexia in adults, as well as the aforementioned lack of data on psychometric properties, such as sensitivity and specificity. For university students with dyslexia, reading still presents a major scientific challenge: understanding how readers with significantly impaired basic reading skills nevertheless manage to cope with the intensive exposure to written language required to obtain a university degree. Because research in this area has focused exclusively on deficits, the required modeling remains to be done. This focus is insufficient to create an appropriate diagnostic tool that allows assessment of not only the deficits of students with dyslexia but also their preserved cognitive skills (i.e., protective factors; see for example Haft, Myers, & Hoeft, 2016), to offer them effective support and promote their integration in educational and work contexts. In addition, while the diagnosis of dyslexia is based on impaired scores on several reading and reading-related tasks, most tasks used for this purpose have been standardized with only a normative sample of unimpaired adults, without any validation procedure on a sample of impaired adult readers.

Only a few studies have reported results from a validation procedure on tests included in batteries for assessing dyslexia in higher education. For instance, in the United Kingdom, Warnington et al. (2013) presented validation results on the *York Adult Assessment–Revised*, a battery consisting of tests of reading, spelling, writing, and phonological skills, providing data from a normative sample of 106 adults without dyslexia and a validation sample of 20 adults with dyslexia. The researchers found the discriminatory power of this battery to be good, with 80% sensitivity and 97% specificity. In addition, for the Dutch language, Callens et al. (2012) and Callens, Tops, Stevens, and Brysbaert (2014) compared the cognitive profiles of 100 university students with dyslexia and 100 students without dyslexia on a variety of tests, using an exploratory factor analysis. They proposed a screening/diagnostic battery for dyslexia in higher education with a reduced number of tasks, considered sufficient to distinguish individuals with dyslexia from those without it, but they did not provide the statistical properties, neither sensitivity nor specificity. They proposed that an assessment battery for dyslexia should include, at minimum, a measure of timed word reading, a word dictation task, a mental calculation task, and a phonological awareness task, in line with evidence that students in higher education with dyslexia continue to have specific problems with these skills (Callens et al., 2012; Callens et al., 2014).

**Conclusion**

The Alouette test appears to be a valid tool for use in screening for dyslexia that should be used in association with
other screening tests in the French language. The methodology used here enabled us to identify cutoff values for hypothesizing the presence of dyslexia. We strongly believe that these methods—including both a validation procedure and a ROC curves analysis—should be applied to all tests whose purpose is to screen or diagnose for specific learning disabilities. This study recommends that a test that presents good psychometric and diagnostic properties for distinguishing dyslexic from nondyslexic adult readers should be created and made available for various languages. To conclude, our results support the use of the Alouette reading test to screen for dyslexia, and we encourage the systematic use of a validation procedure and ROC curves analysis in the construction of screening tests.

Acknowledgments
We thankfully acknowledge all the participants in this study.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note
1. $C$ corresponds to the number of words correctly read and $TL$ to the reading time (temps de lecture in French).

References


