



## The Flamingo test: a new diagnostic instrument for dyslexia in Dutch higher education students

Liset Rouweler , Nelleke Varkevisser , Marc Brysbaert , Ben Maassen & Wim Tops

To cite this article: Liset Rouweler , Nelleke Varkevisser , Marc Brysbaert , Ben Maassen & Wim Tops (2020) The Flamingo test: a new diagnostic instrument for dyslexia in Dutch higher education students, European Journal of Special Needs Education, 35:4, 529-543, DOI: [10.1080/08856257.2019.1709703](https://doi.org/10.1080/08856257.2019.1709703)

To link to this article: <https://doi.org/10.1080/08856257.2019.1709703>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 29 Dec 2019.



[Submit your article to this journal](#)



Article views: 572



[View related articles](#)



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)

ARTICLE



## The Flamingo test: a new diagnostic instrument for dyslexia in Dutch higher education students

Liset Rouweler<sup>a</sup>, Nelleke Varkevisser<sup>a</sup>, Marc Brysbaert<sup>b</sup>, Ben Maassen<sup>a</sup> and Wim Tops<sup>a</sup>

<sup>a</sup>Department of Neurolinguistics, University of Groningen, Groningen, The Netherlands; <sup>b</sup>Department of Experimental Psychology, Ghent University, Ghent, Belgium

### ABSTRACT

In this study, we present a new diagnostic test for dyslexia, called the Flamingo Test, inspired by the French Alouette Test. The purpose of the test is to measure students' word decoding skills and reading fluency by means of a grammatically correct but meaningless text. Two experiments were run to test the predictive validity of the Flamingo Test. In the first experiment, we compared reading times, error rates and, sensitivity and specificity of the Flamingo Test for samples of students with and without dyslexia. In the second experiment, we compared performance on the Flamingo Test with reading performance on two Dutch standard word reading tests: the Leestest Een Minuut voor Studenten (LEMs; 'one-minute word reading test for students') and the Klepel, a one-minute pseudo-word reading test. Again, students with dyslexia and matched non-dyslexic students were included. Our results show that sensitivity and specificity, as well as the positive predictive value (PPV), of the Flamingo Test are high, with even slightly higher PPVs for the Flamingo Test than for LEMs and Klepel. Together with the fact that the test is short and easy to administer, we believe that the Flamingo Test is a valuable new diagnostic instrument to assess reading skills.

### ARTICLE HISTORY

Received 21 August 2019

Accepted 23 December 2019

### KEYWORDS

Dyslexia; Higher Education;  
Text Reading; The Flamingo  
Test; LEMs; Klepel

## Introduction

Good literacy skills are important for academic success and future vocation. Most adults can read and write without effort. However, about five to ten percent of the population fail to attain automatised reading and writing skills (Boulanger 2013). The term for these specific reading and writing difficulties is dyslexia, which is broadly defined by inaccurate and slow reading, and/or by poor spelling skills (Stichting Dyslexie Nederland 2016). Dyslexia is a lifelong impairment and many symptoms persist into adulthood. The profile of adults with dyslexia differs somewhat from that of children, for whom poor accuracy, slow reading and phonological deficits are among the core deficits. Adults, however, mainly face problems with the latter two: slow reading and phonological deficits (Callens, Tops, and Brysbaert 2012; Milne, Nicholson, and Corballis 2003; Swanson and Hsieh 2009).

**CONTACT** Liset Rouweler  [l.i.m.rouweler@rug.nl](mailto:l.i.m.rouweler@rug.nl)  
 Supplemental data for this article can be accessed [here](#).

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Over the last few years, an increasing number of Dutch students with dyslexia entered higher education. It is difficult to give an exact number of the prevalence of dyslexia. Dyslexia International (2017), for example, suggests that 5–10% of the people have dyslexia. Yet, some research suggests that it could even be as high as around 17% (Sprenger-Charolles and Siegel, 2016). Students with dyslexia enrolled in higher education can apply for special facilities and resources. These students have to submit proof of their learning disability, e.g. a former dyslexia certificate, or the student needs to be tested in case the report is not accepted. An assessment of the spelling and reading skills is needed to obtain access to the resources and special arrangements. There are in this case some methods available in Dutch for screening of dyslexia, such as word reading tests and questionnaires (Tamboer, Vorst, and De Jong 2017; Tops et al. 2012). However, there is a lack of validated, relatively short reading instruments for adults with dyslexia. The goal of our study is therefore to present a newly designed diagnostic reading test: the Flamingo Test.

### *Screening for reading problems at university level*

Identifying and diagnosing students with dyslexia at university entrance has been hampered, because the availability of standardised screening tests and questionnaires for adults in Dutch is more limited compared to the relatively large battery of tests available for children (Tamboer, Vorst, and De Jong 2017; Tops et al. 2012). Tamboer, Vorst, and De Jong (2017) showed that a self-report questionnaire had the highest predictive validity in screening for dyslexia, but this method unfortunately lacks the objective comparison universities typically require. An extensive test battery also showed a rather high predictive validity in their study, but it may be questioned whether this is an efficient use of resources. Tops et al. (2012) administered an extensive test battery as well and showed that a short protocol, consisting of a word reading test, a word spelling test and a phonological awareness test, is sufficient to distinguish between students with and without dyslexia.

Currently, the tests that are used most to detect reading problems in Dutch are word reading tests, because the Stichting Dyslexie Nederland (2016) defines dyslexia as a persistent reading and/or spelling problem *at the word level* (a similar definition is used internationally; Lyon, Shaywitz, and Shaywitz 2003). The standard tests are a word reading test (Een-Minuut-Test [One-minute Test]; Brus and Voeten 1994) and a pseudo-word reading test (De Klepel; Van den Bos, Lutje Spelberg, Scheepstra, & De Vries, 1994). Although these tests were initially developed for children, Tops et al. (2012) and Tamboer, Vorst, and De Jong (2017) showed the validity of these tests for Dutch/Flemish<sup>1</sup> young adults and provided norms.

Tops, Nouwels, and Brysbaert (2019) created a new version of the Dutch One minute Test (EMT), specifically designed for adults. Their Leestest Een Minuut voor Studenten (LEMs; 'One Minute Reading Test for Students'), was designed to (1) avoid ceiling effects that often occur with the original EMT, (2) to include more up-to-date words, and (3) to be freely available for research purposes. The LEMs contains 132 words with an increasing level of difficulty, whereas the EMT has only 116 words. The test has been normed on 200 first-year students in higher education and correlates .9 with EMT.

There are two more extensive test batteries available for Dutch adults with a suspicion of dyslexia (De Pessemier and Andries 2009; Van der Leij et al. 2014). De Pessemier and Andries (2009) developed the GL&SCHR, a test battery to identify reading and writing problems in Flemish speaking adolescents and young adults (from 16 to 24 years). The GL&SCHR consists of three main tests, word spelling, spelling rules and text reading, for diagnostic purposes and nine other subtests to test skills that are often associated with reading and writing problems. Even though the GL&SCHR is a validated instrument and the differences between Flemish and Dutch are expected to be small, normative data are lacking for Dutch adolescents and young adults in general. A second test, the Interactive Dyslexia test Amsterdam-Antwerpen (IDAA; Van der Leij et al. 2014), was developed as an online screening instrument and normed for adolescents and adults from the Netherlands and the Flemish-speaking half of Belgium. This test works online and with flashed presentation and taps into the spelling and reading skills.

### *Text reading as a diagnostic instrument*

In other countries, reading aloud short texts is also used for screening and diagnosis. Compared to word reading tests, text reading provides a more natural way of reading, because words are almost never read in isolation. For instance, text reading presents words next to each other in lines of text rather than underneath each other. Both types of tests assess reading time and accuracy, and yield information about the type of errors (Levafrais 1967, 2005).

Yet, researchers in English-speaking countries do not prefer text reading in dyslexia assessment, because they feel that the text contents may obscure the measurement of word decoding skills. Words in context are indeed read faster than words out of context, because the context can be used as a top-down predictor (Jenkins et al. 2003). This means that a reader with poor word decoding skills can use contextual cues as a compensatory mechanism to mask problems with decoding.

An interesting solution to this issue has been presented in the French Alouette Test (Levafrais 1967, 2005). The Alouette Test evaluates lexical decoding under normal reading conditions by using a text that is grammatically and syntactically correct, yet carries no meaning, which makes the predictability of content words very low. As a result, the Alouette Test does not provide contextual cues that the reader can use to compensate for decoding difficulties (Torgesen, Rashotte, and Alexander 2001). Interestingly, the French Alouette Test is much preferred to isolated word reading among French-speaking dyslexia practitioners and researchers (Sprenger-Charolles et al. 2005).

The Alouette Test was developed for the assessment of dyslexia in children, but Cavalli et al. (2018) showed its usefulness for adult assessment by administering the test to a large normative sample of French university students with and without dyslexia. The results showed that the test was good at predicting the diagnosis based on the outcomes of accuracy, speed and efficiency, making the Alouette Test a valid diagnostic tool for adults.

### *Present study*

For this study, we designed a Dutch adaptation of the Alouette Test, called the Flamingo Test. The Flamingo Test is not a direct translation but an adapted version of the Alouette

Test, using the same principles for constructing the test but applied to Dutch. For instance, the name *Alouette* [which means lark] is quite difficult to pronounce in Dutch [le:wəɾɪk] and was replaced by *Flamingo*.

The purpose of our study was threefold. The first goal was to get standardised scores from a reasonable sample of Dutch higher education students with and without dyslexia. We provide data from a normative group, unimpaired readers, and a validation sample of impaired readers. Our second goal was to examine the test's predictive validity and to examine sensitivity and specificity for the different outcome measures (i.e. reading accuracy, reading time and reading efficiency). Finally, our third goal is to compare the *Flamingo Test* to the commonly used tests in the Netherlands, the *LEMs* and *Klepel*, to check whether the *Flamingo Test* is as suitable as a diagnostic instrument as those tests. For this, we calculate the correlations between all three tests and compare the discriminatory power, sensitivity and specificity of each. We designed a first experiment to investigate our first two goals and a second one for our third goal.

## Experiment 1

### *Method*

#### *Participants*

40 students with dyslexia and 63 without dyslexia participated in this study. The students with dyslexia were required to have an official dyslexia certificate. Convergent validation was used to validate the diagnosis: students with dyslexia were required to have a (sub) clinical score (< pc 10) on the word reading test (*LEMs*) and the pseudo word reading test (*Klepel*) and/or the word spelling test of the *GL&SCHR* (De Pessemier and Andries 2009; Stichting Dyslexie Nederland 2016). One dyslexic student did not meet this criterion and was excluded from the study.

Students were recruited from bachelor and master programmes from university and applied science programs. All participants attended higher education in Groningen, a province in the northern part of the Netherlands. Of the participants, 13 students were master students, the majority of the participants were bachelor students. The average age for both groups was 21;8. All students had normal or corrected-to-normal vision and were native speakers of Dutch. The study followed the ethical protocol of the Faculty of Arts of the University of Groningen.

#### *Recruitment procedure*

All participants with dyslexia were recruited via the Student Service Centre by email or through advertising online at various departments of the University of Groningen. Students without dyslexia were asked directly in various departments at the University of Groningen.

#### *The Flamingo test*

The *Flamingo Test* is the Dutch adaptation of the *Alouette Test* (Levafrais 1967, 2005) and evaluates word decoding and reading fluency. The set-up of the *Flamingo Test* is similar to the set-up of the *Alouette Test*, but it is not a one-to-one translation. The test contains 285 words, which should be read aloud within a time limit of 180 seconds. The *Flamingo*

test has no meaningful content, since the text is grammatically correct but the combination of content words is meaningless. Thus, the Flamingo test prevents readers from relying on contextual clues and knowledge of the world (Torgesen, Rashotte, and Alexander 2001).

An English translation of the first two sentences reads as follows: 'Under the moss or on the roof, in living hedges or in a cleft oak, spring makes its nests. Spring with nests in the woods.'

The text is divided in five sections and is accompanied by drawings that can provoke contextual errors (e.g. a drawing of a squirrel [*EEKHOORN*] close to the word *eenhoorn* [unicorn]. The text also includes rare words like *kreupelhout* [thicket] and *capriolen* [caprices] as well as some confusing words that are orthographically or phonologically similar (e.g. *Vredeleen, mijn vriendin* [Vredeleen, my friend]). Furthermore, it contains a few words that are (phonologically) similar to the word suggested by the context (e.g. *blozen* [blushing] instead of *blinken* [blinking] after *zon* [sun]).

### Scoring

Performance on the test was expressed in three different scores: (1) an accuracy score, (2) a reading time score and (3) a reading efficiency score. The accuracy score shows the number of words correctly read by the participant, including words that were read correctly after a self-correction. The maximum score is 285. Secondly, the score for reading time indicates the required reading time in seconds. Thirdly, the reading efficiency score is the number of words correctly read per minute, calculated by the following formula: accuracy score/reading time in minutes. An additional error analysis was also conducted, including substantial errors, e.g. omissions, substitutions, and time-consuming errors, e.g. self-corrections, hesitations, repetitions.

## Results

### Accuracy, time and reading efficiency

The mean scores of the students with dyslexia and control group can be found in Table 1.

Almost every student was able to read the text within 180 seconds: only three students with dyslexia did not. The largest difference between the two groups in terms of effect size was found for reading time, which we evaluated with a Welch's t-test. Because of the ceiling effect for accuracy, the difference between the NonDys group and the Dys group was smaller in terms of effect size, but still significant. As for reading efficiency, the effect size was comparable to the effect size for reading time. Norm scores for the three dependent variables can be found in Table 2.

**Table 1.** Accuracy, time and reading efficiency scores on the Flamingo test.

	Dys		NonDys		<i>t</i>	<i>p</i>	<i>d</i>
	M	SD	M	SD			
Accuracy	274	17.1	283	2.0	158.5	< 0.001*	0.94
Time	150	16.8	111	18.6	54.3	< 0.001*	2.14
Efficiency	112	18.3	156	24.8	31.3	< 0.001*	1.96

Note. Accuracy = number of words correctly read [max. = 285]; Time = reading time in seconds; Reading efficiency = number of words correctly read in one minute; \**p* < .001; Dys = dyslexia group; NonDys = control group; *d* = Cohen's *d*.

**Table 2.** Norm scores per measure based on the NonDys group.

Percentiles	Accuracy	Time	Efficiency
1	≤ 278	≥ 166	≤ 103
5	278	139	122
10	280	132	127
15	281	126	135
20	282	124	137
25	282	123	139
30	282	120	141
35	283	118	144
40	283	115	145
45	284	111	151
50	284	110	155
55	284	108	158
60	284	105	162
65	284	104	163
70	284	103	164
75	284	100	170
80	285	98	172
85	285	92	181
90	285	91	188
95	285	89	191
99	≥ 285	≤ 71	237

Note. Accuracy = number of words correctly read [max. = 285];  
Time = reading time in seconds; Reading efficiency = number of words correctly read in one minute

### Error analysis and comparison

Errors were divided in substantial errors and time-consuming errors. The total number of errors was the sum of the number of substantial errors and time-consuming errors. Only the substantial errors were of influence on the accuracy score. Time-consuming errors were of influence on the reading time and efficiency. The mean number of errors and SD per group and per category are shown in Table 3.

The NonDys group made fewer errors than the Dys group both in total and in the two subcategories. A significant difference between the two groups was found for total errors:  $t(50.0) = 5.94$ ,  $p < .001$ ,  $d = 1.39$ , as well as for time-consuming errors ( $t(63.2) = 4.80$ ,  $p < .001$ ,  $d = 1.05$ ) and substantial errors ( $t(44.9) = 4.38$ ,  $p < .001$ ,  $d = 1.06$ ).

### Sensitivity and specificity

Cut-off scores and, true positives, false negatives, true negatives and false positives were determined for reading efficiency for each test. These scores can be found in Table 4. Cut-off scores for each category were based on the lowest 10% scores of this population.

**Table 3.** Error analysis of the Flamingo Test.

	Dys		NonDys	
	M	SD	M	SD
Substantial errors	7.2 (52%)	6.8	2.3 (41%)	2.3
Time-consuming errors	6.5 (48%)	3.7	3.3 (59%)	2.6
Total errors	13.7	8.1	5.6	3.8

Note. Substantial errors = number of errors, e.g. number of incorrect readings; Time-consuming errors = number of time errors, e.g. hesitations, self-corrections; Total errors = substantial errors + time-consuming errors; Dys = dyslexia group; NonDys = control group; total amount of errors is 100%

**Table 4.** Diagnostic accuracy of the Flamingo Test: correct and incorrect classifications.

	Cut-off	Dys (N = 40)		NonDys (N = 63)	
		True positives	False negatives	True negatives	False positives
Accuracy	< 280	29	11	58	5
Time	> 132	34	6	58	5
Efficiency	< 127	34	6	57	6
Combined score		37	3	63	0

*Note.* Accuracy = number of words correctly read; Time = reading time in seconds; Reading efficiency = number of words correctly read in one minute; true positives = correctly identified with dyslexia; true negatives: correctly identified as non-dyslexic; false negatives = dyslexic students being marked as non-dyslexic; false positives = control students being marked as dyslexic; Combined score = combining the three cut-off points: accuracy, time and efficiency

**Table 5.** Sensitivity and specificity scores of the Flamingo Test.

	Sensitivity	95% Confidence interval sensitivity	Specificity	95% Confidence interval specificity	PPV	NPV
Accuracy	72.5	56.1–85.4	92.1	82.4–97.4	85.3	84.1
Time	87.5	73.2–95.8	92.1	82.4–97.4	87.2	90.6
Efficiency	85	70.2–94.3	90.5	80.4–96.4	87.2	90.5
Combined score	92.5	79.6–98.4	100	94.3–100	100	95.5

*Note.* Cut-off points were set at the lowest 10% scores; Sensitivity = probability that a test result is positive when the diagnosis is present; specificity = probability that a test result is negative when the diagnosis is not present; Accuracy = number of words correctly read; Time = reading time in seconds; Efficiency = number of words correctly read in one minute; Combined score = combining the three cut-off points to check whether the diagnosis was still present; Positive predictor value (PPV) = probability that dyslexia is present when the test is positive; Negative predictor value (NPV) = probability that dyslexia is not present when the test is negative

In addition, we checked whether the false positives and false negatives concern the same students for each category (accuracy, time and efficiency) or whether they concern different students for each category. For this we combined the three categories, so criterion 1 (accuracy) + criterion 2 (reading speed) + criterion 3 (efficiency), in which we took all three cut-off points, i.e. the score that represents the 10<sup>th</sup> percentile, into account. A student was only identified as a false positive or false negative when (1) a dyslexic student received a score on all three criteria above the 10<sup>th</sup> percentile or (2) a non-dyslexic student received a score on all three criteria below the 10<sup>th</sup> percentile.

Based on the cut-off scores, sensitivity and specificity, and positive and negative predictor values (PPV and NPV), were determined for each measure: accuracy, reading time and reading efficiency. These scores can be found in Table 5.

After calculating the PPV and NPV for our sample group, the PPV and NPV were calculated when taking the prevalence of dyslexia into account, which we estimated around 10% based on the numbers of Dyslexia International (2017) and Sprenger-Charolles and Siegel (2016). For accuracy, the PPV was 51% and the NPV was 97%. For reading time the PPV was calculated at 55% and the NPV at 99%. The PPV for reading efficiency was 50% and the NPV was 98%. Lastly, the PPV for the combined score was calculated at 100% and the NPV at 99%.

### Comparison with French data

Although the tests are not entirely identical (e.g. language and the number of words differ), we compared our scores with those of Cavalli et al. (2018). Both the data by Cavalli

**Table 6.** Accuracy, time and reading efficiency of the Alouette test and Flamingo test.

	Alouette					Flamingo				
	Dys (N = 83)		NonDys (N = 63)		<i>d</i>	Dys (N = 40)		NonDys (N = 63)		<i>d</i>
	M	SD	M	SD		M	SD	M	SD	
Accuracy	251	13.4	262	2.2	0.12	274	17.1	283	2.0	0.94
Time	138	24.1	87	11.9	2.68	150	16.8	111	18.6	2.14
Efficiency	113	54.7	184	81.1	1.03	112	18.3	156	24.8	1.96

Note. Alouette accuracy and time data taken from Cavalli et al. (2018); Accuracy = number of words correctly read [max. score Alouette = 265, max. score Flamingo test = 285]; Time = reading time in seconds; Reading efficiency = number of words correctly read in one minute

et al. (2018) and our data are presented in Table 6. Since Cavalli et al. (2018) calculated the efficiency score over 180 seconds instead of 60 seconds, we transformed their scores to the efficiency score as calculated in our study for comparison.

The Alouette Test has a total number of 265 words, compared to 285 words in the Flamingo test. Therefore, we calculated percentages in order to compare the accuracy scores. The percentage of words that were read accurately was identical for the NonDys groups; both groups read 99–100% of the text correctly. For the Dys groups the accuracy scores were almost identical as well, 96.8% for the Dutch group and 94.7% for the French group. The effect size for accuracy was larger for the Dutch Flamingo Test. Reading time is difficult to compare, since the numbers of words in the texts differ, but we found large standardised effect sizes in both languages ( $d = 2.68$  for French compared to  $d = 2.14$  for Dutch). Finally, we compared the reading efficiency between the two tests. For the dyslexia groups the efficiency scores were very similar, 113 words read correctly (French) vs. 112 words read correctly (Dutch) per minute. The efficiency scores between the NonDys groups did differ, however.

## Discussion

The Flamingo Test is able to discriminate between students with dyslexia and students without dyslexia on all three measures, which is also the case for the Alouette Test (Cavalli et al. 2018; Levafrais 1967, 2005). Furthermore, effect sizes are comparable, with slightly higher effect sizes for Dutch on accuracy and reading efficiency.

For accuracy, the NonDys group attained ceiling level scores, with the Dys group scoring somewhat below that. This pattern was also found for the French test (Cavalli et al. 2018). Therefore, the predictive validity of accuracy as a separate measure is not as high as for reading time or efficiency. This shows that accuracy is not the most sensitive marker for dyslexia in adults (Swanson and Hsieh 2009). For instance, reading time was more sensitive: students with dyslexia were more impaired on reading speed than on accuracy, which was also shown by Cavalli et al. (2018) for French and by Swanson and Hsieh (2009). Sensitivity and specificity were the highest for this individual measure.

Similar results were found for the Dutch and French students with dyslexia on reading efficiency. Interestingly, reading efficiency was different in the non-dyslexic groups: French students read the text more efficiently. A possible explanation might be that the Dutch adaptation is a bit more difficult than the Alouette Test. In this context, analyses of the errors are of interest. A significant difference was found in the amount of errors: students with dyslexia made more substantial and time-consuming errors than students

without dyslexia, which was also true for French in Cavalli et al. (2018). However, when taking a more in-depth look, some words in the Dutch version seemed particularly challenging for both groups of students, such as *krabbelt* (error) – *kabbelt* (target). Maybe word frequency plays a role, as *krabbelt* [scribbles] is more frequent in Dutch than *kabbelt* [ripples].

When studying dyslexia, it is helpful to keep the prevalence numbers of dyslexia in mind. We estimated the prevalence number of dyslexia at 10% based on international organisations and previous literature (Dyslexia International 2017; Sprenger-Charolles and Siegel, 2016). Based on our sensitivity and specificity measures we were able to calculate the PPV and NPV for the 10% prevalence criterion. PPV's for the individual measures varied between 50 and 55%. NPV's however varied between 97 and 99%. This indicates that the Flamingo Test is able to classify between the 50 and 55% of the population with dyslexia correctly based on the individual measures. This number however increases enormously when combining all three scores, resulting in a PPV of 100% and a NPV of 99%. This indicates that the Flamingo Test shows the highest predictive validity when combining the cut-off points of all three scores.

At this point we have reason to believe that the Flamingo Test can be used as a diagnostic instrument for dyslexia in Dutch adults. The test discriminates well, and sensitivity and specificity scores are high. However, when we corrected for the prevalence of dyslexia, PPV's dropped to 50–55% on the individual measures. We thus believe that more validation is necessary. In particular, we felt it necessary to compare the Flamingo Test to LEMs and Klepel, tests that are currently used to diagnose dyslexia. This is done in Experiment 2.

## Experiment 2

### Method

#### Participants

51 students with dyslexia and 51 matched controls completed the tests. Of these students, 21 control students and 39 students with dyslexia also participated in Experiment 1. The control participants were matched to the students with dyslexia on age, gender and field of study. An official dyslexia certificate sufficed as being dyslexic. Validation of this diagnosis was done in the same way as in Experiment 1. This resulted in one match being excluded from the study.

Mean age for the control group was 21;4 and the mean age for the dyslexic group was 21;5 years old. All students had normal or corrected-to-normal vision and were native speakers of Dutch. We followed the same recruitment procedure as in Experiment 1.

#### The Flamingo test

The description of the Flamingo Test can be found in Experiment 1.

#### LEMs

The Leestest Een Minuut voor Studenten (LEMs; Tops et al., 2019) is a Dutch word reading task specifically designed for students in higher education and is based on the original EMT. Participants were instructed to read as many words as accurately and quickly as possible within one minute.

## The Klepel

The Klepel (van den Bos et al. 1994) is a Dutch pseudo-word reading test consisting of 116 pseudo-words, i.e. non-existing words that correspond to the Dutch grapheme-phoneme correspondence rules. To avoid ceiling effects, the test was administered in one minute instead of two minutes. The instructions for the Klepel were the same as for the LEMs.

## Scoring

Scoring for the Flamingo Test can be found in the method section of Study 1. For LEMs and Klepel, the number of words read, the number of errors, and the number of words that were read correctly were scored.

## Procedure

This study was part of a larger test protocol and all participants were informed about this protocol before testing. All tests in this 2,5 hour protocol were assessed in a quiet room with one experimental leader.

## Results

### LEMs, Klepel and Flamingo test

Scores for the LEMs, Klepel and Flamingo Test can be found in Table 7. Scores for the LEMs and the Klepel are divided in a raw score (number of words read in one minute) and a reading efficiency score (number of words read correctly in one minute). There was a significant difference between the groups on the raw score and the reading efficiency score on both the LEMs and Klepel.

For the Flamingo Test, there were significant differences between the Dys and NonDys groups in terms of accuracy, reading time and reading efficiency as seen in Table 7.

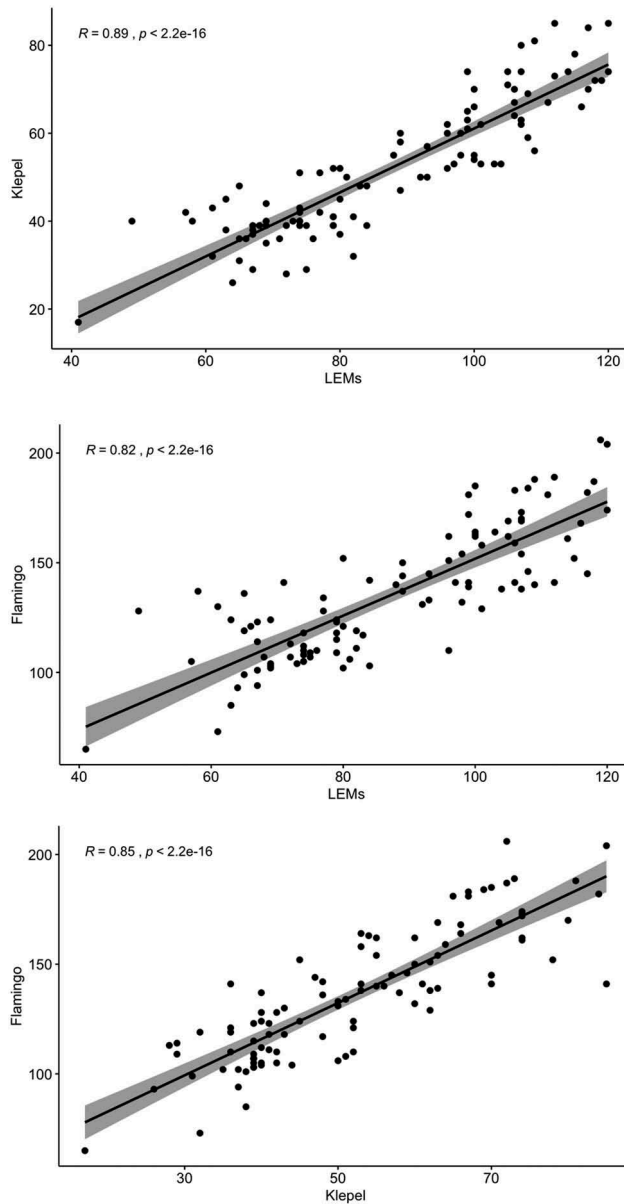
### Correlations LEMs, the Klepel and the Flamingo test

Correlations were calculated between the reading efficiency scores of each task. Figure 1 presents scatterplots of the scores of each pair of tasks. The results indicated strong, significant correlations between the Flamingo Test and the LEMs ( $r = .82, p < .01$ ), the Flamingo Test and the Klepel ( $r = .85, p < .01$ ) and between the LEMs and the Klepel ( $r = .89, p < .01$ ).

**Table 7.** LEMs, Klepel and Flamingo scores.

	Dys		NonDys		<i>t</i>	<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Flamingo							
Accuracy	275.1	14.7	282.4	2.1	187.3	< 0.001*	0.7
Time	147.7	17.5	108.5	14.1	66.8	< 0.001*	2.5
Efficiency	113.6	16.8	158.9	21.8	43.4	< 0.001*	2.3
<b>LEMs</b>							
Raw score	74.2	10.8	104.4	10.0	51.4	< 0.001*	3.0
Reading efficiency	72.4	10.6	103.5	10.0	50.7	< 0.001*	3.0
<b>Klepel</b>							
Raw score	44.6	7.3	67.6	9.6	37.9	< 0.001*	2.7
Reading efficiency	39.8	7.2	64.8	10.0	33.2	< 0.001*	2.9

Note. Raw score = number of words read in one minute; Reading efficiency = number of words read correctly in one minute; Accuracy = number of words read; Time = total reading time in seconds; \* $p < .001$ ; Dys = dyslexia group; NonDys = control group; *d* = Cohen's *d*.



**Figure 1.** Correlations (and regression line with 95% confidence interval) across groups – between the reading efficiency scores of the Flamingo test and the LEMs and Klepel.

### *Sensitivity and specificity*

Cut-off scores and, true positives, false negatives, true negatives and false positives were determined for reading efficiency for the Flamingo Test, LEMs (Top et al., 2019) and Klepel (Tops et al. 2012). These scores can be found in Table 8.

Sensitivity and specificity, and PPV and NPV were determined for the reading efficiency score of each test. These scores can be found in Table 9.

**Table 8.** Diagnostic accuracy of the Flamingo Test: correct and incorrect classifications False positives and false negatives.

	Cut-off	Dys (n = 51)		NonDys (n = 51)	
		True positives	False negatives	True negatives	False positives
Flamingo	<127	44	7	50	1
LEMs	<89	44	7	49	2
Klepel	<50	47	4	47	4

Note. true positives = correctly identified with dyslexia; true negatives: correctly identified as non-dyslexic; false negatives = dyslexic students being marked as non-dyslexic; false positives = control students being marked as dyslexic;

**Table 9.** Sensitivity and specificity scores for reading efficiency.

	Sensitivity	Specificity	95% Confidence interval	95% Confidence interval	PPV	NPV
			sensitivity	specificity		
Flamingo	82.7	98	69.7–91.8	89.4–99.7	97.7	87.5
LEMs	86.5	96	74.2–94.4	86.3–99.5	95.7	87.5
Klepel	92.3	92	81.5–97.9	80.8–97.8	92.3	92.2

Note. Sensitivity = probability that a test result is positive when the diagnosis is present; specificity = probability that a test result is negative when the diagnosis is not present; Positive predictor value (PPV) = probability that dyslexia is present when the test is positive; Negative predictor value (NPV) = probability that dyslexia is not present when the test is negative

After calculating the PPV and NPV for our sample group, the PPV and NPV were also calculated when correcting for the estimated prevalence number for dyslexia (10%). For the Flamingo Test, the PPV was 82% and the NPV was 98%. For the LEMs, the PPV was calculated at 71% and the NPV on 98%. The PPV for the Klepel was 56% and the NPV was 99%.

## Discussion

The second study compared the Flamingo Test with two diagnostic tests for adults: the LEMs and the Klepel. Our results showed that the LEMs, the Klepel and the Flamingo Test were all able to distinguish between students with dyslexia and students without dyslexia with large effect sizes. The reading efficiency measure was used to compare the Flamingo Test to the LEMs and the Klepel. Correlational analyses revealed highly significant, positive correlations between all three tests. Moreover, sensitivity, specificity, and PPV for the reading efficiency measure were high for all three tests.

In terms of effect size, the difference between the dyslexic and the control group was largest for LEMs, closely followed by the Klepel, and then the Flamingo Test. In contrast, PPV was the highest for the Flamingo Test, also when considering the 10% prevalence number of dyslexia, being 82% compared to 71% of the LEMS and 56% of the Klepel.

The LEMs and the Klepel test are already widely used as diagnostic instruments in the Netherlands. In combination with the fact that they are easy and quick to administer, this makes them appealing tests to use. The Flamingo Test is as simple and quick to administer, and appears to be equally valid to the LEMs and the Klepel. Interestingly, the correlation between the Flamingo Test and the other two tests is slightly lower than the inter-correlation of LEMs and the Klepel, suggesting that the Flamingo Test may be tapping into a process not assessed by the other two tests, which is arguably due to the

fact that the words are presented in lines of text and form syntactically coherent sentences. As a result, the main message of our findings is that the Flamingo is a valuable addition to the two existing tests, rather than a replacement of one of them. Indeed, when the results of the three tests are combined, we might find optimal assessment.

## General discussion

Dyslexia is the most prevalent learning disability and there is need for more practical assessment instruments specifically designed for adults. For that reason, we present the Flamingo Test, inspired by the French Alouette Test (Bertrand et al. 2010; Cavalli et al. 2018).

Our study supports the expectation that the Flamingo Test can be useful for both research purposes and clinical practice. It measures the same skills as word list reading and pseudo-word list reading, resulting in high correlations with these tests. This is in line with the claim that tests like the Flamingo Test measure word decoding skills and not higher-level text comprehension. The interesting addition of the new test form is that words are presented in coherent lines of text like in normal reading. One of the advantages of this test form is therefore that reading related visual factors, such as visual discomfort or eye movement problems, such as investigated by Wilkins (2002), Jones et al. (2008), can be investigated in a natural way. In addition, the similarity in results for Dutch and French suggests that the test can easily be adapted for other languages.

As for clinical practice, the test can be used as a short, hands-on diagnostic test for dyslexia in adults, as indicated by the high PPV of the Flamingo Test in both experiments. As a single component, reading time was the strongest marker of dyslexia, followed by reading efficiency. This supports the finding of Swanson and Hsieh (2009) and Callens, Tops, and Brysbaert (2012) that a speed rather than an accuracy deficit, is the core impairment in adult dyslexia. The difference between dyslexics and controls in terms of standardised effect size is larger for reading time than for reading efficiency. This was true in our studies and in Cavalli et al. (2018). Alternatively, our data show that a combination of accuracy, reading time, and reading efficiency result in the best assessment.

Comparisons between the Flamingo Test, the LEMs and the Klepel indicate that all three tests largely measure the same construct, visual word decoding, and can be used together to improve assessment. Given that each variable involves some measurement error and unique processes, combining the measures increases diagnostic accuracy. The Flamingo Test has some advantages compared to LEMs and Klepel. The Flamingo tests measures word decoding in a more natural way: (1) reading from left to right is more logical than vertical word list reading, and (2) it contains function words and syntactic information, with an approximately equal balance between function and content words. Additionally, the paradigm introduced by the Alouette Test exploits a clever way of selectively diminishing top-down conceptual and semantic information, while keeping much of the bottom-up processes intact.

A limitation of our study is that we could not control for the potential influence of pronunciation difficulties. We have no reasons to believe that pronunciation was a problem, but in future research it may be interesting to address this factor. Notice that general differences in pronunciation rate may also account for word list reading and pseudo-word list reading. A second possible limitation is that we do not know what effect

the pictures have on performance. This is a feature of the Alouette Test we kept to maximise the similarity with the test. However, to our knowledge no studies exist that systematically investigated the impact of these pictures, by comparing performance in conditions with and without them.

As a suggestion for future research we recommend comparing the Flamingo Test to tests measuring meaningful text reading. For instance, in meaningful text reading, semantic and conceptual top-down information contribute to the process of word recognition. In addition, Brysbaert (2019) reported an average reading aloud rate in typical readers of 183 words per minute, which is close to the reading rate reported for the Alouette Test (Cavalli et al. 2018; see Table 6), but one standard deviation above the reading rate we observed with the Flamingo test (Table 6). A possible factor may be the length or imageability of the words in the text.

## Note

1. Dutch is the official language of The Netherlands; Flemish is the Dutch variant used in the Northern half of Belgium (Flanders). Although both languages are very similar, there are differences in pronunciation and word use, comparable to the differences between American and British English.

## Acknowledgments

We would like to thank all participants for their contribution in this study. We are also very grateful to our student assistant Tineke Quartel for her help with the data collection.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This research was supported by the University of Groningen, Centre for Language and Cognition.

## References

- Bertrand, D., J. Fluss, C. Billard, and J. C. Ziegler. 2010. "Efficacité, sensibilité, spécificité: Comparaison de différents tests de lecture." *L'Année Psychologique* 110 (2): 299–320. doi:10.4074/S000350331000206X.
- Boulanger, V. 2013. *studeren met een functiebeperking*. Masterproef: KU Leuven. [http://www.siho.be/files/nieuwsbrief/Thesis\\_VincentBoulanger.pdf](http://www.siho.be/files/nieuwsbrief/Thesis_VincentBoulanger.pdf)
- Brus, B. T., and M. J. M. Voeten. 1994. *Een-Minuut-Test (EMT)*. Harcourt Test Publishers.
- Brysbaert, M. 2019. "How Many Words Do We Read per Minute? A Review and Meta-analysis of Reading Rate." *Journal of Memory and Language* 109: 104047. doi:10.1016/j.jml.2019.104047.
- Callens, M., W. Tops, and M. Brysbaert. 2012. "Cognitive Profile of Students Who Enter Higher Education with an Indication of Dyslexia." *PloS One* 7 (6). doi:10.1371/journal.pone.0038081.
- Cavalli, E., P. Colé, G. Leloup, F. Poracchia-George, L. Sprenger-Charolles, and A. El Ahmadi. 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. doi:10.1177/0022219417704637.
- De Pessemier, P., and C. Andries. 2009. *GI&Schr. Test voor gevorderd Lezen en Schrijven*. Antwerpen: Garant Uitgevers.
- Dyslexia International. 2017. "Better Training, Better Teaching." Accessed October 23 2019. <https://www.dyslexia-international.org/wp-content/uploads/2016/04/DI-Duke-Report-final-4-29-14.pdf>
- Jenkins J. R., L. S. Fuchs, P. Van den Broek, C. Espin, and S. L. Deno. 2003. "Sources of Individual Differences in Reading Comprehension and Reading Fluency." *Journal of Educational Psychology* 4: 719–729. doi:10.1037/0022-0663.95.4.719.
- Jones, M. W., M. Obregón, M. L. Kelly, and H. P. Branigan. 2008. "Elucidating the Component Processes Involved in Dyslexic and Non-dyslexic Reading Fluency: An Eye-tracking Study." *Cognition* 109 (3): 389–407. doi:10.1016/j.cognition.2008.10.005.
- Leij, A., J. I. van Der, Bekebrede, A. Geudens, K. Schraeyen, G. M. Schijf, H. Garst, H. Willems, V. Janssens, E. Meersschaert, and T. J. Schijf. 2014. *Interactieve Dyslexietest Amsterdam-Antwerpen: Handleiding*. Uithoorn: Muiswerk Educatief.
- Levafrais, P. 1967. *Test de l'alouette: Manuel*. Paris, France: Les éditions du centre de psychologie appliquée.
- Levafrais, P. (2005). "Alouette-R. Paris: Les Éditions du centre de Psychologie Appliquée." <https://www.nuffic.nl/publicaties/vind-een-publicatie/onderwijssysteem-belgie.pdf>
- Lyon, G. R., S. E. Shaywitz, and B. A. Shaywitz. 2003. "Part I Defining Dyslexia, Comorbidity, Teachers' Knowledge of Language and Reading; a Definition of Dyslexia." *Annals of Dyslexia* 53: 1–14. doi:10.1007/s11881-003-0001-9.
- Milne, R. D., T. Nicholson, and M. C. Corballis. 2003. "Lexical Access and Phonological Decoding in Adults Dyslexic Subtypes." *Neuropsychology* 17: 362–368. doi:10.1037/0894-4105.17.3.362.
- Sprenger-Charolles, L., P. Colé, D. Béchennec, and A. Kipffer-Piquard. 2005. "French Normative Data on Reading and Related Skills from EVALEC, a New Computerized Battery of Tests (End Grade 1, Grade 2, Grade 3, and Grade 4)." *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology* 55 (3): 157–186. doi:10.1016/j.erap.2004.11.002.
- Sprenger-Charolles, L., and B. Siegel. 2016. "Prevalence and Reliability of Phonological, Surface, and Mixed Profiles in Dyslexia: A Review of Studies Conducted in Languages Varying in Orthographic Depth." *Scientific Studies of Reading* 15 (6): 498–521. doi:10.1080/10888438.2010.524463.
- Stichting Dyslexie Nederland. 2016. *Diagnose dyslexie. Brochure van de Stichting Dyslexie Nederland. Herziene versie*. Bilthoven: Stichting Dyslexie Nederland.
- Swanson, H. L., and C. J. Hsieh. 2009. "Reading Disabilities in Adults: A Selective Meta-analysis of the Literature." *Review of Educational Research* 79: 1362–1390. doi:10.3102/0034654309350931.
- Tamboer, P., H. Vorst, and P. De Jong. 2017. "Six Factors of Adult Dyslexia Assessed by Cognitive Tests and Self- Report Questions: Very High Predictive Validity." *Research in Developmental Disabilities* 73: 148–163.
- Tops, W. (2012). "Dyslexia in Higher Education: Research in Assessment, Writing Skills, and Metacognition". PhD Diss, Ghent University. Accessed. [http://crr.ugent.be/papers/Doctoraat\\_wim\\_tops.pdf](http://crr.ugent.be/papers/Doctoraat_wim_tops.pdf)
- Tops, W., M. Callens, J. Lammertyn, V. Van Hees, and M. Brysbaert. 2012. "Identifying Students with Dyslexia in Higher Education." *Annals of Dyslexia* 62 (3): 186–203. doi:10.1007/s11881-012-0072-6.
- Tops, W., A. Nouwels, and M. Brysbaert. 2019. "Leestest 1-minuut Onder Nederlandse studenten." *Stem-, Spraak- en Taalpathologie* 24: 1–22. Accessed. <https://sstp.nl/article/view/26110/33076>
- Torgesen, J. K., C. A. Rashotte, and A. A. Alexander. 2001. "Principles of Fluency Instruction in Reading: Relationships with Established Empirical Outcomes." In *Dyslexia, Fluency, and the Brain*, 333–356. Timonium, MD: York Press.
- Van den Bos, K. P., H. C. Lutje Spelberg, A. J. M. Scheepstra, and J. R. de Vries. 1994. *De Klepel: Een test voor de leesvaardigheid van pseudowoorden*. Amsterdam: Pearson.
- Wilkins, A. 2002. "Coloured Overlays and Their Effects on Reading Speed: A Review." *Ophthalmic and Physiological Optics* 22 (5): 448–454. doi:10.1046/j.1475-1313.2002.00079.x.