Identifying students with dyslexia in higher education

Wim Tops – Maaike Callens – Jan Lammertyn – Valérie Van Hees – Marc Brysbaert

Ghent University, Belgium

Corresponding Author Address:

Wim Tops
Department of Experimental Psychology
Ghent University
Henri Dunantlaan 2
B-9000 Gent
Belgium
Tel. +32 9 264 94 31
wim.tops@ugent.be

Abstract

An increasing number of students with dyslexia enter higher education. As a result, there is a growing need for standardized diagnosis. Previous research has suggested that a small number of tests may suffice to reliably assess students with dyslexia, but these studies were based on post hoc discriminant analysis, which tends to overestimate the percentage of systematic variance, and were limited to the English language (and the Anglo-Saxon education system). Therefore, we repeated the research in a non-English language (Dutch) and we selected variables on the basis of a prediction analysis. The results of our study confirm that it is not necessary to administer a wide range of tests to diagnose dyslexia in (young) adults. Three tests sufficed: Word reading, word spelling, and phonological awareness, in line with the proposal that higher education students with dyslexia continue to have specific problems with reading and writing. We also show that a traditional postdiction analysis selects more variables of importance than the prediction analysis. However, these extra variables explain study-specific variance and do not result in more predictive power of the model.

Identifying students with dyslexia in higher education

A growing group of students with dyslexia enter higher education. Guidance protocols and educational arrangements in primary and secondary education are optimized so that students can get past their difficulties and function according to their talents (Tzouveli, Schmidt, Schneider, Symvonis, & Kollias, 2008). As a result, there is an increasing need for standardized diagnosis of young adults. If institutions want to grant program adjustments and other compensatory measures to students with dyslexia, it is necessary to have objective criteria so that the measures are perceived as fair. Students come with a variety of indications of their learning disability and a subgroup may not have been formally examined at all. Also the students themselves often have questions about their strengths and weaknesses and how to best organize their studies. According to Mapou (2008) manifestations of adult learning disabilities may be subtle and remained manageable in secondary education because pupils developed compensatory strategies to mask their limitations. However, these limitations may become more of an issue in higher education because of the much higher study load. Finally, lecturers in higher education are proponents of valid and reliable assessment as well. It makes them more willing to grant special arrangements like extra exam facilities.

Although considerable efforts have been made to develop relevant screening tests and diagnostic tools, there still is much uncertainty about how the screening can be organized efficiently. As a result, there is a lack of standardization across institutes (Johnson, Humphrey, Mellard, Woods, & Swanson, 2010) and there remain many questions about the extent to which findings in English can be extrapolated to other languages (e.g., De Pessemier, & Andries, 2009).

To have more information about the specificity of adult dyslexia, Hatcher, Snowling and Griffiths (2002) compared the cognitive and literacy skills of 23 university students with dyslexia to those of 50 matched control students. All participants completed 17 tasks assessing reading and writing, processing skills, phonological skills, verbal fluency, verbal abilities, non-verbal abilities and self-reported problems in attention and organization. Hatcher et al. (2002) found that the dyslexic students performed worse on all but the two tasks measuring general cognitive abilities, namely the WAIS vocabulary test (Wechsler, 1997) and the Raven test of non-verbal reasoning (Raven, 1938). In order to determine which tests discriminated best between dyslexics or control students, Hatcher et al. (2002) ran a discriminant analysis. About 95% of the students could be classified correctly on the basis of four tests only: spelling, word reading, verbal short term memory and writing speed. Thus, according to Hatcher et al. (2002) a short assessment with four tests is enough to classify adults with dyslexia.

Another interesting study investigating a broad range of skills in students with dyslexia in higher education was the meta-analysis by Swanson and Hsieh (2009) based on 52 published articles. Swanson and Hsieh made a total of 776 comparisons between participants with and without dyslexia. The most important problems of adults with dyslexia were reading, spelling and phonological problems. Furthermore, adults with dyslexia also seemed to have difficulties with verbal long-term memory and arithmetic. Swanson and Hsieh (2009) did not find differences in general intelligence, non-verbal reasoning, cognitive monitoring, visuo-motor skills, verbal and visual perception, social and personal skills, or neuropsychological measures (e.g., EEG).

The findings of Hatcher et al. (2002) and Swanson and Hsieh (2009) suggest that a limited number of tests may suffice to reliably assess dyslexia in higher education. However, the data are far from optimal. The number of students in Hatcher et al. (2002) was small and based on a single English university with rather high entrance requirements. Swanson and Hsieh’s (2009) analysis seems more secure to generalize but is also largely limited to the

English language (only 5% of the studies concerned another language) and is hampered by the fact that data from several different types of studies were aggregated. For instance, some of the perceptual data were based on psychophysical studies of low-level vision and auditory capacities, rather than on measures relevant for educational settings. Indeed, a caveat that must be kept in mind when reading the outcome of a meta-analysis is that it may be comparing apples and oranges given that data from very different sources have been combined (e.g., Sharpe, 1997). Because of the high stakes of correct assessment, one would like to have a broader empirical basis about the number of tests to be administered.

The question about which tests are needed for diagnosis further relates to the theoretical issue about whether students with dyslexia in higher education form a homogeneous group or comprise several subgroups. If different types of dyslexia exist, various criteria must be used. Several authors have indeed suggested the existence of subtypes (e.g., Castles and Coltheart, 1993; Lorusso, Facoetti, & Bakker, 2011), either on the basis of differences in performance on a series of tasks or on the basis of a theoretical analysis of the processes involved in successful reading. For instance, van der Schoot, Licht, Horsley, & Sergeant (2002) made a distinction between *guessing readers* and *spelling readers*, based upon performance criteria initially developed by Bakker (1981). Another typology was proposed by Castles and Coltheart (1993), who started from the assumption that words can be read aloud in two different ways, either by directly converting letters into sounds or by activating the written word in a mental lexicon and deriving the corresponding phonology. On the basis of this analysis, Castles and Coltheart predicted the existence of dyslexics who would have major problems with the reading of new words and non-words (*phonological dyslexia*) and dyslexics who would be mostly impaired on the reading of irregular words (*surface dyslexia*).

The above distinctions strongly focus on impaired reading, whereas dyslexic students in higher education mainly complain about their spelling difficulties. According to Fink (1998), high-performing adults with dyslexia can be divided in two groups. One group has a slow reading rate and difficulty in spelling, whereas the other group has compensated for their impaired reading and shows deficits in spelling only. Indeed, the writing tests were among the most informative in Hatcher et al. (2002; see also Holmes, & Carruthers, 1998; Shaywitz, Fletcher, Holahan, et al., 1999).

Finally, it must be kept in mind that although impaired reading and writing are the center problems of dyslexia, they probably are but the most visible impairments of a larger neuropsychological deficit (Habib, 2000). Neuropsychological research has provided evidence that, although the core deficit in dyslexia is phonological in nature, dyslexic people often present associated problems in a wide range of neurocognitive domains, such as spoken language and auditory temporal processing (e.g., Heath, Hogben, & Clark, 1999), arithmetic (e.g., Perkin, & Croft, 2007), visual attention (e.g., Facoetti, Paganoni, Turatto, Valentina, & Gian, 2000), short term verbal memory and working memory (e.g., Swanson, 1999), time-estimation, automatization, and fine-grained motor skills (e.g., Nicolson, Fawcett, & Dean, 2001). It is not clear to what extent these deficits are a consequence of the difficulty to manipulate speech sounds, or whether they point to a larger variety of cognitive and neuropsychological impairments in individuals with dyslexia, requiring the identification of a more detailed pattern of weaknesses and strengths (Crew & D'Amato, 2010).

For the above reasons we felt that a replication of the Hatcher et al. (2002) study was warranted. First, we wanted to see whether indeed a small number of tests sufficed to reliably diagnose higher education students with dyslexia, or whether an adequate assessment required a much wider battery of tests. Second, we wanted to make sure that Hatcher et al.'s (2002) findings applied to contexts outside British education. These involve a different language (Dutch) as well as differences in entrance requirements. Indeed, alphabetical languages differ in the degree of complexity in the mappings between spelling and sound (Borgwaldt & Hellwig, 2004; Van den Bosch, Content, Daelemans, & de Gelder, 1994) and countries differ in the entrance requirements they impose on students coming to higher education. As a rule, entrance requirements are lower in systems that see selection as part of the curriculum (as happens in Belgium, with more than half of the students not successfully completing their degree) than in systems based on the master-apprentice model (once admitted, the apprentice is expected to complete successfully, as happens in British higher education).

Thirdly, we also wanted to know to what extent it is possible to generalize predictions made by our model (our selection of tests) to future populations. We do not want a model that is only valid for the data upon which it is based; we also want our model to perform well on new data too. An important limitation of the study of Hatcher et al. (2002) is that the results were based on a post hoc discriminant analysis. In such an analysis authors first administer a series of tests and then examine how well the scores allow them to classify the participants. In this type of analysis the more test scores one has the better the prediction becomes, because the test scores are combined in such a way that they optimally account for the pattern of performances observed in the specific group tested. The drawback of this procedure is that it tends to overestimate the percentage of systematic variance, because sample-specific variance (noise) is used for model fitting. As a result, using the same criteria for a new group of participants is likely to result in significantly worse assessment.

In the present study we will select variables based on prediction results rather than “postdiction” results (Gauch, 2002). In such an analysis, one examines to what extent it is possible to use the scores of one group of participants (the training data) to predict the performance of *another* group (the test data). This avoids the problem of model overfitting.

Both in a predictive and post-hoc model the model fit increases over the first few predictors included. However, whereas in a post-hoc model the fit keeps on increasing (because of overfitting), in a predictive model the fit starts to decrease after a few variables have been entered, a phenomenon which Gauch (2002) called “Ockham’s hill”. The reason for the decrease in performance is that after a certain point the model starts to explain noise in the group tested rather than variables systematically affecting performance. Therefore, the number of significant variables in a predictive model usually is lower than the number in a post-hoc analysis. Models with few parameters may be underfitting reality, but models with additional parameters tend to overfit spurious noise (Gauch, 2002).

Following Hatcher et al. (2002), the main aim of the present study was to develop an efficient and optimal diagnostic protocol for the classification of students with dyslexia. By comparing the new data with those of Hatcher et al., we wanted to see to what extent the findings can be generalized to a non-English language. As far as we know, no similar studies have been published.

Method

Participants

Two hundred first-year bachelor students in higher education participated in the study, both non-university college students and university students. They received a small financial compensation for their participation. All had normal or corrected-to normal vision and were native speakers of Dutch. The sample consisted of one hundred students with dyslexia and a control group of hundred students with no known neurological or functional deficiencies.

The students in the dyslexia condition were students who had approached the institute at which they studied for special educational measurements because of dyslexia. The standard procedure in such cases was that the evidence was examined by a specialized remediation service, called *Cursief*, and that the students were retested if needed. So, all students in the dyslexia condition met the formal criteria of dyslexia in accordance with the definition of SDN (Stichting Dyslexie Nederland [Foundation Dyslexia Netherlands], 2008). According to this definition, dyslexia is defined as an impairment characterized by a persistent problem in learning to read and/or write words or in the automatization of reading and writing. The level of reading and/or writing has to be significantly lower than what can be expected based on the educational level and age of the individual. For many students, the deficit also proved resistant to instruction, given that they had taken remedial teaching in primary and secondary education. We ran our tests on the first 100 students who went to the special educational support service in the academic year 2009-2010, who were confirmed as dyslexic, and who were willing to take part in our study (nearly all students did). The mean age of the group was 19 years and 4 months [18 – 23;5 years].

For each student in the dyslexia condition we searched for a control student matched on age, gender, and field of study. For the recruitment, we relied on the student with dyslexia or we contacted study coaches within the different departments. This action was repeated until all participants of the control group were found. The mean age of the control group was 19 years and 11 months [17;9 – 21;6 years]. There was no significant difference in age between the groups of students with and without dyslexia ($t(198) = 0.91$; $p = .36$). There were 46 male and 54 female subjects in each group, from whom 34 students were studying at Ghent University and 66 students took a (non-university) college program.

Table 1:
General Information About the Student Groups With and Without Dyslexia

	Students with dyslexia	**Students without dyslexia**
Number of participants	100	100
Mean age	19 years 4 months *(SD 1.0)*	19 years 11 months *(SD 0.7)*
Gender	46 male	46 male
	54 female	54 female
Number of college students	66	66

College fields of study		
Educational sciences	16	16
Health and behavioral sciences	21	21
Management	9	9
Sciences and Engineering	19	19
Other	1	1
Number of university students	34	34
University fields of study		
Arts and humanities	5	5
Health and behavioral sciences	19	19
Sciences and Engineering	10	10

Tests

In order to determine which tests contribute to an efficient diagnostic protocol for young adults with dyslexia, we administered a large number of tests, covering a wide variety of cognitive domains such as verbal and non-verbal reasoning, processing speed, reading and writing, phonological awareness, executive functions, and memory. All tests had been developed and tested previously for research with adult dyslexics and were mostly adaptations of English tests.[1]

Verbal and non-verbal reasoning. We administered the Dutch version of the Kaufman Adolescent and Adult Intelligence Test, (KAIT) (Dekker, Dekker, & Mulder, 2004), which is a translated and standardized version of the American Kaufman Adolescent and Adult Intelligence. Both fluid IQ (problem solving) and crystallized IQ (reasoning and memory retrieval of stored information) were measured.

[1] Unfortunately, this excluded the Dyslexia Adult Screening Test (DAST), developed in the UK for our target population (Nicolson & Fawcett, 1997), because this test is not available in Dutch. The DAST includes 11 subtests, such as literacy measures, phonological processing, rapid naming, working memory, non-verbal reasoning, verbal fluency, and postural balance.

Vocabulary We used a subtest from the Test for Advanced Reading and Writing, a recently developed test battery for the diagnosis of (young) adults with dyslexia, also called GL&SCHR (De Pessemier, & Andries, 2009). Participants are asked to give definitions of low frequency words like the Dutch equivalents of *anonymous* or *simultaneous*.

Speed of processing. We used a paper and pencil test, called CDT (Dekker, Dekker, & Mulder, 2007), which contains 960 digits from 0 to 9 presented in 16 columns. Students had three minutes to underline as many fours and to cross out as many threes and sevens as possible. The aim of the test was to measure processing speed (the number of correctly processed digits) and accuracy (number of missed and incorrectly processed digits) in a task of selective attention with a considerable task-switching load.

Short-term memory. We used short-term memory span tests for phonological, visual, and lexical items, and a test in which the participants had to reproduce randomly presented series of letters and digits in ascending order (GL&SCHR; De Pessemier, & Andries, 2009).

Phonological awareness We used a spoonerism task (two words were presented auditorily and the first letters had to be switched, e.g. Harry Potter became Parry Hotter) and a reversal task (participants had to judge if two spoken words were reversals or not, e.g. rac-car). Both time and accuracy were taken into account (GL&SCHR; De Pessemier, & Andries, 2009).

Rapid naming. Letters, digits, colors, or objects were individually presented centrally on a computer screen and participants had to name them as rapidly as possible (GL&SCHR;

De Pessemier, & Andries, 2009). The participant determines the pace by pressing the Enter button. Accuracy and speed are measured.

Arithmetic. We used the *Tempo Test Rekenen* [Test for Fast Fact Retrieval] *(TTR;* de Vos, 1992), a standardized test of arithmetic in Flanders. It is designed to examine the rate at which participants mentally perform simple mathematical operations. There are five lists, consisting of addition, subtraction, multiplication, division below 100, and a random sequence of all four operations. Participants are given one minute per list to solve as many problems as possible.

Reading skills.

- Word reading was assessed using the EMT or One Minute Test (Brus, & Voeten, 1991). The test consists of 116 Dutch words of increasing difficulty printed in four columns. The participant has to read aloud as many words as possible in one minute trying to minimize reading errors. The raw score is the number of words read correctly. We also administered an English version of the EMT, namely the One Minute Test or OMT (Kleijnen, & Loerts, 2006). This test is comparable to the Dutch version.
- We administered a Dutch pseudoword reading test called *De Klepel* (van den Bos, Spelberg, Scheepsma, & de Vries, 1999). It contains 116 pseudowords that follow the Dutch grapheme-phoneme correspondence rules. The raw score is the number of pseudowords read correctly within two minutes.
- The silent reading test (Henneman, Kleijnen, & Smits, 2004) consists of silently reading a simple Dutch text and making a précis of the text. The primary aim was to calculate the average number of words silently read per minute.

- Text reading was investigated by a subtest of the GL&SCHR (De Pessemier, & Andries, 2009). Participants had to read out loud a text about fear of failure with increasing difficulty. Both substantial errors and time consuming errors were taken into accounts as well as the total reading time.
- For *text comprehension*, a text was read out by the computer and also presented in printed form. Afterwards, participants had to answer questions about the text (GL&SCHR; De Pessemier, & Andries, 2009).

Writing skills.

- We made a distinction between the spelling of exception words and the knowledge of spelling rules. We used two subtests from the GL&SCHR (De Pessemier, & Andries, 2009). For *Word Spelling* 20 Dutch exception words were used. *Spelling Rules* were tested with a proofreading task consisting of 20 Dutch words or sentences which the participant had to correct. The *Word Spelling* subtest is computer paced. In the first phase, 20 words are read out loud at a rate of one word per two seconds. Students are asked to write down as many of the words as they can. If necessary, they can skip a word and continue with the next one. Afterwards, missed words and words they felt unsure about are repeated as often as required, so that students can correct and supplement their initial spellings without time limits. Because of the strong time constraints in the first phase, tis test can also be used as an indication of *writing speed*.
- Because higher education involves academic language, we also administered an advanced Dutch *Sentence Dictation* test, developed and used at the University of Leuven (Ghesquière, 1998). It consists of 12 paragraphs with exception words and difficult spelling rules (e.g. for the verbs). Also the correct use of capitals and punctuation marks is taken into consideration.

- Given the importance of English in higher education, we also included an English word dictation test. We used *WRAT-III English Word Dictation* (Wilkinson, 1993), a commonly used test for English word spelling.
- All students were asked to write a précis about the text which was used in the silent reading test. In this way, we gathered information about their free writing skills.
- We also administered the subtest *Morphology and Syntax* from the GL&SCHR (De Pessemier, & Andries). The same principles as in the *Spelling Rules* test are applied to sentence writing (e.g., the correct spelling of homophonic verb forms, correct use of punctuation and capitalization).

Procedure

The complete test protocol was administered during two sessions of about three hours each. The protocol was divided into two counterbalanced parts. The order of the tests in part one and two was fixed and chosen to avoid succession of similar tests. There was a break halfway each session. If necessary, students could take additional breaks. Students with dyslexia started with part one or two according to an AB-design. Their control student always started with the same part. All tests were administered individually by three test administrators[2] according to the manual guidelines. Testing occurred in a quiet room with the test administrator seated in front of the student.

[2] The test administrators were the two first authors and a test psychologist. To standardize the administration each administrator read the manuals of tests, had a practice session, and was observed by the others during the first ten sessions.

Results

Differences in test scores between dyslexic students and controls

Table 2 shows the results of the 10 variables with the biggest differences between the students with and without dyslexia. The results are reported as effect sizes. The sign of the d-values is so that positive d-values represent worse performance of the students with dyslexia compared to the control group.

Table 2:
Ten Variables With the Biggest Differences Between the Students With and Without Dyslexia

Word Spelling Dutch	
Weighted score	2.28
Number correct words	2.05
Word Reading Dutch	
Correctly read words	1.97
Total number of read words	1.87
Sentence Dictation Dutch	
Number of errors	2.10
Pseudoword reading Dutch	
Correctly read items	1.59
Total number of read items	1.50
Word spelling English	
Number correct words	1.50
Phonological awareness	
Spoonerisms time	1.42
Word Reading English	
Correctly read words	1.40
Total number of read words	1.36
Phonological awareness	
Reversals time	1.30
Text reading	
Reading time	1.29
Mental calculation	
Mixed operations	1.13

Seven of the ten variables with the biggest differences between the two groups were reading and spelling related. Surprisingly, a big difference was also observed for an arithmetic test, namely rapid fact retrieval of mixed operations. Finally, there were big differences in the time students with dyslexia needed to perform the phonological awareness tasks like the spoonerism and the reversal task.

Table 3:
Five Variables With the Smallest Differences Between the Students With and Without Dyslexia

Symbol memory (KAIT)	0.03
Symbol learning (KAIT)	0.07
Auditory comprehension (KAIT)	0.09
Logical reasoning (KAIT)	0.12
Fluid IQ (KAIT)	0.13

Table 3 shows the five variables with the smallest differences between the students with dyslexia and the students without dyslexia, which were subtests of the KAIT. There was no difference between both groups for symbol memory, symbol learning, auditory comprehension, logical reasoning, and fluid intelligence. Overall, there was a large agreement with the outcomes of the meta-analysis by Swanson and Hsieh (2009).

Postdictive analysis to discriminate dyslexic students from normal readers

For the traditional, postdiction analysis, we used logistic regression (Peng, Lee, & Ingersoll, 2002) to fit a classification model to the data. That way, we could examine how well the participants could be classified as having dyslexia or not. To avoid multicolinearity, we first calculated the variance inflation factor of all variables (tests) mentioned above (n = 29). Two variables were excluded, namely total IQ (because TIQ is the sum of CIQ and FIQ) and RAN letters (because that variable highly correlated with RAN digits). The remaining 27 variables were entered stepwise and removed backwards using the likelihood ratio statistic.

The results of the logistic regression analysis are reported in Table 4, using the log odds, the standard error, the level of significance and the adjusted odds ratios (with 95%

confidence intervals). The best fit was a model with seven variables, including literacy skills (word dictation, sentence dictation, word reading), phonological awareness and verbal reasoning.

Table 4:
Results of the Stepwise Backwards LT Logistic Regression Analysis (postdiction)

		95% CI for Odds Ratio		
	B (*SE*)	**Lower**	**Odds Ratio**	**Upper**
Included				
Intercept	2.53 (*7.37)*			
Dutch sentence dictation	0.09 (*0.04)***	1.02	1.10	1.18
Dutch word spelling	-0.13 (*0.04*)**	0.81	0.88	0.94
PA spoonerisms accuracy	0.33 (*0.15*)*	1.03	1.40	1.89
PA reversals accuracy	-0.41 (*0.15*)**	0.50	0.66	0.89
PA reversals time	0.05 (*0.02*)**	1.01	1.05	1.09
Dutch word reading	-0.10 (*0.03*)**	0.85	0.91	0.97
CIQ	0.12 (*0.05*)*	1.02	1.13	1.25

Note. CIQ = crystallized IQ (KAIT); PA = Phonological awareness;
* *p*<.05; ** *p*<.01.

This model resulted in a correct classification of 93.8% of the participants as being dyslexic or not. This means that when the results of 27 variables were used to classify the participants into the two designated groups (with and without dyslexia) and when multicolinearlity was ruled out, a model with seven variables came out as the best and could assign 93,8% of the participants correctly.

The classification table (Table 5) shows the number of students with and without dyslexia that could be predicted successfully on the basis of the logistic regression analysis (postdiction). Ninety-three out of 100 students with dyslexia were correctly classified (true positives), leaving seven false negatives. In the control group, the data of seven students could not be used because of missing values on one or the other test. Of the remaining participants, 88 students were correctly classified (true negatives) whereas five were false positives, who were classified as having dyslexia without ever having sought help for dyslexia.

Predictive model to discriminate dyslexic students from normal readers

Secondly, we constructed a predictive model that is capable of drawing more general conclusions about the population of higher education students with dyslexia. Before using statistical techniques to select variables for our predictive model, we manually selected promising variables out of the larger pool of available tests. From the tests mentioned earlier, we selected the 10 tests with the largest differences between the groups (Table 2). To avoid multicolinearity problems between highly intercorrelated variables, we inspected the correlations between the variables, as shown in Table 5.

Table 5: Correlation matrix of the ten variables with the largest effect size

	SD	**MC**	**EWS**	**DWSws**	**DWSnc**	**DTR**	**PAspotime**	**PArevtime**	**DWRnc**	**DWRtotal**	**DPRnc**	**DPRtotal**	**EWRnc**	**EWRtotal**
SD	--	.482	.737	.794	.809	.608	.635	.476	.642	.670	.542	.603	.573	.561
MC		--	.484	.516	.528	.258	.424	.403	.538	.543	.524	.533	.484	.495
EWS			--	.709	.708	.483	.584	.433	.618	.637	.581	.626	.698	.664
DWSws				--	.949	.480	.528	.431	.623	.648	.582	.626	.532	.523
DWSnc					--	.479	.528	.414	.614	.639	.555	.609	.544	.526
DTR						--	.506	.414	.434	.484	.472	.563	.470	.442
PAspotime							--	.648	.515	.537	.543	.572	.504	.520
PArevtime								--	.463	.482	.538	.544	.400	.425
DWRnc									--	.995	.741	.744	.701	.790
DWRtotal										--	.754	.767	.713	.795
DPRnc											--	.970	.635	.665
DPRtotal												--	.644	.669
EWRnc													--	.981
EWRtotal														--

Note : SD = Sentence Dictation (number of errors) ; MC = Mental Calculation (mixed operations) ; EWS = English Word Spelling (number of correct answers) ; DWSws = Dutch Word Spelling weighted score : DWSnc = Dutch Word Spelling number of correct responses ; DTR = Dutch Text Reading (reading time) ; PAspotime = Phonological Awareness (spoonerisms time) ; PArevtime = Phonological Awareness (reversals time) ; DWRnc= Dutch Word Reading (number of correct responses) ; DWRtotal = Dutch Word Reading (total number of responses) ; DPRnc = Dutch Pseudoword Reading (number of correct responses) ; DPRtotal = Dutch Word Reading (total number of response)s ; EWRnc= English Word Reading (number of correct responses) ; EWRtotal = English Word reading (total number of responses).

When the correlation between two variables was equal to or higher than .70, the variable with the lowest effect size was excluded. As a result, the following variables were omitted: Dutch sentence dictation, English word dictation, Dutch nonword reading, English word reading and text reading. Next, we added the variables corresponding to the best predictors in the study of Hatcher et al. (2002). These variables were verbal short term memory and writing speed. Together with these two variables of Hatcher et al. (2002), the five following variables were included: Dutch word spelling, Dutch word reading, phonological awareness (both spoonerisms time and reversals time) and mental calculation (mixed operations).

In the next step, we searched for a model that combined maximum predictive power (in terms of prediction accuracy) with the smallest number of predictors. To assure the generalizability of our model we used a resampling method called 10-fold cross validation (Kuhn, 2008). Cross validation is a technique in which the available data set is split up in a training set and a test set. The training set is used to fit a logistic regression model. The predictive power of this model is then tested using the test set. This method assures that data used for fitting the model is never used to test the model. Furthermore, all folds serve several times as training sets, but only once as test set. In our case, the initial dataset was split up in 10 sets (folds) of equal size. This way, the procedure of training and testing a model could be repeated 10 times. On each iteration the model was trained using 9 folds and the predictive power was tested on the remaining fold. Once all iterations were completed, the predictive measures obtained for each of the 10 iterations were averaged.

Nested within the 10 fold cross validation iterations, we also tested models with less predictors than the seven initially selected. After the full logistic regression model with seven predictors was fit, the next model was fit with the same training and test data but with the least important predictor left out. This process was repeated until only one predictor was left in the model. The order by which predictors were left out of the model was based on the z-values of the predictors in the previous model. Based on these values a ranking was made indicating the importance of the variables in the model. Eventually, this resulted in 10 prediction scores for each of the seven variable subset sizes. For each subset size, the predictive score was averaged over the 10 folds.

Predictions were based on a preset probability cut-off value of .50, meaning that cases with a predicted probability larger or equal than .50 were classified as having dyslexia. The predictive power of the model was assessed using prediction accuracy (proportion of true positives and true negatives classifications to the total number of classifications).

Out of all models tested, the model with three predictors, namely Dutch word reading, Dutch word spelling and a phonological awareness task (reversals time) came out as the best. For this subset of predictors, the average prediction accuracy on the test data was 90.9% (95% CI [87.1, 94.8]). Table 6 shows the classification outcomes of the prediction (and the postdiction) model. Ninety-one out of 100 students with dyslexia were correctly classified (true positives), leaving 8 false negatives. In the dyslexic group, the data of one student could not be used because of missing values on two variables. Of the remaining participants, 90 students were correctly classified (true negatives) whereas 10 were false positives (i.e., they were “detected” by the model as dyslexic without any other diagnosis of dyslexia).

Table 6:
Classification Table Based on the Predictive Model with 3 Predictor Variables Versus the Postdictive Model with 7 Predictor Variables.

Reference		**Prediction**	
		Dyslexic	***Non-dyslexic***
Dyslexic	Prediction (postdiction)	91 (93)	8 (7)
Non-	Prediction	10 (5)	90 (88)

dyslexic	(postdiction)

To further illustrate the predictive capabilities of our final model, we fitted a model with the three selected variables on the complete data set (See Table 7). Using the parameters of this model to predict all cases resulted in a sensitivity of 0.97 and a specificity of 0.87. When plotting the sensitivity versus 1 - the specificity (i.e. a ROC curve), the area under the curve equals .97, confirming the predictive quality of the model.

Table 7:
Results of the Logistic Regression Using Dyslexia Group as Dependent Variable and the Three Selected Variables as Predictors.

		95% CI for Odds Ratio		
	Logg Odds (*SE*)	**Lower**	**Odds Ratio**	**Upper**
Included				
Intercept	18.63 (*3.86*)			
Dutch word spelling	-0.14 (*0.03*)**	0.82	0.87**	0.91
Dutch word reading	-0.09 (*0.02*)**	0.87	0.92**	0.96
Reversals time	0.04 (*0.02*)*	1.01	1.04*	1.08

Note. *p < .05; ** p < .01; LRT $\chi 2(3) = 187.71$, p < .001

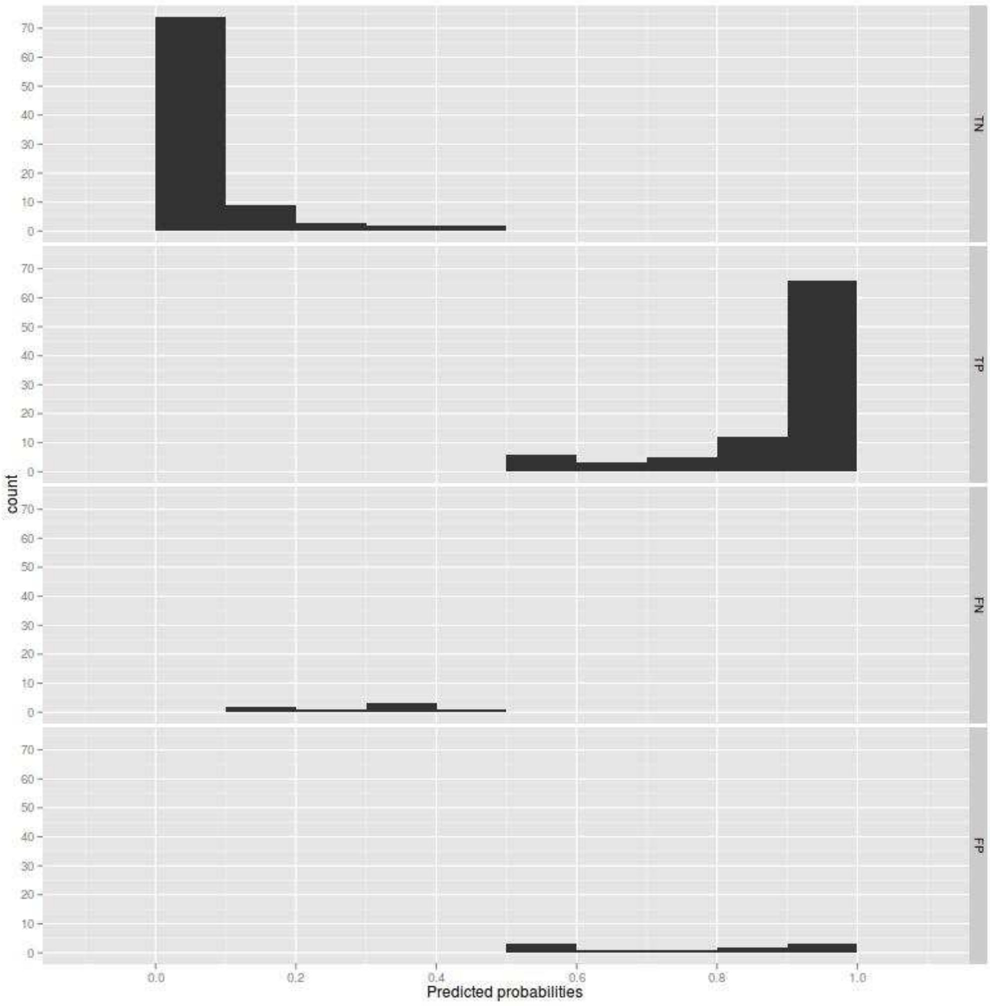

Figure 1. Classification distribution of the true negatives, true positives, false negatives and false positives

To get a better idea of the classification distribution, we calculated the predicted probability scores for each of the participants on the basis of the parameter estimates obtained by this model. Figure 1 shows the distributions of the predicted probability score in the various classification conditions. For the true negatives and true positives, the data are as expected: A clear segregation is present between both groups, with quite extreme scores (either very low or very high). However, the data for the false negatives and the false positives are more surprising, because here we see data all over the range. It is not the case that all false negatives were participants slightly below the .5 criterion (two of the seven participants had a predicted probability below .2, bringing them very much in line with the control group). Similarly, not all false positives were students slightly above the .5 criterion. Five of the 10 participants had probabilities above .8, putting them squarely within the dyslexic range. It will be interesting to see whether these deviating scores are due to random measurement noise (in which case they should not be found in a replication with similar tests), could be the result of an initial misclassification, or point to deviating patterns in a small percentage of individuals (both in those who seek compensation measures for dyslexia and those who do not).

To examine whether some of the misclassifications could be explained on the basis of intelligence, in a final step we added total IQ, crystallized IQ and fluid IQ one-by-one to the model. These variables were of no influence. The same was true for another non-reading test we added, namely the test for processing speed (CDT; Dekker, Dekker, & Mulder, 2007; see above). This did not affect performance either. So, it looks like we cannot improve the power of the predictive model based on the three variables listed in Table 6.

Prediction equation

A further interesting aspect of the analysis is that we can easily calculate a person's predicted probability of dyslexia given the three test scores. All we need to do is to multiply the person's scores with the parameter estimates from Table 6. This calculation involves two steps. First, the log-odds of dyslexia are calculated on the basis of the parameters. This is simply done by entering the test scores into the regression equation. For example, for a person with a score of 94 on the word spelling test, a score of 72 on the word reading test and a score of 116 on the phonological awareness test (reversals), the log odds value is $18.63079 - 0.13644*94 - 0.08514*72 + 0.03973*116 = 4.284$.

Second, from the log_odds, two other values are calculated: the odds of having dyslexia and the probability of having dyslexia. The first value is given by the equation *Odds_dyslexia = exp(log_odds)*. So, for our example the odds of the person having dyslexia equal exp(4.284) = 72.5, meaning that the person is 72.5 times more likely to be classified as dyslexic than not to be. The probability of having dyslexia is calculated with the equation:

P(dyslexia) = exp(log_odds dys of person x) / (1 + exp(log_odds dys of person x))

So, the probability of the person in the example having dyslexia is exp(4.284) / (1 + exp(4.284)) = 0.986.

Discussion

We designed a study to obtain an empirically based diagnostic protocol for young adults with dyslexia. We made a disctinction between postdiction and prediction. For postdiction, we used stepwise (backward) logistic regression analyses, whereas for prediction, we constructed a new model based on tenfold cross validation.

By large our results replicate those of Hatcher et al. (2002). A small number of tests suffice to diagnose students with dyslexia in higher education. These tests involve word reading, word spelling, and phonological awareness (as assessed with the word reversal test, in which participants have to decide whether two words are each other’s reversal). It was expected that students with dyslexia required more time to identify the correct spelling. The fact that two of the three tests were literacy tests, confirms that undergraduate students with dyslexia continue to have serious problems with reading and writing into adulthood, as was also shown in previous research (Everatt, 1997; Erskine, & Seymour, 2005; Felton et al., 1990; Hanley, 1997; Hatcher et al. 2002; Lefly, & Pennington, 1991; Swanson, & Hsieh, 2009). Further interesting is that no extra information seems to be gained from tests involving sentences and coherent text material than from tests involving simple word reading and writing. This considerably simplifies the assessment. In addition, our data show that not only reading and spelling remain deficient in adulthood. Also a more underlying phonological deficit persists in highly functioning adults with dyslexia. We found that adults with dyslexia were not capable to compensate entirely for their phonological deficits.

There is indeed a growing consensus that the primary deficit in dyslexia is phonological in nature (see Vellutino, Fletcher, Snowling, and Scanlon, 2004 for a review). However, there is little agreement about the exact underlying mechanisms making that someone with dyslexia can or cannot bypass poor phonology. Poor phonological awareness is responsible for establishing poor spelling–sound mappings and, further, phonological decoding deficits, which manifest themselves in slow and/or effortful word reading and poor spelling abilities (Kemp, Parilla, & Kirby, 2008).

We further showed that a traditional postdiction analysis lists more variables of importance (seven) than a prediction analysis (three). At first sight, the fact that the former analysis classifies more participants appropriately (94%) than the latter (91%) may be taken as evidence that it is superior. However, this is not true, because the results of the analysis only apply to the specific groups of participants tested. If the equation from Table 4 were used to classify two new groups of 100 participants, it is not expected to result in better performance than the simpler equation from Table 2 (in all likelihood the reverse would be observed).

Postdiction models are data-driven. This means that their results are dataset dependent. Several factors contribute to this. First, the regression weights are adjusted to optimal fit the data at hand. Second, all variance is assumed to be systematic variance that can be predicted. Finally, the category demarkation value is adjusted to optimize the prediction. The impact of these factors becomes stronger the more variables one includes in a postdiction analysis. As it happens, we can further “improve” the postdiction performance by including all 24 variables in the regression analysis. Then we “correctly classify” 95.8% of the participants. However, the expected performance of such a model for new participants is quite bad, because of the instability of the regression weights (which can easily be noticed by comparing the SEs in Table 6 to those in Table 4). The fact that the postdiction model in Table 6 classifies fewer people correctly than maximally possible, is due to the fact that variable selection in logistic regression is based on the best fit of the model and not on the highest classification accuracy. The backward stepwise procedure stops from the moment the elimination of a variable results in a significanly reduction of model fit. It does not stop as soon as the prediction accuracy drops.

This is the most important limitation of postdiction: the lack of generalizability to new (but similar) samples. When postdiction is used instead of prediction, chances are real that in another study the dataset will lead to other variables selected, other weights given to these variables, and another demarkation value for categorisation. None of these elements are interesting for clinical practice. In contrast, our predicton analysis allows clinicians to enter the data of a new student from the same population into the equation of Table 6 and to obtain a valid combined estimate of dyslexia probability (as we showed in the example above). Furthermore, probability values above .50 can safely be interpreted as an indication of dyslexia, whereas lower values can be interpreted as an indication against dyslexia (i.e., the criterion is fixed). Although Hatcher et al. (2002) did not explain very much in detail the method they used to discriminate students with dyslexia from those without reading and spelling impairments, we can assume that it was postdictive. They were able to correctly categorize 95% of the students using four variables, close to what we observed in our postdiction analysis.

Do the results of our prediction analysis imply that there are only three differences between dyslexics and controls? The study of Gauch (2002) suggests that this is not necessarily true. Simulations indicate that in data with a relatively large degree of noise, the best predictive model generally contains fewer variables than the model used to generate the data. Therefore, the results of the present study apply to the prediction of dyslexia in educational settings, but not to the causes of dyslexia. We show that we can reduce the initial battery of 24 variables to 3 without loss of predictive power. This is important information for educational settings (which can considerably simplify their procedures), but it does not mean that students with dyslexia differ on these variables only. In reality, there may be more variables and the measures we used may also turn out not to be the most optimal. This is particularly true for the variables that were highly intercorrelated (Appendix A). For instance, the high correlation between Dutch sentence dictation and Dutch word spelling ($r = .8$, close to the reliability of the tests) means that both tests can be swapped without much loss of predictive power. Such a finding is not problematic for pragmatic assessment purposes, but obviously is less optimal when it comes to discerning the "true" deficits underlying dyslexia. This relates to the question of how learning disabilities have to be diagnosed and what purpose is served by such a diagnosis.

Another important message of our study is that more tests did not improve our classification quality. Although the postdiction analysis "suggested" that more students could be classified correctly if the full set of 27 variables was used, the prediction analysis clearly showed that such impression was an illusion, due to the fact that sample-specific variance was being modeled by the ever increasing number of predictors. This is a strong argument against the pressure clinicians sometimes experience from the authorities to add more and more tests to their assessment battery in order to achieve "better" assessment. Of the 24 extra tests we added, none made a significant contribution.

At the same time, our data (together with those of Hatcher et al., 2002) form a firm basis for further research into valid predictors of dyslexia for higher education. The fact that our 24 extra variables did not improve the classification power, does not exclude the possibility that some variable may do so. Each study involves choices (which can be criticized). The same is true for the present study. Maybe we overlooked a critical variable or we operationalized it badly? Researchers thinking they have a better candidate can now severely reduce the number of tests they have to run in addition to one they are interested in. Moreover, we think we have brought for the first time in dyslexia research the expertise together that is needed to run prediction analyses (e.g., to make sure that the new variable is indeed making a difference) and to use the outcome in order to turn multiple test data into easy-to-interpret dyslexia probabilities (or odds). This is likely to be of profit to dyslexia researchers, clinical practitioners, students, and university authorities alike. [3]

A final intriguing aspect of our study is that although the vast majority of students fall within the remit of the phonological deficit hypothesis of dyslexia, some 5% have clearly deviating results. These are nearly equally divided over false positives and false negatives (see the two lower panels of Figure 1). It will be interesting to examine whether these exceptions are statistical errors, or whether indeed a small segment of dyslexics has a deficiency that consistently deviates from the general pattern. This is a typical example where case studies can augment the findings of a group analysis. It is also a reminder that student-centered facilities are more than the blind application of a series of tests (however useful they are for the initial assessment).

[3] The remediation service in Ghent is already using the outcome of our study (Table 6) for the initial assessment of new students with a suspicion of dyslexia.

Acknowledgments

This study was made possible by an Odysseus Grant awarded by the Government of Flanders to MB. The authors thank Valérie Van Hees and Charlotte De Lange from Cursief for their help in the study and the recruitment of participants. They also thank Joke Lauwers for her assistance in testing the participants, and two anonymous reviewers for their helpful suggestions.

References

Alonso-Búa, B., Díaz, F., & Ferraces, M.J., (2006). The contribution of AERPs (MMN and LDN) to studying temporal versus linguistic processing deficits in children with learning difficulties. *International Journal of Psychophysiology, 59,* 159-167. doi:10.1016/j.ijpsycho.2005.03.020

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed., text revision). Washington,DC: Author.

Bakker, D.J. (1981). A set of brains for learning to read. In K.C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 56-71). Rowley, MA: Newbury House.

Borgwaldt, S., & Hellwig, F. (2004). Onset entropy matters: Cross-linguistic sound-spelling analyses and empirical validation. *Australian Journal of Psychology, 56,* 106-107.

Brus, B., & Voeten, M. (1991). Eén-minuut-test vorm A en B, schoolvorderingstest voor de technische leesvaardigheid bestemd voor groep 4 tot en met 8 van het b asisonderwijs. Verantwoording en handleiding. Lisse, The Netherlands: Swets & Zeitlinger.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use and calculation of confidence intervals based on central and noncentral distributions. *Educational and Psychological Measurement, 61,* 530-572. doi: 10.1177/0013164401614002

Dekker, R., Dekker, P.H., & Mulder, J.L., (2004). *Kaufman Adolescent and Adult Intelligence Test – Nederlandstalige Versie:* Handleiding. PITS: Leiden.

Dekker, R., Mulder, J.L. & Dekker, P.H. (2007) De ontwikkeling van vijf nieuwe Nederlandstalige tests. Leiden, The Netherlands: Pits.

De Pessemier, P., & Andries, C. (2009). Test voor Gevorderd Lezen en Schrijven. Antwerpen, Belgium –Apeldoorn, The Netherlands : Garant.

Desoete, A. (2009). Metacognitive prediction and evaluation skill and methematical learning in third-grade students. *Educational Research and Evaluation,* 15, 435-446.

de Vos, T. (1992). *Tempo Test Rekenen*. Amsterdam, The Netherlands: Pearson Education.

Dykman, R.A., & Ackerman, P.T. (1991). Attention-deficit disorder and specific reading disability: separate but often overlapping disorders. *Journal of learning disabilities,* 24, 95-103.

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self- regulation and co-regulation. *European Psychologist, 13,* 277-287. doi: 10.1027/1016- 9040.13.4.277

Everatt, J., Bradshaw, M.F., Hibbard, P.B., et al. (1997). Visual processing deficits and dyslexia. *Perception, 26,* 1330-1330.

Facoetti, A., Paganoni, P., Turatto, M., Valentina, M., & Gian, G.M. (2000). Visual-spatial attention in developmental dyslexia. *Cortex, 36,* 109-123. doi: 10.1016/S0010-9452(08)70840-2

Fawcett, A.J., & Nicolson, R.I. (1998). *Dyslexia Adult Screening Test.* London, United Kingdom: The Psychological Corporation.

Ferrari, M. (2009). A comparative assessment of the cognitive functioning of adults with childhood histories of learning disability and adults with noncognitive disorders. *Journal of developmental and physical disabilities, 21,* 355-368. doi: 10.1007/s10882-009-9148-6

Fink, R.P. (1998). Literacy development in successful men and women with dyslexia. *Annals of Dyslexia, 48,* 311-346. doi: 10.1007/s11881-998-0014-5

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-235). Hillsdale, NJ: Erlbaum.

Flavell, J.H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, 34, 906–911. doi: 10.1037/0003-066X.34.10.906

Fletcher, J.M., Francis, D.J., Morris, R.D., Lyon, G.R. (2005). Evidence-based assessment of learning disabilities in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34,* 506-522. doi: 10.1207/s15374424jccp3403_7

Gauch, H. G. 2002. *Scientific Method in Practice*. Cambridge, United Kingdom: Cambridge University Press.

Germano, E., Gagliona, A., & Curatolo, P. (2010). Comorbidity of ADHD and dyslexia. *Developmental neuropsychology,* 35, 475-493. doi: 10.1080/875656412010494748

Ghesquière, P. (1998). *Algemene toets gevorderde spelling van het Nederlands (AT-GSN), verantwoording en handleiding. Rapport van een specialisatiejaar: onderzoek AT-GSN-dictee* [General Test in Dutch Advanced Spelling (AT-GSN), justification and manual. Report of a specialisation year: study on AT-GSN dictation]. Unpublished manuscript, University of Leuven, Belgium.

Habib, M. (2000). The neurological basis of developmental dyslexia: An overview and working hypothesis. *Brain, 123,* 2373-2399. doi: 10.1093/brain/123.12.2373

Hatcher, J., Snowling, M.J., Griffiths, Y.M. (2002). Cognitive assessment of dyslexic students in higher education. *British Journal of Educational Psychology, 72,* 199-133. doi: 10.1348/000709902158801

Heath, S.M., Hogben, J.H., Clark, C.D. (1999). Auditory temporal processing in disabled readers with and without oral language delay. *Journal of Child Psychology and Psychiatry, 40,* 637-647. doi: 10.1017/S0021963099003947

Henneman, K., Kleijnen, R., Smits, A. (2004). *Protocol Dyslexie Voortgezet Onderwijs : Deel 2 - Signalering, diagnose en begeleiding*. KPC Groep, Expertisecentrum Nederlands, Werkverband Opleidingen Speciaal Onderwijs.

Holmes, V.M., & Carruthers, J. (1998). The relation between reading and spelling in skilled adult readers. *Journal of Memory and Language, 39,* 264-289. doi: 10.1006/jmla.1998.2583

Johnson, E.S., Humphrey, M., Mellard, D.F., Woods, K., & Swanson, H.L. (2010). Cognitive processing deficits and students with selective learning disabilities: a selective meta-analysis of the literature. *Learning disabilities Quarterly, 33,* 3-18.

Jones, M.W., Obregón, M., Kelly, M.L, & Branigan, H.P. (2008). Elucidating the component processes involved in dyslexic and non-dyslexic reading fluency: an eye-tracking study. *Cognition, 109,* 389-407. doi: 10.1016/j.cognition.2008.10.005

Kemp, N., Parrila, R., & Kirby, J. R. (2009). Phonological and Orthographic Spelling in High- functioning Adult Dyslexics. *Dyslexia, 15,* 105-128. doi: 10.1002/dys.364

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of statistical software, 28,* 1-26.

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition, 63,* 315-334. doi: 10.1016/S0010-0277(97)00005-X

Lefly, D.L., & Pennington, B.F. (1991). Spelling-errors and reading fluency in compensated adult dyslexics. *Annals of dyslexia, 41,* 143-162.

Livingstone, M. S., Rosen, G. D., Drislane, F. W., & Galaburda, A. M. (1991). Physiological and anatomical evidence for a magnocellular deficit in developmental dyslexia. *Proceedings of the National Academy of Sciences, 88,* 7943–7947.

Lorusso, M.L., Facoetti, A., Bakker, D.J. (2011). Neuropsychological Treatment of Dyslexia: Does Type of Treatment Matter? *Journal of Learning Disabilities, 44,* 136-149. DOI: 10.1177/0022219410391186

Mapou, R.L. (2008). Comprehensive evaluation of adults with learning disabilities. In L.E. Wolf, H.E. Schreiber, & J. Wasserstein (Eds.), *Adult Learning Disorders. Contemporary Issues.* (pp. 248-273). Hove, NY: Psychology Press, Taylor & Francis Group.

McLoughlin, D., Leather, C., & Stringer, P. (2002). *The adult dyslexic. Interventions and outcomes.* London, United Kingdom: Whurr Publishers.

Nicolson, R.I., Fawcett, A.J., & Dean, P. (2001). Developmental dyslexia: the cereballar deficit hypothesis. *Trends in Neurosciences, 24*, 508-511.

Nicolson, R.I. & Fawcett, J. A. (1997). Development of objective procedures for screening and assessment of dyslexic students in higher education. *Journal of Research in Reading, 20,* 77-83. doi: 10.1111/1467-9817.00022

Perkin, G., & Croft, T. (2007). The dyslexic student and mathematics in higher education. *Dyslexia, 13*, 193-210. doi: 10.1002/dys.334

Raven, J.C. (1938). *Progressive Matrices: a perceptual test of intelligence*. Londen, United Kingdom: H.K. Lewis.

Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review, 17*, 881-901.

Shaywitz, S.E., Fletcher, J.M., Holahan, J.M., Shneider, A.E., Marchione, K.E., Stuebing, K.K., Francis, D.J., Pugh, K.R., & Shaywitz, B.A. (1999). Persistence of dyslexia: The Connecticut Longitudinal Study at Adolescence. *Pediatrics, 104*, 1351-1359. doi: 10.1542/peds.104.6.1351

Smythe, I. & Everatt, J. (2001). A new dyslexia checklist for adults. In I, Smythe (Ed.), *The Dyslexia Handbook 2001*. (pp. 73-78). British Dyslexia Association, United Kingdom.

Snowling, M. J. (2000). Dyslexia (2[e] ed.). Oxford, United Kingdom: Blackwell.

Stichting Dyslexie Nederland (2008). *Diagnose dyslexie. Brochure van de Stichting Dyslexie Nederland. Herziene versie* [Diagnosis of dyslexie. Brochure of the Dyslexia Foundation of the Netherlands. Revision]. Bilthoven, The Netherlands: Stichting Dyslexie Nederland.

Swanson, H.L. (2009). Reading comprehension and working memory in learning-disabled readers: Is the phonological loop more important than the executive system? *Journal of Experimental Psychology, 72,* 1-31.

Swanson, H.L., & Hsieh, C.J. (2009). Reading Disabilities in Adults: A Selective Meta-Analysis of the Literature. *Review of educational Research, 79,* 1362-1390. doi: 10.1348/000712603321661859

Trainin, G., & Swanson, H.L. (2005). Cognition, metacognition, and achievement of college students with learning disabilities. *Learning Disability Quarterly*, 28, 261-272. DOI: 10.2307/4126965

Tzouveli, P., Schmidt, A., Schneider, M., Symvonis, A., & Kollias, S. (2008). Adaptive reading assistance for the inclusion of students with dyslexia: The AGENT-DYSL approach. *8th IEEE International Conference on Advanced Learning Technologies,* 167-171. doi: 10.1109/ICALT.2008.236

van den Bos, K., Spelberg, H., Scheepsma, A., & de Vries, J. (1999). De klepel vorm A en B, een test voor leesvaardigheid van pseudowoorden. Verantwoording, handleiding, diagnostiek en behandeling . Lisse, The Netherlands: Swets & Zeitlinger.

Van den Bosch, A., Content A., Daelemans, W., & De Gelder, B. (1994). Measuring the complexity of Writing Systems. *Journal of Quantitative Linguistics, 1,* 178-188.

Van der Sluis, S., de Jong, P.F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading and arithmetic. *Intelligence, 35,* 427-449. doi:10.1016/j.intell.2006.09.001

Vanderswalmen, R., Vrijders, J., & Desoete, A. (2010). Metacognition and spelling performance in college students. In A. Efklides, & P. Misailidi (eds.), *Trends and Prospects in Metacognition Research.* (pp. 367-394). New York, NY: Springer.

Vellutino, F.R., Fletcher, J.M., Snowling, M.J. & Scanlon, D.M. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*, 2-40. doi: 10.1046/j.0021-9630.2003.00305.x

Wechsler, D. (1997). *WAIS-III administration and scoring manual.* San Antonio, TX: The Psychological Corporation.

Wilkinson, G.S. (1993). *Wide Range Achievement Test*. Wilmington, DE: Jastak Assessment.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, *91*, 415-438. doi: 10.1037/0022-0663.91.3.415