Dyslexia

From diagnoses to theory

Tamboer, P.

Publication date
2019

Document Version
Final published version

License
Other

Dyslexia

From Diagnoses to Theory

Peter Tamboer

Dyslexia

From diagnoses to theory

Peter Tamboer

Dyslexia

From diagnoses to theory

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties ingestelde commissie,

in het openbaar te verdedigen in de Agnietenkapel

op donderdag 21 maart 2019 te 14.00 uur

door Peter Tamboer

geboren te Amsterdam

Promotiecommissie:

Promotor(es):	Prof. Dr. F. J. Oort Universiteit van Amsterdam Prof. Dr. P. F. de Jong Universiteit van Amsterdam
Copromotor(es):	Dr. H. S. Scholte Universiteit van Amsterdam
Overige leden:	Prof. Dr. L. A. van der Ark Universiteit van Amsterdam Prof. Dr. M. W. van der Molen Universiteit van Amsterdam Dr. W. Tops Rijksuniversiteit Groningen Dr. P. J. F. Snellings Universiteit van Amsterdam Prof. Dr. P. Ghesquière KU Leuven

Faculteit der Maatschappij- en Gedragswetenschappen

Admissions of ignorance and temporary mystification are vital to good science.

Richard Dawkins

Contents

1

General Introduction

Methods for investigating dyslexia

At present, after about half a century of scientific research, there is still no consensus about what causes dyslexia, and the number of studies about dyslexia is steadily increasing every year. One of the main topics in dyslexia research is how to distinguish people with dyslexia from those without, while underlying causes are still unknown. In this thesis, various methods for diagnosing dyslexia will be investigated.

Although underlying causes are unknown, much has been learned about dyslexia in the past decades, and many theories of dyslexia have been proposed. It is widely accepted that dyslexia affects learning to read and write, and that typical difficulties persist into adulthood. The most prominent theory of dyslexia for many years was the phonological deficit theory, which posits that dyslexia is caused by impairments in phonological information processing, probably caused by problems in the access to or fuzziness of phonological representations of spoken words (e.g. Vellutio et al., 2004; Shaywitz & Shaywitz, 2005; Snowling & Hulme, 2005). Recently, attention was drawn to visual and attentional deficits (e.g. Bosse, Tainturier & Valdois, 2007), although the relevance regarding causes of dyslexia is debated (e.g. White et al., 2006). Other theories are, for instance, the cerebellar or automaticity theory (e.g. Nicolson & Fawcett, 1990; 1995) and the magnocellular theory of dyslexia (Stein and Walsh, 1997).

One thing that these theories have in common is that they describe dyslexia as a disorder that results from one single deficit, which explains all other symptoms. However, several studies showed that various symptoms tend to be characteristic only for subgroups of people with dyslexia (Bosse, Tainturier & Valdois, 2007; Ramus et al., 2003). Moreover, none of the known symptoms has yet been found to be characteristic for all people with dyslexia. Alternatively, Pennington (2006) proposed the so-called multiple deficit theory (MDT) of dyslexia. This theory assumes that dyslexia cannot be explained by a single deficit but, instead, is the outcome of multiple causes, either genetic or environmental. Some causes, so-called risk factors, increase the likelihood of dyslexia, whereas protective factors decrease the likelihood of dyslexia. Dyslexia is assumed to be the outcome (poor reading) of the interplay of many risk and protective factors. According to this view, it is possible that not all people with dyslexia share one common symptom.

An important issue for all theories is that so many symptoms have been described for subgroups of people with dyslexia, that it is hard to assess which of these symptoms contribute to the development of dyslexia and which symptoms are merely the result of dyslexia. Below, a list of symptoms is provided, which is probably still incomplete. But it shows the overwhelming number of findings in regard to dyslexia research. A theoretical review of all known symptoms is in preparation (Tamboer & Vorst, in prep.).

1. Reading (International Dyslexia Association, 2002)
2. Spelling (Blomert, 2005)
3. Auditory perception and temporal processing (Tallal, 1980; Tallal, Miller & Fitch, 1993; McArthur & Bishop, 2001; Schulte-Körne & Bruder, 2010)
4. Phonological representations (Serniclaes et al., 2004; Blomert, Mitterer & Paffen, 2004)
5. Allophonic speech perception (Serniclaes et al., 2001; Serniclaes et al., 2004; Bogliotti et al., 2008; Dufor et al., 2009).
6. Access of phonological representations (Ramus & Szenkovits, 2008)
7. Rapid naming (Wolf & Bowers, 1999; Wolf, Bowers & Biddle, 2000; Vaessen, Gerretsen & Blomert, 2009)
8. Morphological awareness (Siegel, 2008)
9. Procedural learning (Nicolson et al., 2010)
10. Working memory (Savage, Lavers & Pillay, 2007; de Jong, 1998; Kibby et al., 2004)
11. Magnocellular processing (Stein & Walsh, 1997; Chase et al., 2007)
12. Automatic anchoring (Ahissar, 2007)
13. Perceptual noise exclusion (Sperling, Lu & Manis, 2004; Sperling et al., 2005; Sperling et al., 2006)
14. Motion perception (Benassi et al., 2010)
15. Reversal errors (Orton, 1925; Lachmann & Geyer, 2003; Willows & Terepocki, 1993)
16. Coordination (e.g. Lachmann, 2002; Lachmann & Geyer, 2003)
17. Thinking in pictures (Davis & Braun, 1995)
18. Eye-movements, binocular coordination (Kirkby et al., 2008)
19. Redundant bilateral advantage (Henderson, Barca & Ellis, 2007)

20. Encoding of letter position within words (Ziegler et al., 2008; Friedmann, Dotan & Rahamim, 2010)
21. Word-length effect (Marinus & de Jong, 2010).
22. Visual crowding (e.g. Lorusso et al., 2004; Martelli et al., 2009)
23. Shifting covert attention (Buchholz & Davies, 2005)
24. Switching attention in dichotic listening (Kershner, 2016)
25. Task switching (Poljac et al., 2010)
26. Visual attention span (e.g. Bosse, Tainturier & Valdois, 2007; Romani et al., 2011)
27. Multisensory spatial attention (e.g. Facoetti et al., 2010)
28. Reading music notes (Hébert et al., 2007; Miles, Westcombe & Ditchfield, 2008; Tunmer & Greaney, 2010)

In conclusion, there are many reported symptoms, but there is still no theory of dyslexia that clearly explains causal relations between these symptoms. Currently, it is unclear whether there is a single deficit that can explain all symptoms or whether there are numerous independent causes that may result in dyslexia. The result of these different views is that definitions of dyslexia vary to a large extent. According to the MDT of Pennington (2006), the outcome of many possible causes is poor reading; thus, poor reading or reading disorder is the same as dyslexia. On the other hand, a single deficit theory advances the hypothesis that dyslexia does not always result in poor reading. For example, highly intelligent people with dyslexia may overcome most difficulties, such as learning to read, whereas dyslexia still affects other abilities.

The aim of this thesis is to determine the best method for diagnosing dyslexia, while considering that there are numerous symptoms and various theoretical perspectives and definitions. Regarding the investigation of diagnostic methods, it is better not to take a position within the theoretical debate. For instance, many researchers still consider dyslexia as a disorder of phonological abilities. If I would agree with this point of view, it might result in a selection of solely phonological tests without also considering tests of other cognitive abilities, which could possibly show strong predictive power. Therefore, in regard to the studies of this thesis, the investigation of diagnostic methods will be approached from a fully exploratory perspective. I did not embrace any specific hypothesis or theory about dyslexia beforehand. Moreover, when preferring one theory over the other, the danger exists that a personal bias distracts from clear methodological reasoning.

The three core themes of an exploratory methodological approach are measurement, theory, and prediction. In general, each scientific discipline starts with measurement: the documentation of observations. The analysis of measurements can lead to specific hypotheses. These hypotheses can be tested, for instance, by making predictions that follow from the hypotheses. In the scientific field of dyslexia, these processes are the same. In this thesis, I will also turn the order around and investigate whether the study of prediction can result in theoretical insights. Below, I will summarise some issues about the development of theories, the measurement of symptoms, and predictions of dyslexia, and how these are interrelated.

How to find a cause of dyslexia? (Development of theory)

Although the cause or causes of dyslexia are unknown, many theories have been postulated. However, how were such theories developed? Regarding dyslexia, two types of research methods can be distinguished: the behavioural approach and the brain approach.

The behavioural approach consists of collecting data (usually with tests) about the symptoms of dyslexia and trying to determine what each person with dyslexia has in common. The main finding of this approach is that there is no single symptom of dyslexia that uniquely distinguishes all people with dyslexia from those without. This is also found for the symptom of reduced phonological awareness (Bosse, Tainturier & Valdois, 2007), which is usually reported as a causal underpinning of dyslexia. Moreover, when people with dyslexia do not share one single characteristic, dyslexia cannot be reliably predicted; therefore, any conclusions about its causes should be drawn with great caution. Although it is widely accepted that dyslexia affects learning to read and write, some children with high intelligence and good schooling, for instance, might overcome such reading and writing difficulties. Whether these children should be regarded as dyslexic, varies between theories of dyslexia.

The brain approach consists of researching the brain itself by using imaging techniques, such as magnetic resonance imaging (MRI) or functional MRI (fMRI). However, the brain is large and complex, while almost no complex cognitive function is known to be positioned in one single area of the brain. Furthermore, it should be accounted for that learning to read and write and any symptoms of dyslexia influence the development of the brain. The main finding of this approach is that many areas may be involved in dyslexia, while it is impossible to pinpoint a single area as the basis of a cause. It is like a large fire: the larger the fire becomes, the harder it is to find the origin of inflammation.

How to determine all symptoms of dyslexia? (Measurement)

In general, behavioural symptoms can be measured with tests and brain imaging techniques. Determining all symptoms of dyslexia is easier than to finding a cause; however, it is not as easy as one might think. At first glance, it seems to be a simple job: recruit some people with and without dyslexia and measure their performances on various tasks or compare the anatomy and functionality of their brains. As mentioned above, at least 28 symptoms have been reported so far in the literature, and the number of areas in the brain that are assumed to be related to dyslexia is increasing with each new MRI study.

One problem is that we are never sure whether all people from a sample with dyslexia really have dyslexia. Some people may be included, who suffer from difficulties that are typical for dyslexia but stem from something else, such as low intelligence, bad schooling, or Attention Deficit Hyperactivity Disorder (ADHD). As a result, it is hard to distinguish for instance spelling errors that are typical of dyslexia and spelling errors that are typical for all people who experience ‘normal’ language difficulties.

A second problem is that some people with dyslexia are overlooked because they do not suffer or do not suffer severely enough from typical difficulties. This can be the result of highly developed compensation strategies because of high intelligence, good schooling, or individual training. As a result, we cannot be sure whether all symptoms are measured. It may well be that highly compensating people with dyslexia suffer from other lesser-known symptoms that are overlooked when the focus is only on well-known symptoms.

In numerous studies, these problems result in different decisions regarding inclusion and exclusion of participants from the dyslexia sample, such as: participants who also suffer from ADHD, participants with low intelligence or low schooling, participants with high intelligence and compensate for well-known difficulties, left-handed participants, participants with a history of depression or anxiety, and so on. Thus, if samples are qualitatively different, how can we evaluate the reported symptoms between studies? How do we determine which of the reported symptoms represent core difficulties of dyslexia? Moreover, how do we determine whether the excluded participants have dyslexia or not?

Difficulties of interpretation that result from sample issues are common in almost all studies of dyslexia. Additional problems arise when measuring the brain with MRI. If many people with dyslexia show poor performances on spelling tests, it is clear that spelling is somehow related to dyslexia. However, what does it mean when brain scans of people with dyslexia show less or more than an average amount of grey matter in one specific area? A large volume of grey matter can mean that someone is good at something, which resulted in the growth of grey matter. On the other hand, it may also point to a weakness: grey matter will increase as a result of extra training in doing something difficult. Or it may be the case that somebody has a specific weakness but uses another ability to compensate for this weakness resulting in more grey matter in an area that is not related to the original weakness.

How should dyslexia be diagnosed? (Prediction)

At first glance, it may also seem easy to determine how dyslexia should be diagnosed. Determine the main consequences of dyslexia, determine whether a person suffers from these consequences, and then categorise that person as dyslexic or not: this is the usual process of diagnosis, both in research and in clinical settings. Many studies report high classification accuracy of various test batteries of about 90%. However, this still means that many people cannot reliably be diagnosed.

On a more general note, though, any reported classification accuracy is unreliable because we do not know what dyslexia is. Classification accuracies are based on the comparison between the prediction of a test battery and what is known beforehand about the participants from, for instance, diagnostic centres that specialise in diagnosing dyslexia. Yet, how can they know for sure who has dyslexia and who does not?

To make matters even more complicated, the best test batteries and criteria of dyslexia also differ across countries because each language has characteristics that are problematic for people with dyslexia in different ways. Some people mainly suffer from phonological difficulties, while others mainly suffer from visual attention difficulties. Some researchers have even proposed various subtypes of dyslexia. So, how does one diagnose dyslexia when there are subtypes?

Conclusion

The main conclusion is that much is still unknown about the cause or causes of dyslexia and that existing procedures for diagnosing dyslexia cannot distinguish all people with and without dyslexia with complete certainty. This is a problem for psychologists who must diagnose children, for parents who must spend a lot of money for diagnoses and additional help, but also for scientific studies when people with dyslexia are selected. Looking for a cause of dyslexia is searching for a needle in a haystack without knowing the best strategy. During this search, therefore, we find that all kinds of symptoms are interrelated in an endless complexity.

The best way to deal with these problems is to accept that finding the cause or causes of dyslexia might be far away, not to have any preferences for theories or research methods beforehand, and to remain persistent in finding better ways of diagnosing dyslexia in the endless complexity that arises from the literature.

Accountability and goals of this thesis

As long as the cause or causes of dyslexia remain unknown, diagnoses of dyslexia can never be perfect. Various complications of diagnosing dyslexia were mentioned in the previous section. Nevertheless, preceding the studies of this thesis, I believe that there are various reasons to investigate whether the methods for diagnosing dyslexia can be improved.

One reason is that the knowledge of dyslexia is constantly growing, and with that, the methods for diagnosing dyslexia should be improved as well. A second reason is that diagnostic methods are usually based on theories and rarely the other way around. Studying diagnostic methods might contribute to theories of dyslexia and may ultimately lead to better diagnoses. A third reason is that we found various objections against existing statistical procedures of diagnosing dyslexia. These objections and issues are summarised below.

Diagnosing dyslexia with tests

In all countries that acknowledge dyslexia as a serious issue, diagnoses are based on test batteries. Although some results seem to be quite good, there are some issues that need a discussion.

Firstly, the choice of tests differs between most test batteries across countries, which does not allow for the unification into a single method based on tests. The choice of tests depends on typical difficulties of each language, which vary to a large extent (Brunswick, 2010; Ziegler et al., 2010). For instance, many typical difficulties depend on the orthographic depth of a language. In shallow orthographies (e.g. Spanish, Italian, and Finnish), there is a one-to-one relationship between most of its graphemes and phonemes, while in deep orthographies (e.g. English and French), there is a less direct correspondence between letters and sounds. Also within one language, the choice in tests differs sometimes, usually resulting from different theoretical views of the researchers or psychologists, or because different tests are available for measuring the same ability. In this thesis, I will try to test as many symptoms of dyslexia as possible, if necessary, by creating new tests.

Secondly, the inclusion and exclusion criteria in samples differ between studies. For instance, in some studies, people with both dyslexia and ADHD are excluded from further analyses, while in other studies these people are included. Comorbidity between dyslexia and ADHD is widely accepted (Willcutt et al., 2010), which might compromise diagnoses of dyslexia. Comparisons between studies are severely hindered by different choices about exclusion or exclusion criteria. In the studies of this thesis, we will exclude only participants with a foreign background, or participants who clearly did not take the tests seriously.

Thirdly, an extremely difficult issue in dyslexia research is the way researchers deal with intelligence. Low intelligence of participants compromises the performances on tests of dyslexia. Participants with low intelligence are usually excluded, but cut-off scores vary between studies. In this thesis, it is not necessary to exclude participants based on low scores on intelligence measures because only students will participate.

Fourthly, the reliability of the tests used for diagnoses is usually high, but validity is often not reported, or at least questionable. In many studies, it is assumed that a certain test measures something without reporting convergent or divergent validity. For example, it is generally assumed that a sound deletion test measures phonological abilities. This is probably true, but the possibility cannot be excluded that this test also measures something else that is unknown. This is not criticism, but it only points out the argument that it is almost impossible to make a test that exclusively measures only one single ability. Moreover, if other abilities are involved, these may bias the interpretation of the results of the test. In this thesis, we will investigate underlying dimensions of tests, using explorative and confirmative factor analyses on a very large battery of tests, an approach that has hardly been done before.

Fifthly, when using a test for diagnosing dyslexia, one must decide about how low a performance should be to assume that a participant has dyslexia. These cut-off scores vary between studies. Sometimes a cut-off score is determined by one standard deviation below the average of people without dyslexia, but sometimes by one-and-a-half standard deviation below average. When batteries of tests are used, one also must decide for each test separately. Furthermore, it must be decided how the results of all tests determine dyslexia. Something like, ‘at least three out of five tests scored below one standard deviation below average’, is a completely arbitrary decision. Moreover, it is very questionable that a person who performed just above the cut-off score is assumed not to have dyslexia. In this thesis, we will not make decisions about cut-off scores, but we will use prediction analyses that assign weights to tests based on iterative analytic procedures.

Finally, in relation to all issues mentioned above is the issue of a criterion of dyslexia. How can we know whether a diagnosis based on a test battery is accurate without being certain that the criterion used during development of the battery is completely reliable? We can test a diagnostic test battery on a group of people with and without dyslexia of whom we are sure (after various exclusion criteria), thus, people with severe dyslexia and high performing controls. However, we still do not know how accurate this battery is when applied to other ‘moderately performing’ people. It cannot be avoided that reasoning goes in circles. Related to the issue of a criterion is the issue of prevalence. For instance, we want to know the prevalence of dyslexia in a country and apply a test battery on a random population sample. Choices about cut-off scores often implicitly depend on assumptions about the prevalence that already exist. Estimates of prevalence vary between studies to a large extent, ranging from three to more than ten percent. This is a large range and affects all steps of any diagnostic method. In this thesis, we will investigate the relation between diagnostic accuracy and various criteria of dyslexia, both stringent and flexible criteria.

Diagnosing dyslexia with questionnaires

The use of self-report questionnaires as a tool for diagnosing dyslexia does not need a lengthy summary in this introduction, since it has hardly been investigated. Only a few very short questionnaires are reported in the literature (e.g. Lefly & Pennington, 2000; Mortimore & Crozier, 2006; Schulte-Körne, Deimel & Remschmidt, 1997; Wolff & Lundberg, 2003). Considering the amount of research performed on dyslexia in the past decades, this is highly surprising.

Even more surprising is the scepticism of some scientists regarding the use of questionnaires. The main argument against the use of questionnaires is that socially desirable answers would compromise both construct and predictive validity. However, this is not supported by scientific studies.

Diagnosing dyslexia with MRI

About the use of neuro-imaging techniques such as MRI for diagnosing dyslexia we can be short in this introduction as well. This has never been studied. Here, this lack of research has nothing to do with scepticism but with the fact that neuro-imaging research is still in its infancy.

A few studies in the past decade have shown various differences between people with and without dyslexia regarding anatomical structures of grey matter and its functionality (e.g.

Pernet et al., 2009a; Richlan, Kronbichler, & Wimmer, 2009). Although the significance of these results is questionable, as I argued above, it still is remarkable that so many areas across the brain were found to be related with dyslexia. This seems to be consistent though with the numerous symptoms and theories accompanying dyslexia. Because of these widespread findings, the question is justifiable whether some differences in the brain may be common for all people with dyslexia.

Conclusion and goals

In conclusion, there are numerous methodological issues that raise questions about the reliability of diagnoses of dyslexia using test batteries on general populations in various languages. Nevertheless, I think that these issues can be addressed with methodological tools and that diagnostic methods can be improved. We will investigate various cut-off scores and criteria of dyslexia (see chapters 2 and 5), factor scores (see chapters 3 and 5), self-report questions (see chapters 2, 4 and 5), and MRI-findings (see chapters 6 and 7) and compare all results with existing theories of dyslexia. I expect that studying predictive and construct validity of tests, questionnaires and brain-imaging results may not only provide more insights in diagnosing dyslexia but also in understanding the theories regarding causal underpinnings of dyslexia.

2

Identifying dyslexia in adults: An iterative method using the predictive value of item scores and self-report questions

Methods for identifying dyslexia in adults vary widely between studies. Researchers have to decide how many tests to use, which tests are considered to be the most reliable, and how to determine cut-off scores. The aim of this study was to develop an objective and powerful method for diagnosing dyslexia. We took various methodological measures, most of which are new compared to previous methods. We used a large sample of Dutch first-year psychology students, we considered several options for exclusion and inclusion criteria, we collected as many cognitive tests as possible, we used six independent sources of biographical information for a criterion of dyslexia, we compared the predictive power of discriminant analyses and logistic regression analyses, we used both sum scores and item scores as predictor variables, we used self-report questions as predictor variables, and we retested the reliability of predictions with repeated prediction analyses using an adjusted criterion. We were able to identify 74 dyslexic and 369 non-dyslexic students. For 37 students, various predictions were too inconsistent for a final classification. The most reliable predictions were acquired with item scores and self-report questions. The main conclusion is that it is possible to identify dyslexia with a high reliability, although the exact nature of dyslexia is still unknown. We therefore believe that this study yielded valuable information for future methods of identifying dyslexia in Dutch as well as in other languages, and that this would be beneficial for comparing studies across countries.

Keywords: Adult dyslexia, classification, criterion, cross validation, item scores, self-report

Based on: Tamboer, P., Vorst, H. C. M., & Oort, F. J. (2014). Identifying dyslexia in adults: An iterative method using the predictive value of item scores and self-report questions. *Annals of Dyslexia*, *64*(1), 34-56.

Introduction

Dyslexia is considered to be a developmental disorder that persists into adulthood, with estimates of prevalence varying between 3 and 18 %. Dyslexic children are usually relatively reliably identified with poor reading and spelling abilities despite adequate intelligence, motivation or schooling. However, the identification of dyslexic adults is usually problematic because, for instance, school records of reading and spelling abilities are not always available. The lack of objective methods for identifying adult dyslexia forces researchers to make arbitrary decisions about the process of selecting dyslexic and non-dyslexic participants. Any attempt to find an objective and standardized selection method for dyslexia is complicated by the development of an overwhelming quantity of theories and hypotheses of dyslexia in decades of research. In a recent study, Ramus and Ahissar (2012) discussed how the existence of so many theories complicates the interpretation of poor and normal performances of dyslexics in a broad range of tasks. They argued that any theory of dyslexia (e.g. phonological, visual, attentional) runs the risk of overgeneralizing poor performances – predicting poor performance in many more situations than observed – while normal performances are overlooked. For example, poor performances may be confounded by minor intellectual abilities or other factors. Normal performances on the other hand, may wrongly be interpreted as an indication of being not dyslexic, while the effects of dyslexia may just as well be too small to be detected in tasks that are too ‘easy’, especially in the case of people who are highly intelligent or people who underwent extra training during their school days or in college.

As a result of various interpretations of performances, researchers who need to select dyslexic and non-dyslexic adults are faced with various problematic issues, such as: which tests to use, which tests are considered to be the most reliable predictors of dyslexia, and how to determine cut-off scores. In a recent attempt to arrive at an objective method that deals with these issues, Tops et al. (2012) used a predictive analysis and found that the combination of only three tests (word reading, word spelling, and phonological awareness) sufficed for the identification of dyslexic and non-dyslexic students in higher education with more than 90% accuracy. However, some issues in the process of identifying dyslexic adults remained unresolved in this study as well.

One issue is which exclusion and inclusion criteria should be applied to selected groups of dyslexics and non-dyslexics. For example, significant differences in general intelligence between groups are usually not accepted. The result is that in many studies, some participants – who are suspected of being dyslexic – are removed from further analyses because of low general intelligence. Aside from the disadvantage of the fact that dyslexic people with low intelligence are seldomly selected for studies of dyslexia, it is still unclear to what extent measurements of intelligence are affected by dyslexia in general.

A second issue is how to determine an objective criterion of dyslexia that can be used in a prediction analysis. The assumption in, for example, a discriminant analysis is that criterion groups exist beyond any doubt, so that a new sample of people can be classified based on the behaviour of these criterion groups in certain tasks. However, as long as we do not know exactly what causes dyslexia, we can never be absolutely sure of any individual being dyslexic or not. Even for very clear cases of dyslexia – individuals showing all known symptoms – alternative explanations such as low intelligence can never be ruled out completely. Realizing this, a criterion of dyslexia can never be determined with absolute certainty. An additional danger is that a criterion of dyslexia is based on the same kind of tasks as the tasks which are used in the predictive analysis. In the study of Tops et al., the dyslexic students were previously examined by a specialized remediation service and retested if necessary. However, the authors did not report specific details about the tests which were used for a criterion of dyslexia, or how cut-off scores were determined.

A third issue is which tests should be administered. Tops et al. found that a combination of three different tests resulted in the most reliable prediction, which is consistent with the view that dyslexia is characterized by multiple deficits. However, some other cognitive impairments which have been described for dyslexics were not incorporated in the study of Tops et al. The number of tests needed for the best diagnosis of dyslexia remains unclear, but it seems safe to say that researchers who administer only a few tests for selection purposes run the risk of selecting false positives or false negatives because any single poor or normal performance does not necessarily mean that someone is dyslexic or not.

A fourth issue is how to determine which tests are the most important ones. This is a difficult issue for researchers who make use of an extended battery of tests. With the existence of so many theories of dyslexia, the risk exists that choices between tests are influenced by theoretical insights favoured by researchers. For example, one researcher might consider a poor performance on a phonological test to be the most important indication of dyslexia and a poor performance on a visual test less important, while another researcher might decide the other way around. This is not a reproach of researchers because as long as we do not know exactly what causes dyslexia, any decision can only be subjective. This subjectivity is strengthened by the fact that the performances of dyslexics on specific language tests vary widely between languages and across samples.

A fifth issue, related to the previous one, is how to determine cut-off scores for various tasks. This issue is complicated by the difficulty of interpreting poor and normal performances, as discussed by Ramus and Ahissar (2012). An additional complication is that it is unknown whether dyslexia is a distinct trait – just like handedness – or a disorder that varies in severity, representing the left-hand side of a normal distribution. Various theoretical views of dyslexia contribute to this issue. Some genetic and neurobiological findings support the view of one underlying deficit that develops differently, depending on the nature of schooling and training during childhood (e.g. Richlan, Kronbichler, & Wimmer, 2011). However, just as much support can be collected for the existence of subtypes of dyslexia (Bosse, Tainturier, & Valdois, 2007; Castles & Coltheart, 1993), which in turn can be distinct traits or traits that are normally distributed. Thus, to determine cut-off scores for cognitive measurements is a puzzling assignment for researchers leading to arbitrary choices.

A sixth issue is that the choice of which analyses to use is arbitrary. As an alternative to determining cut-off scores, discriminant analyses (DA) or logistic regression analyses (LR) can be used. As also explained by Tops et al. (2012), the general danger of prediction analyses is that the more predictors are used – which would be preferable for the identification of dyslexics – and the smaller the samples are, so a model fit will be based more and more on sample-specific variance. Considering this danger, the choice between DA and LR is not easy because the assumptions between these analyses differ considerably. For example, DA assumes normal distributions of explanatory data while LR does not. Pohar, Blas and Turk (2004) evaluated both methods in various situations and concluded that DA should be preferred in some cases and LR in others.

In this study, we propose a method of diagnosing adult dyslexia which takes into account the above issues as much as possible. Thus, the aim was to identify dyslexic and non-dyslexic students with a high reliability and without having a preference for any theory of dyslexia beforehand. We took nine methodological measures, with most of them being new compared to previous methods. To address the first issue, we carefully considered several options for exclusion and inclusion criteria (e.g. intelligence, age, psychological health, fraudulent behaviour on tests). To address the second issue, we used six independent sources of biographical information for a criterion. We also varied the criterion by applying different decision rules for the classification into the groups of dyslexics and non-dyslexics. In this way, we were able to compare predictions based on severe dyslexia with predictions based on moderate dyslexia. To address the third issue, we collected as many cognitive tests as possible, which together covered the entire spectrum of cognitive deficits in dyslexia. To address the fourth issue, we used two statistical techniques – DA and LR – which do not depend on subjective preferences for certain theories of dyslexia. To address the fifth issue, we not only used sum scores as predictor variables in the analyses, but also single item scores. As discussed by Ramus and Ahissar (2012), we assumed that dyslexic people do not necessarily make more mistakes than others, but that they make specific types of mistakes. These specific mistakes could be overlooked in prediction analyses using sum scores as predictors, but not, maybe, when item scores are used as predictors. To address the sixth issue, we did not choose between DA and LR beforehand. Instead, we compared the predictive power of the analyses with each other, by evaluating their assumptions for each separate prediction. Finally, to enhance the reliability of the predictions as much as possible, we took three additional measures. One additional measure was that we used a large sample of 495 Dutch first-year psychology students. With such a large sample, it was possible to use many predictors in the analyses, which were needed when using single items as predictors. A second additional measure was that we used self-report questions as predictor variables in the analyses. We assumed that this could enhance the reliability of predictions, because a self-report of very specific language difficulties might be less vulnerable to the influences of intelligence, schooling and compensation strategies. A third additional measure was that we repeated predictions on different criterion groups. We reasoned that when consistency between separate predictions is high for a participant, the reliability of the classification would be enhanced.

Method

Participants and sample characteristics

For this study, data of 1110 first-year psychology students from the University of Amsterdam were used. This group is the summation of two separate groups of first-year students of 2009 (548 students) and 2010 (562 students). We excluded 615 students from any further analyses, more than half of the original group. At first sight this seems strange and may raise questions about the objectivity of procedures. Therefore, we will explain the exact procedures that resulted in this large exclusion of students (see ‘Procedure’).

The remaining group of 495 students consisted of 125 males and 370 females. Mean age was 19.7 (1.5) years with a minimum age of 17.8 years, and a maximum age of 25.9 years. There were 414 right-handed students (84.5%), 53 left-handed students (10.8%), 23 ambidextrous students (4.7%), and 5 students without handedness data.

Tasks

Questionnaires

The *SES Questionnaire* (socio-economic status) is a questionnaire that is administered at the University of Amsterdam every year. In this study, we used information from this questionnaire about age, gender, nationality, language background, history of health problems, type of education, and school grades.

The *Handedness & Orientation Questionnaire* – designed for this study and partly based on an existing questionnaire (van Strien, 1992) – aims to acquire all information about hand preference and orientation. We administered this questionnaire for two reasons. Years ago, many researchers assumed that dyslexia and left-handedness might be correlated. However, there is only a limited amount of empirical data available regarding this correlation, with inconsistent results (Saviour et al., 2009; van Strien, 1992). Nevertheless, at present, it is still common to report percentages of left-handedness in studies of dyslexia. The questionnaire consists of two parts. In the first part, questions are asked about hand preference in general (right-handed, left-handed, ambidextrous), hand preference in writing (right, left, both, right but as a child a preference for left), hand preferences of family members, and about 22 specific preferences (five response categories) such as: 'Which hand do you use for brushing your teeth?' The second part consists of ten statements about orientation between left and right in general (yes, no, a little) such as: 'As a child it was hard for me to determine the difference between left and right'.

The *Dyslexia Questionnaire* – designed for this study – aims to acquire general information that can be used to create a criterion of dyslexia. There are three parts: (1) a full self-report of dyslexia; (2) former test results from school or of an official institute of dyslexia; and (3) information about dyslexic family members (biological mother, father, sisters, or brothers).

The *Language Preference Questionnaire* – designed for this study – aims to acquire information about general language difficulties. The term 'dyslexia' is not mentioned in this questionnaire. There are six subscales, each consisting of ten statements with seven response categories. Here, we give an example of each subscale. Reading: ‘Every week I read a book’; Speaking: ‘I like to talk fast’; Writing: ‘I keep a diary’; Mental representations: ‘I remember faces easily’; Memory: ‘I remember phone numbers easily’; Foreign languages: ‘I don't like the fact that most study books are written in English’.

The *Communication Questionnaire* (Kramer & Vorst, 2007) aims to acquire information about specific language difficulties. The term ‘dyslexia’ is not mentioned in this questionnaire. The questionnaire is designed according to a 7×5×4 facet design. There are seven subscales representing different aspects of language, each consisting of 20 statements with seven response categories. Here, we give an example of each subscale. Reading: ‘Sometimes I skip a letter, which results in reading a different word’; Writing: ‘Sometimes I forget to write down a syllable’; Speaking: ‘While speaking, I sometimes exchange similar words’; Listening: ‘I hear a story exactly like someone tells it’; Copying: ‘When I copy out a text, I sometimes exchange letters with similar sounds’; Dictating: ‘I make mistakes in a dictation, because I don't hear the correct sounds’; Reading aloud: ‘When reading aloud, I sometimes repeat a part of the text’. All statements can also be categorized into five subscales representing sounds, letters, words, sentences and text. A leading thought during the creation of statements was that four typical mistakes might distinguish dyslexics from others: skipping (forgetting), adding, changing, and exchanging.

Intelligence tests

Six tests were based on Guilford's Structure of Intellect Model: (1) *Vocabulary* (cognition of semantic units: knowing and understanding words and concepts); (2) *Verbal Analogies* (cognition of meaningful verbal relations); (3) *Conclusions* (cognition of meaningful symbolic relations, or the ability to understand and structure difficult situations and the evaluation of semantic implications); (4) *Numeric Progressions* (cognition of symbolic systems in progressions of numbers); (5) *Speed of Calculation* (the ability to assess simple symbolic rules); and (6) *Hidden Figures* (spatial intelligence). For general (non-verbal) intelligence we used *Advanced Progressive Matrices Set 2* (Raven, Court, & Raven, 1979).

Short-term memory test

The *Short-Term Memory Test* – designed for this study – aims to measure the capacity of short-term memory. We used the concept of digit span: the number of digits a person can retain and recall. There are four subtests: *numbers* and *letters*, both *forward* and *backward*. Each subtest consists of 24 series: 6 of 4, 6 of 5, 6 of 6, and 6 of 7 items for the subtests *numbers* and *letters forward*, and 6 of 3, 6 of 4, 6 of 5, and 6 of 6 items for the subtests *numbers* and *letters backward*. The numbers and letters are presented one by one, for one second each on a computer screen. The participants have to retype these numbers and letters after the last one of a series has been presented. About half of all series consists of some typical difficulties for dyslexics, either phonological, visual, or both. For example, a typical phonological confusion is between the numbers seven and nine which resemble each other phonologically in Dutch (zeven/negen). Typical visual confusions are between the numbers six and nine and the letters [m] and [w]. The letters [p], [d], and [b] resemble each other phonologically as well as visually.

Specific language tests

Eleven (six auditory and five visual) language tests were designed for this study. We incorporated many typical difficulties for dyslexics in these tests. For the six auditory language tests, instructions and all test items were read out aloud in a well-trained female voice and can be heard through headphones and read on a computer screen at the same time. Item responses have to be typed within a limited time and can be started while listening. For all tests, high scores represent good performances and low scores represent poor performances.

Dutch Dictation (auditory) aims to measure spelling abilities in the Dutch language (10 sentences, maximum score 10×4=40).

English Dictation (auditory) aims to measure spelling abilities in the English language (10 sentences, maximum score 10×2=20). It can be assumed that Dutch students are familiar with the ordinary English words we used.

Missing Letters (auditory) also aims to measure spelling abilities in the Dutch language (10 sentences, maximum score 10×2=20), but in a slightly different way. For each sentence, two words are repeated while these words are shown on the computer screen with a few letters left out of the word.

Pseudowords (auditory) aims to measure spelling abilities of pseudowords – non-words that sound like real words (30 words, maximum score 30). Participants have to decide whether the non-words they hear are spelled correctly on the computer screen. Usually, pseudowords are admitted the other way around through participants reading aloud the words themselves. The reason for changing this was that it would be practically impossible to get all students in private sessions for this way of testing.

Sound Deletion (auditory) aims to measure phonological abilities (20 words, maximum score 20). Participants have to decide whether the difficult Dutch words they hear are pronounced correctly, and if not, which letter is missing or has been added (there is a choice between three words). For example the existing word ‘fietsenstalling’, which means bicycle shed, is read out as ‘fiestenstalling’. The possible answers are: ‘fietsentalling’, ‘fiestensalling’ and ‘fiestenstalling’.

Spoonerisms (auditory) (see also Hazan et al., 2009) also aims to measure phonological abilities (20 words, maximum score 20). A spoonerism is a word that consists of two existing smaller words and that still consists of two small existing words when the first letters of both small words are exchanged. For example, participants hear the word ‘kolen-schop’ which has to be altered to ‘scholen-kop’.

Incorrect Spelling (visual) is the third test in our study that aims to measure spelling abilities in the Dutch language, again in a different way (40 words, maximum score 40). All words are flashed on a computer screen for 50ms. Participants have to decide whether the words are spelled correctly or not.

Dutch–English Rhyme Words (visual) aims to measure the ability to recognize similar-sounding nouns in Dutch and English (40 words, maximum score 40). Dutch–English word pairs are shown on a computer screen with the Dutch words on the right. Participants have to decide whether the words rhyme with each other or not. Typical confusion may arise in this test because the non-rhyming items have the same vowels, such as ‘Deep-Reep’.

Letter Order (visual) aims to measure the ability to read words as a whole (20 sentences, maximum score 20×2=40; time limit of 5 min). Theoretical hypotheses about reading words as a whole are described in the dual route model of reading (for an extended description see de Groot et al., 1994). The idea for this test comes from a well-known text: ‘Aoccdrnig to rscheearch at Cmabrigde uinervtisy, it deosn't mttaer waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteres are at the rghit pclae. The rset can be a tatol mses and you can sitll raed it wouthit a porbelm. Tihs is bcuseae we do not raed ervey lteter by it slef but the wrod as a wlohe’. We created 20 sentences based on the same principle: the order of the letters of the words was changed, apart from the first and last letters. The words in the sentences are more difficult towards the end of the test. The sentences have to be typed in with all words correctly spelled. There are no words that consist of typical dyslexic spelling difficulties.

Counting Letters (visual) aims to measure the effects of global reading (two sentences, maximum score 6+8=14). The idea for this test is based on a well-known language joke.

Count the number of times that the letter [f] appears in the following sentence: ‘Finished files are the result of years of scientific study combined with the experience of years’. Many people only see the [f] three times. It has been suggested that the [f] in [of] is overlooked because it sounds more like a [v]. Another suggestion is that [of] is overlooked completely as a result of global reading. We created two Dutch sentences based on the same principle. In the sentence ‘Het deftige hondje van de man en de vrouw drinkt water uit de kraan’ participants have to count the letter [d] (6). In the sentence ‘Met de neus en de mond is het niet moeilijk en zelfs gemakkelijk een liedje te neuriën’ participants have to count the letter [n] (8).

Mirror Reading (visual) aims to measure the ability to read in mirror image (20 words, maximum score 20; 10 sentences, maximum score 10×4=40). On a computer screen, the words and sentences are presented in mirror image, so that they would appear normally written when held in front of a mirror. Thus, the letters are also shown as a mirror image!

Reliability

The reliability of all tests and questionnaires was determined based on a sample of 495 students (see ‘Procedure’). Table 1 shows the values of Cronbach's alpha. For *Letter Order* and *Mirror Reading* we did not calculate the reliability because the items were varying in difficulty and there was a time limit. For *Counting Letters*, only one score was acquired. The majority of students performed on all tests and questionnaires. A group of 26 students did not have a score on *Spoonerisms* because they did not understand the instructions of the test.

Procedure

Collection of data

All data were collected at the University of Amsterdam during five sessions of three hours each, in which tests and questionnaires of various studies were administered. These sessions took place on midweek evenings with one or two weeks between each session. All students were obligated to participate because these sessions were part of the first-year study programme. All students were informed about the general nature of the tests and questionnaires in advance according to a standard protocol. However, regarding the tests and questionnaires related to dyslexia neither the students nor the surveillants knew about the true purpose. Afterwards, the students received a more detailed debriefing. Anonymity was guaranteed by the standard protocol of the University of Amsterdam. All students had up to three weeks after their debriefing to request that their test results were not used.

Table 1

Reliabilities of (sub-) tests and (scales of) questionnaires (Cronbach`s alpha)

(Sub-) test / (Scales of) questionnaire	N items	N subjects	Cronbach's α
Language Preference Questionnaire	60	489	0.721
Handedness & Orientation Questionnaire	22	489	0.964
Communication Questionnaire	141	494	0.984
QC subscale reading	20	494	0.892
QC subscale writing	20	494	0.902
QC subscale speaking	20	494	0.905
QC subscale listening	20	494	0.910
QC subscale copying	20	494	0.942
QC subscale dictating	21	494	0.939
QC subscale reading aloud	20	494	0.942
Short-Term Memory Numbers Forwards	24	487	0.767
Short-Term Memory Numbers Backwards	24	482	0.804
Short-Term Memory Numbers Total	48	480	0.863
Short-Term Memory Letters Forwards	24	492	0.773
Short-Term Memory Letters Backwards	24	485	0.815
Short-Term Memory Letters Total	48	485	0.873
Dutch Dictation	10	494	0.670
English Dictation	10	493	0.425
Missing Letters	10	490	0.627
Pseudowords	30	489	0.518
Sound Deletion	20	489	0.847
Spoonerisms	20	469	0.906
Incorrect Spelling Test	40	493	0.584
Dutch–English Rhyme Words	40	493	0.788
Letter Order	20	492	x
Counting Letters	24	482	x
Mirror Reading Words	20	492	x
Mirror Reading Sentences	10	492	x

x: Reliability could not be calculated
Counting Letters: only one score was acquired
Letter Order and *Mirror Reading*: items varying in difficulty, time limit for the test

We made sure that our tests and questionnaires were spread out as much as possible, ensuring that they were not administered in one session but over all sessions and positioned between tests of other studies. We were also able to make sure that the order of tests was about the same in both years. However, we had no influence on the positioning of the standard intelligence tests. In Appendix A, 39 test results can be found for the 2009 and 2010 groups. A few significant differences were found, but not more than could be expected when performing many separate *t*-tests. However, most intelligence tests were performed better by the first year which might be due to differences in the positioning of these tests (different sessions or different positions within one session).

Exclusion of participants

Of 1110 participants, 615 were excluded from any further analyses, leaving 495 students. We performed this process of exclusion in three successive steps.

In the first step, we excluded 361 students on three grounds: health, nationality, and missing data, leaving 749 students. A small group of students had a history of moderate to severe head trauma or severe general health problems (3%). A large group of students was not ‘completely Dutch’ (25%). This group consisted of foreign students, students who had not lived in the Netherlands for their entire youth, and students with one foreign parent (only when they indicated Dutch not as their primary mother language). Also excluded were students who did not participate in a large number of tests and questionnaires (10%). There was quite some overlap between the last two groups, which makes sense because, for example, most foreign students did not participate in the language tests.

In the second step, we excluded 70 students aged between 26 and 54 years, leaving 679 students aged from 17 to 25 years. We excluded older students because we wanted our sample to be as homogeneous as possible, so that effects resulting from differences in long-term experience with language could be prevented.

In the third step, we excluded 184 students who we suspected were not serious in a few of the tests or questionnaires which were crucial for our study (27% of 679). This seems a lot. However, we realized that the test sessions – which lasted three hours each – required much concentration and could have been exhausting for some students, leading to rushing through tests or questionnaires. After excluding a small group of very obvious cases of fraudulent behaviour, we applied an objective procedure to detect as much fraudulent behaviour as possible. First, we applied several statistic procedures that can detect answer patterns. Second, we excluded students whose registered test durations were so short that it would be theoretically impossible to even read the question in such a short time. This was, in many cases, probably due to test instructions that were not well understood. Finally, we decided that students with at least four fraudulent results in separate tests or questionnaires were excluded. In the remaining cases, we only removed the scores of single tests or questionnaires.

We did not exclude participants based on intelligence scores for two reasons. First, in a sample of students, it can be assumed that general intelligence is within the normal range. Second, it is not clear to what extent specific measurements of intelligence correlate with dyslexia.

Results

In the first paragraph, we describe how we determined the criterion groups of dyslexics (D-group) and non-dyslexics (ND-group). In the second paragraph, we describe the procedure of identifying dyslexic and non-dyslexic students. In the third paragraph, we describe various characteristics of the D-group and the ND-group by comparing these groups with the original criterion groups and by comparing test results between groups.

Criterion and classification

Ideally, when using prediction analyses, a perfect criterion is required. This criterion separates two small groups of people so that other subjects can be classified based on the behaviour of the criterion groups in certain tests. To determine a perfect criterion of dyslexia is of course impossible when the exact nature of dyslexia is unknown. The only thing we can do is try to approximate a perfect criterion, which will always be based on arbitrary decisions.

In studies of dyslexia in the Netherlands, dyslexics are often selected based on the so called ‘dyslexieverklaring’, an official document that can be acquired by children after an extended testing procedure that is performed by a specialized educational psychologist. Although this document is widely accepted to be a reliable criterion of dyslexia in the Netherlands, it has three drawbacks. First, it cannot be ruled out that – where there is some doubt – some psychologists will decide to give a child a ‘dyslexieverklaring’. Second, the reliability of this document suffers from various issues mentioned in the introduction. For instance, different psychologists use qualitatively different tests, a different number of tests and/or subjective standards in the choice of cut-off scores. Third, the use of the ‘dyslexieverklaring’ as a criterion of dyslexia for further research is not appropriate when the same kind of tests are investigated as were used for the criterion. In other words: the predictive power of a test can never objectively be investigated when a similar test was used in a criterion.

In this study, we took three approaches for improving the reliability and objectivity of a criterion. First, alongside the ‘dyslexieverklaring’, we used five other independent sources of biographical information as indications of dyslexia. Second, these other sources of information were not based on test results, so that we would not compare predictions and a criterion that were based on the same information. And third, we used different criterion groups, starting with strict criterion groups and then adjusting them based on information from various predictions. The advantage of this was that we could compare predictions based on the behaviour of severely impaired dyslexics and ‘super controls’ with predictions based on the behaviour of groups which included mildly impaired dyslexics and poor-performing non-dyslexics. The use of additional information from the predictions itself for the criterion is of course a statistical pitfall, especially because we reasoned that we should be careful in using the same information twice. However, we also reasoned that this would be justified as long as not too many false positives and false negatives were detected. An additional advantage is that the reliability of predictions would increase in the case of high consistency between predictions, just as when applying cross-validation procedures. For the first criterion groups, we only used biographical indications of dyslexia which were acquired from the *Dyslexia Questionnaire*. The indications and decision rules for classification are presented below. Students with inconsistent criteria were classified into a separate group.

Indications of dyslexia:

1. Formal diagnosis of dyslexia by a qualified educational psychologist (yes, no)

In the Netherlands, the so called ‘dyslexieverklaring’ is an official written document that a dyslexic child can acquire after an extensive individual test session performed by a qualified educational psychologist. Despite some drawbacks, as mentioned above, we consider this document as a strong indication of dyslexia in this study.

2. Other test results from school or institute (dyslexic, non-dyslexic, maybe-dyslexic)

Until recently, the cost of acquiring an official document of dyslexia was high and it was common in many schools to arrange a testing procedure within the school. These test results are less reliable than an official document, but still provide important information about the history of language difficulties of an individual student. We considered a positive result ('dyslexic') and an indication of doubt ('maybe-dyslexic') as a strong indication of dyslexia.

3. Extra language lessons in primary school (yes, no)

Some students were never officially or unofficially tested for dyslexia, but reported a history of extra language lessons during childhood. We believe that any sign of difficulties with language should be incorporated into a criterion. However, we considered extra lessons as a weak indication of dyslexia because language difficulties might just as well be the result of general learning difficulties. On the other hand, the importance of this indication should not be underestimated in our study, because it can be expected that most university students do not have general learning difficulties.

4. Dyslexia in family (mother, father, brothers, sisters) (yes, no, not sure)

Heritability estimates of dyslexia in twin studies vary between 40% and 80% (Hensler et al., 2010; Scerri & Schulte-Körne, 2010; Schumacher et al., 2007). A child with a dyslexic parent has a 40%–60% chance of developing dyslexia, a risk that increases if there are more dyslexic family members (Ziegler et al., 2005). These percentages clearly show the usefulness of information about family members. We considered having one or more dyslexic family member as a strong indication of dyslexia, and being not sure about having one or more dyslexic family member as a weak indication.

5. General self-assessment of dyslexia (no, maybe, moderate, severe)

A general self-assessment of dyslexia consisted of only one question: 'Are you dyslexic?' We considered the answers 'severe' and 'moderate' as strong indications of dyslexia and 'maybe' as a weak indication. We assumed that this indication is relatively reliable, because most students in the Netherlands know what dyslexia is. It can therefore be expected that highly educated students in particular – who have just finished school – are capable of assessing for themselves to what extent they were different from other children. In other words: they probably know that something was wrong when they experienced difficulties performing tasks which were easy for most of the other children.

6. Specific self-assessment of dyslexia: five statements (yes, no, a little)

There are five statements: (1) ‘As a child, I experienced difficulties with reading’; (2) ‘As a child, I experienced difficulties with spelling’; (3) ‘I still experience difficulties with reading’; (4) ‘I still experience difficulties with spelling’; and (5) ‘I experience difficulties with writing’. The answers were scored as follows: yes (2), a little (1) and no (0). After adding up the five scores, the total score ranged from 0 to 10 with 10 representing severe difficulties with language. We marked scores of six or higher as a strong indication of dyslexia and a score of three, four or five as a weak indication.

Decision rules and classification into criterion groups:

Dyslexic (D-group) (N=33)

We required the following: a formal diagnosis of dyslexia, together with at least three strong or two strong and two weak indications of dyslexia.

Non-dyslexic (ND-group) (N=256)

We required the following: no formal diagnosis of dyslexia, no strong indications and not more than one weak indication of dyslexia.

Of 495 students, 206 did not satisfy the strict requirements to be classified into either of the two criterion groups. Thus, for these students indications of dyslexia were inconsistent. Most of these students had one strong or more than one weak indication of dyslexia. Three students were of special interest, because they had a formal diagnosis of dyslexia, but not the number of indications needed to be classified into the D-group. As mentioned before, we accounted for the possibility that some students might have received a formal diagnosis on false grounds.

Table 2 shows the number of times that various indications are represented in the two criterion groups. In the ND-group, only 32 weak indications were reported (all for different students), mostly regarding family members who might be dyslexic. Thus, 77.5% of this group did not have any indication of dyslexia. In the D-group all students reported a formal diagnosis, which was required. Strong indications derived from former test results, general self-assessment, and specific self-assessment, and are common. And about half of these students reported dyslexic family members.

Table 2

Criterion groups: specific indications of dyslexia

Indication of dyslexia	Dyslexic (N=33)		Non-dyslexic (N=256)	
Former diagnosis (required for D-group)	33	100%	0	0%
Former test result (strong indication)	31	94%	0	0%
Extra lessons (weak indication)	21	64%	3	1%
Dyslexia in family (strong indication)	16	48%	0	0%
Dyslexia in family (weak indication)	11	33%	25	10%
Self-assessment general (strong indication)	32	97%	0	0%
Self-assessment general (weak indication)	1	3%	2	1%
Self-assessment specific (strong indication)	27	82%	0	0%
Self-assessment specific (weak indication)	5	15%	2	1%

Predictions

The aim of this study was to identify dyslexic and non-dyslexic students with a high reliability. This can be accomplished by classifying as much as possible students with high consistency between the criterion and separate predictions. We therefore carried out predictions based on different predictor variables (sum scores, factor scores and single item scores of tests and questionnaires) and repeated the predictions using criterion groups which were acquired after intermediate adjustments. After each round of predictions, we classified all students based on decisions which are described below.

Sum scores and factor scores

For sum scores of questions, we determined the total scores of seven subscales of the *Communication Questionnaire*: reading, writing, speaking, listening, copying, dictating, and reading aloud. Attempts to arrive at meaningful factor scores failed as a result of the large quantity of items (20 for each subscale).

For sum scores of tests, we determined the total scores of 13 tests: specific language tests and *Short-Term Memory Letters* and *Numbers*. After an exploratory factor analysis, these 13 sum scores loaded on four components which together explained 54% of all variance. The factor scores of these components were used for a separate prediction. Appropriate names for these factors are ‘Spelling’ (*Dutch Dictation*, *English Dictation*, *Missing Letters*, *and*

Incorrect Spelling), ‘Phonology’ (*Pseudowords* and *Sound Deletion*), ‘STM’ (*Short-Term Memory Numbers* and *Letters*) and ‘Order’ (*Letter Order*, and *Mirror Reading*). The tests *Spoonerisms*, *Dutch–English Rhyme Words*, and *Counting Letters* were involved in various factors. Table 3 shows the factor loadings of each test in a rotated component matrix.

Table 3

Rotated component matrix of 13 tests

Test	Component			
	Spelling	Phonology	STM	Order
Dutch Dictation	**0.73**	0.18	0.05	0.20
English Dictation	**0.62**	0.08	0.15	0.08
Missing Letters	**0.65**	0.18	0.08	0.23
Pseudowords	0.16	**0.69**	0.13	0.23
Sound Deletion	0.01	**0.81**	0.10	-0.12
Spoonerisms	0.26	0.11	0.10	0.35
Incorrect Spelling	**0.73**	-0.11	0.17	0.14
Dutch–English Rhyme Words	0.25	0.10	0.42	0.31
Letter Order	0.12	0.05	0.10	**0.66**
Counting Letters	**0.51**	-0.11	0.35	-0.27
Mirror Reading	0.07	-0.06	0.16	**0.72**
Short-Term Memory Numbers	0.14	0.14	**0.85**	0.16
Short-Term Memory Letters	0.16	0.12	**0.86**	0.16

Extraction Method: Principal Component Analysis
Rotation Method: Varimax with Kaiser Normalization

Item scores

The advantage of using single items as predictor variables instead of sum scores is that they account for the possibility that the mistakes dyslexic people make are qualitatively different from ‘normal’ mistakes. We could, for example, imagine a highly intelligent dyslexic person who performs relatively well on a language test, but who makes mistakes on most of the typical dyslexic items. This is an example of ‘overlooking normal performances’, as described by Ramus and Ahissar (2012). The reverse situation would be a non-dyslexic person with low intelligence who performs relatively poorly on a language test, but not necessarily on the typical dyslexic items. This person might falsely be diagnosed as dyslexic using traditional diagnosing methods.

All available questions were used as predictors in two separate analyses, 141 questions from the CQ and 60 questions from the *Language Preference Questionnaire*. Furthermore, 242 items of 10 tests were selected as predictors in a third analysis. We did not select items of *Mirror Reading* because the performance on this test depended merely on speed of performance, nor of the *Short-Term Memory Tests* because the number of items of these tests would enlarge the total number of items too much. For *English Dictation* and *Missing Letters*, both sentences as separate words were used as predictors (every item was a sentence that consisted of two sub-items which were words).

The danger of many predictor variables is that a model fit will for a large part be based on sample-specific variance. In this study, the number of selected items exceeded the number of cases in the smallest group, the D-group (both the original and the adjusted criterion D-group). Therefore, we selected only a relatively small number of items for a third round of predictions. We used two subsets of items for two predictions: one that contained 88 test items and lone that contained 80 questions of the *Communication Questionnaire, Language Preference Questionnaire*, and the *Orientation Questionnaire*. We selected items with a high predictive value, and preserved the original facet design as much as possible.

Characteristics of discriminant analysis (DA) and logistic regression analysis (LR)

We used the *stepwise method* of DA and LR to predict group membership (dyslexic or non-dyslexic), because we wanted to acquire a reduced set of predictors while we did not want to assign some predictors higher priority than others. We entered the criterion groups as criterion in the analysis. For the DA, we set prior probabilities to *all groups equal*. A cross-validation procedure was chosen by selecting the option *leave-one-out classification* in order to prevent a single case being classified partly based on its own behaviour, and to correct for the effects of outliers. All analyses were carried out in SPSS.

Choices between predictions in this study

For all predictions, we carried out both DA and LR. Pohar, Blas and Turk (2004) evaluated DA and LR in various situations. We considered their recommendations to be a reliable guideline for evaluating various predictions. According to these researchers DA should be preferred over LR in the case of small sample sizes, when violations of normality of predictor variables are not too bad (skewness within the interval [−0.2, 0.2]) and in the case of categorical explanatory variables with four or more answer categories. Based on these recommendations we made a choice between DA and LR for each separate prediction.

1. For a prediction based on the sum scores of the *Communication Questionnaire*, DA should be preferred because the scores (except one) were distributed with skewness within the interval [−0.2, 0.2].

2. For a prediction based on the sum scores of the tests, LR should be preferred because most predictors were distributed with skewness outside the interval [−0.2, 0.2] (but within the interval [−0.4, 0.4]).

3. For a prediction based on the single questions of the *Communication Questionnaire*, DA should be preferred because all items had seven answer categories with most of them showing a relatively limited skewness of distribution.

4. For a prediction based on the item scores of the tests, LR should be preferred because most test items were dichotomous.

5. For a prediction based on the single questions of the *Language Preference Questionnaire*, DA should be preferred because all items had five answer categories, and most items were normally distributed.

6. For a prediction based on the factor scores of the tests, LR should be preferred because the factor scores were distributed with skewness below −0.5.

7. For a prediction based on a selection of questions of the *Communication Questionnaire* and *Language Preference Questionnaire*, DA should be preferred because all items had five or seven answer categories with most of them showing a relatively limited skewness of distribution.

8. For a prediction based on a selection of item scores from the tests, LR should be preferred because most test items were dichotomous.

Decision rules

Two of the steps that we took to acquire reliable predictions were that we made sure that various predictions were independent and that we repeated the predictions on adjusted criterion groups. In this way, we expected to acquire test-retest reliability, with the additional benefit that predictions would not only be based on the behaviour of extreme groups of dyslexics and non-dyslexics, but also on the behaviour of dyslexics who only show mild impairments. After various predictions, the criterion groups were adjusted following certain decision rules. All students were reassigned six times and this mostly affected the students who were not classified into the first criterion groups. Each step of reassignment is described below. Table 4 shows the results of all these steps.

Table 4

Reassignment (RA) procedures applied on criterion groups

	Criterion	RA1	RA 2	RA 3	RA 4	RA 5	Final RA
Dyslexic	33	49	73	46	62	71	74
Non-dyslexic	256	255	237	227	322	347	369
Remaining Students	206	191	185	222	111	77	52

Reassignment 1:

We adjusted the original criterion groups based on six predictions (sum scores *Communication Questionnaire* and tests; factor scores tests; item scores *Communication Questionnaire, Language Preference Questionnaire* and tests), so that the consistency between the criterion and the predictions was high. First, we added up the six predictions which resulted in a score that ranged from zero (low chance of dyslexia) to six (high chance of dyslexia). Next, we reassigned all students. The requirements for the new criterion of dyslexia were a prediction score of four or higher and not a previous criterion of non-dyslexia. The requirements for the new criterion of non-dyslexia were a prediction score of zero and not a previous criterion of dyslexia. The main result of this reassignment was that the D-group was enlarged.

Reassignment 2:

We repeated all predictions with the new criterion groups, and repeated exactly the same reassigning procedure, so that we again acquired new criterion groups. The main result was that the group of dyslexics was enlarged even more. This might be explained by the fact that the predictions were now also based on the behaviour of moderately impaired dyslexics.

Reassignment 3:

Another possibility is to add up all predictions of the two previous reassignment procedures. This resulted in scores that ranged from zero to 12. We determined a cut-off score of eight or higher for the new D-group, and zero for the new ND-group. Thus, at this point, we had evaluated the original criterion groups based on two separate reassignment procedures and on a prediction score of these two procedures together. Only with complete consistency between these three different classifications and following the same requirements regarding consistency with the original criterion, students were classified as dyslexic or as non-dyslexic. The group of dyslexics now became smaller again.

Reassignment 4:

Next, we carried out predictions based on two selections of items (questions and test items). We added up these predictions. A score of two represented consistency regarding dyslexia and a score of zero represented consistency regarding non-dyslexia. The requirements for the new criterion of dyslexia were a prediction score of two and not a previous criterion of non-dyslexia. The requirements for the new criterion of non-dyslexia were a prediction score of zero and not a previous criterion of dyslexia. The main result of this reassignment was that both the D-group and the ND-group was enlarged.

Reassignment 5:

After the last procedure, 111 students were still not yet classified. For these students, we compared the last predictions with all previous ones again, and classified students in the case of high consistency. This was the case for 34 students. They had inconsistent predictions according to reassignment 4 but consistent predictions according to reassignment 3.

Final reassignment:

We repeated the exact procedures of reassignment 4 and 5 on the new groups. Thus, we repeated the two predictions based on single items (as in reassignment 4), and again we added up these predictions. A score of two represented consistency regarding dyslexia and a score of zero represented consistency regarding non-dyslexia. Also students with a score of one but with consistent predictions before (as in reassignment 5), were classified as dyslexic or non-dyslexic. After this procedure, only 52 students remained without a reliable classification.

Predictions

Below, all predictions are summarized. Sum scores, factor scores and items which were inserted in the regression equation are reported as well. Tables 5 and 6 show how the original criterion groups were predicted. For most intermediate predictions a few students could not be predicted because of missing variables. The number of students with missing predictions varied across predictions because, for each prediction, the predictor variables that were inserted in the equation differed.

Predictions first round:

1. Sum scores *Communication Questionnaire* (DA)	Reading, Writing, Listening, Dictation
2. Sum scores tests (LR)	*Dutch Dictation*, *English Dictation* *Missing Letters*, *Spoonerisms*, *Short-Term Memory Numbers*
3. Questions *Communication Questionnaire* (DA)	26 items
4. Item scores tests (LR)	12 items
5. Questions *Language Preference Questionnaire* (DA)	13 items
6. Factor scores tests (LR)	Spelling, Memory, Order

Predictions second round (the same predictions of the first round repeated):

7. Sum scores *Communication Questionnaire* (DA)	Reading, Writing, Listening, Dictation, Reading aloud
8. Sum scores tests (LR)	*Dutch Dictation*, *English Dictation* *Missing Letters*, *Spoonerisms*, *Short-Term Memory Numbers*
9. Questions *Communication Questionnaire* (DA)	20 items
10. Item scores tests (LR)	8 items
11. Questions *Language Preference Questionnaire* (DA)	12 items
12. Factor scores tests (LR)	Spelling, Memory, Order, Phonology

Predictions third round:

13. Selection questions (DA)	15 items
14. Selection test items (LR)	13 items

Predictions fourth round (the same predictions of the third round repeated):

15. Selection questions (DA)	15 items
16. Selection test items (LR)	24 items

Table 5

Intermediate predictions of the original criterion group of dyslexics (N = 33), leading to a final identification

Predicted as	Predictions first round						Predictions second round						Predictions third round		Predictions fourth round		Final identification
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Dyslexic	25	16	29	23	29	16	26	20	29	19	26	18	31	25	30	26	33
Non-dyslexic	6	9	3	5	4	7	5	5	3	9	7	5	1	1	2	0	0
Missed Predictions	2	8	1	5	0	10	2	8	1	5	0	10	1	7	1	7	0

Table 6

Intermediate predictions of the original criterion group of non-dyslexics (N = 256), leading to a final identification

Predicted as	Predictions First Round						Predictions Second Round						Predictions Third Round		Predictions Fourth Round		Final identification
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Dyslexic	25	2	3	10	12	2	34	10	10	16	24	4	7	9	4	1	1
Non-dyslexic	231	234	253	236	244	232	222	226	246	230	232	230	249	235	251	243	255
Missed Predictions	0	20	0	10	0	22	0	20	0	10	0	22	0	12	1	12	0

Conclusion

From Table 4, it becomes clear that all except 52 students could be classified with high consistency. The original D-group was enlarged from 33 to 74, and the original ND-group was enlarged from 256 to 369. For 15 of the 52 not classified students, too many missed predictions made a reliable classification impossible. For the other 37 students, predictions were too inconsistent for a reliable classification.

Table 5 shows that all dyslexic students according to the original criterion were identified as dyslexic. Table 6 shows that all non-dyslexic students according to the original criterion were identified as non-dyslexic except for one. It should be noted that the false positives and false negatives according to the separate predictions were in most cases not the same students. Therefore, we can conclude that the false positives and false negatives of the separate predictions are mainly shortcomings of these predictions themselves, while the summation of repeated predictions is consistent with an independent criterion.

We further should note that most of the 37 students, who could not be classified, could also not be classified in a majority of the six reassignment procedures. A majority of 21 of these students could not be classified in all reassignment procedures; ten students were classified as dyslexic once; one student was classified as non-dyslexic once; four students were classified as dyslexic three times; and one student was classified as dyslexic five times. Thus, only five students could be classified more than one time.

Another conclusion from Tables 5 and 6 is that the predictions based on a selection of items are the most reliable because these predictions resulted in the smallest number of false positives and negatives compared to the original criterion groups.

Differences between groups

Table 7 – similar to Table 2 – shows the number of times that various indications of dyslexia are represented in the final groups of dyslexics and non-dyslexics. Remarkable is that two students with a formal diagnosis were classified as non-dyslexic, whereas it might be expected that predictions would be inconsistent for these students. The former diagnosis was probably provided to the students on false grounds. All remaining percentages make sense. We should note that the percentage of dyslexic students who have a strong indication of dyslexic family members is about the same for the criterion D-group (48%) as for the final classification D-group (43%). This supports the reliability of the classification of dyslexics that do not have a former diagnosis.

Table 7

Final classification results: specific indications of dyslexia

Indication of dyslexia	Dyslexic (N=74)		Non-dyslexic (N=369)	
Former diagnosis	33	45%	2	1%
Former test result (strong indication)	36	49%	4	1%
Extra lessons (weak indication)	36	49%	16	4%
Dyslexia in family (strong indication)	32	43%	43	12%
Dyslexia in family (weak indication)	13	18%	31	8%
Self-assessment general (strong indication)	33	45%	0	0%
Self-assessment general (weak indication)	19	26%	15	4%
Self-assessment specific (strong indication)	49	66%	4	1%
Self-assessment specific (weak indication)	13	18%	32	9%

Group differences of sample characteristics (gender, age and handedness) and sum scores of tests and questionnaires were calculated between the criterion groups and between the final groups. Furthermore, we wanted to know whether differences exist between the dyslexics according to the original criterion and the dyslexics that were added after the predictions, and likewise, between the non-dyslexics according to the original criterion and the non-dyslexics that were added after the predictions. This resulted in three large tables which can be found in Appendix B. Here, we only present a summary of the results. The significance of differences between the various groups was determined with post hoc comparisons of an ANOVA analysis (Tukey, 0.05 level).

First, we calculated the differences between the three original criterion groups of dyslexics (33), non-dyslexics (256) and students without a criterion (206). Sample characteristics did not differ between the groups. The ND-group performed better than the D-group (most $p<0.005$) on all subscales of the *Communication Questionnaire*, on all *Short-Term Memory Tests*, on all specific language tests (except *Sound Deletion*), on three school grades (Dutch, English and other languages), and on the intelligence tests *Vocabulary*, *Verbal Analogies*, *Speed of Calculation*, and *Hidden Figures*. No significant differences were found for *Raven Progressive Matrices*, *Conclusions*, *Numeric Progressions*, and two other school grades (mathematics and remaining courses). All performances of the students without a criterion were higher than those of the D-group and lower than those of the ND-group, with some of these differences significant and some not significant.

Second, we calculated the differences between the two final groups of dyslexics (74), non-dyslexics (369) and non-identified students (37). Again, sample characteristics did not differ between the groups. The ND-group performed better than the D-group (all $p<0.0005$) on all subscales of the *Communication Questionnaire*, on all *Short-Term Memory Tests*, on all specific language tests, on three school grades (Dutch, English and other languages) (all $p<0.005$), and on the intelligence tests *Vocabulary*, *Verbal Analogies*, *Speed of Calculation*, and *Hidden Figures* (all $p<0.0005$), *Numeric Progressions* ($p=0.021$), and *Raven Progressive Matrices* ($p=0.022$). No significant differences were found for *Conclusions* and two other school grades (mathematics and remaining courses). The performances of the students that could not be identified (NI-group) were not, as might be expected – like those of the students without a criterion – somewhere in between the performances of the D-group and the ND-group. Instead, the performances of the NI-group on the specific language tests were sometimes close to the performances of the D-group and sometimes close to those of the ND-group. Furthermore, the performances on the subscales of the *Communication Questionnaire* were almost exactly the same as those of the D-group, being lower than those of the ND-group ($p<0.0005$).

Third, we calculated the differences between the original criterion group of dyslexics (33) and the later-identified dyslexics (41), and between the original criterion group of non-dyslexics (255, minus the false negative!), and the later-identified non-dyslexics (114). Sample characteristics did not differ between the groups. The two groups of dyslexics only differed on *Sound Deletion* with the later-identified dyslexics having lower scores than the original dyslexics ($p=0.036$). But statistically, the finding of only one significant difference out of 39 comparisons is not relevant. The two groups of non-dyslexics only differed on *Incorrect Spelling* with the later-identified non-dyslexics having lower scores than the original non-dyslexics ($p=0.042$). This is again statistically not relevant. However, the later-identified non-dyslexics also performed worse than the original non-dyslexics on all scales of the *Communication Questionnaire* (all $p<0.05$). Cross-comparisons between the two groups of dyslexics and the two groups of non-dyslexics for these last results showed that the performances of the later-identified non-dyslexics were still significantly better than of both groups of dyslexics (all $p<0.05$).

Discussion

With a new method of diagnosing dyslexia in adults, 89.5% of 495 students could be identified as dyslexic (74) or non-dyslexic (369) with high reliability. The main characteristic of this method is that it is iterative: a criterion of biographical information was adjusted after repeated predictions. The most remarkable finding was that the most reliable predictions were acquired using test items (24) and self-report questions (15) as predictor variables instead of sum scores. We will discuss the reliability of the method, and the pros and cons of the method together with recommendations for future research.

Reliability of results

The main support for the reliability of the classification method is that there was a high consistency between separate predictions, which can be interpreted as an extended cross-validation procedure. However, one might argue that with this method only sample-specific features were successfully identified, while it is uncertain whether these features are symptomatic of dyslexia. A first objection to this is that it is not known what the most characteristic features of dyslexia are in general. Furthermore, we believe that in this study the reliability of the classification method is also supported by the fact that the classification results were highly consistent with an independent criterion and by the fact that the predictions based on test results and the predictions based on a self-report of dyslexia were highly consistent with each other, especially when single test items and questions were used as predictor variables.

That the predictions based on items were the most reliable ones in our study was what we expected, especially after reading the discussion of Ramus and Ahissar (2012) regarding the interpretation of poor and normal performances. Sum scores are vulnerable to the overgeneralizing of certain mistakes, while it is unclear which mistakes are representative for the difficulties that accompany dyslexia. For instance, while phonological impairments are generally accepted, their exact nature remains unclear, with interpretations varying from deficits in the phonological lexicon or deficits in processes of retrieval from the lexicon (e.g. Blomert, Mitterer, & Paffen, 2004; Blomert & Willems, 2010). Thus, a collection of items can successfully distinguish dyslexics and non-dyslexics by identifying specific difficulties that accompany dyslexia. A requirement, of course, is that in the analysis items must be implied that together represent as many typical difficulties of dyslexia as possible. This was indeed the case for the two collections of predictor items. From the consistency between the predictions based on test items and the predictions based on self-report questions we conclude that, apparently, students are capable of identifying the exact nature of difficulties themselves, which is most remarkable when we realize that the questions were part of a questionnaire in which no link to dyslexia was made. We therefore conclude that the predictive power of self-report questions may have been underestimated so far.

The fact that all 289 students with a strong criterion of dyslexia or non-dyslexia were classified correctly, except for one mismatch – a student without a criterion of dyslexia who was predicted as dyslexic – strongly supports the reliability of the method because the original criterion and the classification results were independent from each other. Two remarks should be made here. First, one might argue that there is no complete independency because the official document of dyslexia might have been partly based on the same type of tests as used in our prediction analyses. However, a requirement for a criterion of dyslexia was high consistency between biographical indications of dyslexia. In fact, we excluded three students with an official document from the criterion dyslexic group because other indications pointed to no dyslexia. Second, questions may arise about the independency between the biographical self-report questions of the criterion and the questions of the *Communication Questionnaire*. We emphasize here that these questions are both quantitatively and qualitatively different from each other. The biographical questions are merely questions that request information about a history of language-related difficulties, while the questions of the *Communication Questionnaire* request information about very specific difficulties in the use of language which are not necessarily related to dyslexia, and without mentioning that the questions were meant to measure dyslexics' difficulties.

More support for the reliability of the method can be derived from a comparison of characteristics and test results between the original criterion groups and the dyslexics and non-dyslexics that were identified with predictions only. For instance, the most objective indication of dyslexia in the criterion is family members being dyslexic. We found that 48% of the criterion dyslexics and 40% of the added dyslexics had a dyslexic parent, brother or sister, which is consistent with heritability estimates of dyslexia in twin studies varying between 40 and 80 % (Hensler et al., 2010; Scerri & Schulte-Körne, 2010; Schumacher et al., 2007). Furthermore, there were no differences between the two groups of dyslexics in performances of tests or subscales of questions, except for one, which was merely the result of making many post hoc comparisons. However, the later-identified non-dyslexics performed worse than the criterion non-dyslexics on all subscales of the *Communication Questionnaire* (but still better than the groups of dyslexics). An explanation might be that the later-identified non-dyslexics have overcome language difficulties which are not necessarily related to dyslexia. The fact the predictions based on questions for this group were consistent with other predictions underlines that using sum scores as predictors of dyslexia might be misleading.

Finally, a few remarks should be made about the differences between groups in measurements of intelligence. Generally, it is assumed that dyslexics and non-dyslexics do not differ in general intelligence. We found that the non-dyslexics in this study outperformed the dyslexics in some specific intelligence tests. For some tests, such as *Vocabulary* and *Verbal Analogies*, this result makes sense because these tests partly depend on language skills. However, it was unexpected that the non-dyslexics also performed better than the dyslexics on the *Raven Progressive Matrices*, while no difference was found for *Conclusions*, tests which are both assumed to measure general intelligence. On the other hand, school grades were as expected, with lower school grades of language courses for dyslexics, but no differences in school grades of mathematical and other courses. An explanation for the difference in the *Raven Progressive Matrices* might be that performances on this test are actually influenced by dyslexia to some extent, but with this influence being so small that it usually remains undetected in small samples.

Pros and cons of the method and recommendations

A shortcoming of the classification method in this study was that of 495 students 10.5% could not be identified. For 15 students, this was due to too many missing variables. Before our analyses, we excluded many students because of fraudulent behaviour. However, in the case of students having only a few suspicious scores, only the suspicious scores were deleted. Apparently, this led to insufficient information for a reliable prediction for 15 students. Thus, of the remaining 480 students, 8% could not be identified due to inconsistent predictions. One explanation is that the method in this study fell short as a result of using many predictor variables for some separate predictions. This might have resulted in the overestimation of dyslexics or non-dyslexics in separate predictions, leading to inconsistency between predictions. Another explanation is that there will always be a few dyslexics who manage to overcome some difficulties as a result of highly developed compensation strategies and a few non-dyslexics who just by chance perform poorly on some specific items or tests. A third explanation is that some students overestimate or underestimate their difficulties in the self-report questions, which also leads to inconsistency between predictions. From a theoretical point of view, a fourth explanation might be derived from the unresolved issue as to whether dyslexia is a distinct trait or a trait which is normally distributed. Maybe there is a small group of people who only show a few symptoms of dyslexia, which might be caused by something else than dyslexia.

We took nine methodological measures with the aim to improve the reliability of diagnosing dyslexia in adults, with most of them being new compared to previous methods. Although most of them showed clear benefits, we found reasons to believe that some measures can be improved in future studies, especially in studies with other samples than in our study. First, the advantage of applying strict exclusion and inclusion criteria for students to participate in further analyses was that it became clear that there might always be a group of people who cannot be identified for reasons other than differences in age, health, or intelligence. However, for samples of a general population instead of students, this might be a more complicated issue. Second, the criterion in this study was reliable as a starting point of further analyses. Third, the advantage of using many tests and questionnaires appeared to be that most of the symptoms were included in the most reliable predictions. However, one consideration would be to test for even more symptoms. For instance, attentional symptoms of dyslexia were mainly included in the self-report questions but not in the tests (while it cannot be excluded that some items of the tests depended partly on attention abilities). Fourth, the use of two statistical techniques (discriminant analysis and logistic regression analysis) which do not depend on subjective preferences for certain theories of dyslexia was successful by applying the *stepwise method*. In this way, we acquired a reduced set of predictors without assigning some predictors higher priority than others. Fifth, we avoided the use of cut-off scores by using items as predictor variables. These predictions proved to be more reliable and consistent than predictions based on sum scores. Sixth, instead of choosing between the analyses beforehand, a comparison between both analyses for each separate prediction resulted in more consistency between predictions than would have been the case using only one analysis. Seventh, using a large sample made it possible to use many predictors in the analyses, which is needed when using items as predictors. Eighth, using self-report questions as predictor variables supported the reliability of the identifications, because the resulting predictions were highly consistent with predictions based on test items. Ninth, the results showed that the consistency and reliability of predictions increased with the use of repeated predictions. An additional advantage of this study was that no productive tasks were used, which are generally time-consuming. For future methods of identifying dyslexia, this is an advantage in a practical sense.

Conclusion

This study showed that the reliability of diagnosing dyslexia in adults can be improved using an independent criterion in combination with using test items and self-report questions as predictor variables. The main characteristic of the analyses in this study was that the reliability could be improved with repeated predictions. We believe that this study provides new tools for diagnosing dyslexia for future studies, including in other languages. One clear disadvantage of the method in this study is that it requires a large sample and a long testing time. We have provided recommendations for future studies of which the most important one is that samples other than students' should be investigated.

Acknowledgements

The authors thank Ineke van Osch for her help with the development of tests, Paul Brouwer and Nihayra Leona for their work and patience in programming many of our newly developed tests, and Jan Hoogeboom for his tremendous efforts in the processing of data.

Appendix A

Differences in scores between students of 2009 and students of 2010

	Group 1, 2009 (N=241)			Group 2, 2010 (N=254)			
Test / Questionnaire	N	Mean	SD	N	Mean	SD	*p*
Tests of Intelligence							
Vocabulary	238	13.5	3.9	253	12.7	3.6	0.03*
Verbal Analogies	238	21.6	5.2	253	22.4	4.7	0.07
Conclusions	223	19.6	7.9	253	19.1	7.1	0.46
Numeric Progressions	240	13.3	3.6	253	12.4	3.0	0.00*
Speed of Calculation	241	16.5	6.8	251	15.1	6.3	0.02*
Hidden Figures	227	20.9	6.3	254	18.7	5.4	0.00*
Raven Progressive Matrices	230	22.2	4.0	252	21.1	4.4	0.01*
School Grades							
School Grade Dutch	199	6.7	0.8	217	6.8	0.8	0.52
School Grade English	200	6.7	1.2	215	6.9	1.1	0.05
School Grade Other Languages	210	6.8	0.7	216	6.9	0.7	0.36
School Grade Mathematics	214	6.6	0.8	214	6.4	0.8	0.01*
School Grade Remaining Courses	218	6.9	0.8	218	6.9	0.6	0.66
Short-Term Memory (STM)							
STM Numbers Forwards	238	17.1	3.6	249	17.5	3.6	0.11
STM Numbers Backwards	235	19.0	3.6	247	18.9	3.8	0.90
STM Numbers Total	235	36.0	6.4	245	36.6	6.7	0.38
STM Letters Forwards	239	14.9	3.8	253	14.5	3.8	0.22
STM Letters Backwards	235	16.5	4.5	250	16.3	4.2	0.49
STM Letters Total	235	31.5	7.5	250	30.8	7.2	0.28

* Group 1 performing better
** Group 2 performing better

Appendix A

Differences in scores between students of 2009 and students of 2010

	Group 1, 2009 (N=241)			Group 2, 2010 (N=254)			
Test / Questionnaire	N	Mean	SD	N	Mean	SD	*p*
Questionnaire Communication (QC)							
QC total	241	731	109	253	749	93	0.05
QC subscale Reading	241	99	18	253	103	12	0.00**
QC subscale Writing	241	106	17	253	108	15	0.07
QC subscale Speaking	241	101	16	253	104	13	0.01**
QC subscale Listening	241	104	16	253	106	13	0.21
QC subscale Copying	241	107	19	253	107	17	0.80
QC subscale Dictating	241	110	20	253	113	17	0.09
QC subscale Reading Aloud	241	105	19	253	107	17	0.24
Specific Language Tests							
Dutch Dictation	240	34.8	3.0	254	34.9	2.9	0.49
English Dictation	240	19.2	1.1	253	19.3	1.0	0.16
Missing Letters	238	17.5	1.8	250	17.5	1.8	0.79
Pseudo Words	239	26.4	2.1	250	26.4	1.9	0.82
Sound Deletion	237	18.8	2.0	250	18.7	2.3	0.54
Spoonerisms	222	16.4	3.7	229	16.0	4.0	0.30
Incorrect Spelling	240	36.3	2.8	253	36.2	2.6	0.69
Dutch-English Rhyme Words	240	37.0	3.4	253	37.6	2.7	0.03*
Letter Order	240	36.3	3.2	252	35.3	4.1	0.00*
Counting Letters	236	12.1	1.9	246	12.1	1.9	0.87
Mirror Reading Part 1	241	18.9	1.9	251	18.6	2.4	0.21
Mirror Reading Part 2	241	19.7	1.0	251	19.8	0.8	0.26
Mirror Reading Total	241	38.5	2.6	251	38.4	2.9	0.53

* Group 1 performing better
** Group 2 performing better

Appendix B1

Differences between the three criterion groups (33 dyslexics, 256 nondyslexics, and 206 maybe dyslexics)

	Dyslexic (N=33)				Maybe dyslexic (N=206)				Non-dyslexic (N=256)			
	N	Mean	SD	Zscore	N	Mean	SD	Zscore	N	Mean	SD	Zscore
Sample characteristics												
Age (years)		19.8	1.3			19.7	1.6			19.7	1.5	
Age (months)		238	16			237	20			236	18	
Gender (male)	10 = 30.3%				53 = 25.7%				62 = 24.2%			
Gender (female)	23 = 69.7%				153 = 74.3%				194 = 75.8%			
Handedness (right)	28 = 84.8%				172 = 83.5%				214 = 83.6%			
Handedness (left)	2 = 6.0%				22 = 10.7%				29 = 11.3%			
Handedness (ambidextrous)	2 = 6.0%				9 = 4.4%				12 = 4.7%			
Handedness (missing data)	1 = 3.0%				3 = 1.5%				1 = 0.4%			
Tests of Intelligence												
Vocabulary	33	10.6	3.7	-0.65	202	12.8	3.7	-0.07	256	13.6	3.7	0.14
Verbal Analogies	32	20.7	4.2	-0.27	204	21.1	5.1	-0.17	255	22.9	4.8	0.17
Conclusions	29	17.9	7.3	-0.19	200	18.8	7.3	-0.07	247	19.9	7.7	0.08
Numeric Progressions	33	12.0	2.8	-0.24	204	12.3	3.5	-0.14	256	13.3	3.2	0.14
Speed of Calculation	32	13.7	5.7	-0.32	204	14.3	6.0	-0.23	256	17.3	6.8	0.22
Hidden Figures	29	16.7	5.5	-0.51	204	19.4	6.0	-0.06	248	20.4	5.9	0.11
Raven Progressive Matrices	31	21.0	4.6	-0.14	201	21.3	4.4	-0.09	250	22.0	4.1	0.09
School Grades												
School Grade Dutch	27	6.3	0.5	-0.61	173	6.6	0.8	-0.16	216	6.9	0.8	0.20
School Grade English	26	6.3	1.3	-0.42	173	6.7	1.1	-0.11	216	7.0	1.1	0.14
School Grade Other Languages	27	6.4	0.6	-0.74	177	6.8	0.7	-0.15	222	7.0	0.7	0.21
School Grade Mathematics	23	6.7	0.7	0.26	183	6.5	0.9	-0.08	222	6.5	0.7	0.04
School Grade Remaining Courses	28	6.9	0.6	0.06	185	6.8	0.8	-0.12	223	6.9	0.6	0.09

Appendix B1

Differences between the three criterion groups (33 dyslexics, 256 nondyslexics, and 206 maybe dyslexics)

	Dyslexic (N=33)				Maybe dyslexic (N=206)				Non-dyslexic (N=256)			
	N	Mean	SD	Zscore	N	Mean	SD	Zscore	N	Mean	SD	Zscore
Short-Term Memory (STM)												
STM Numbers Forwards	31	14.5	3.6	-0.78	202	16.7	3.8	-0.17	254	18.1	3.2	0.23
STM Numbers Backwards	30	16.0	4.8	-0.80	196	18.4	4.0	-0.16	256	19.8	3.0	0.22
STM Numbers Total	30	30.4	7.5	-0.89	196	35.1	7.0	-0.18	254	37.9	5.4	0.24
STM Letters Forwards	31	11.9	3.2	-0.73	205	13.8	3.9	-0.22	256	15.7	3.5	0.27
STM Letters Backwards	30	13.7	4.6	-0.63	200	15.8	4.7	-0.14	255	17.2	3.8	0.18
STM Letters Total	30	25.6	7.0	-0.75	200	29.7	7.8	-0.19	255	32.9	6.4	0.24
Questionnaire Communication (QC)												
QC total	32	644	93	-0.95	206	705	97	-0.35	256	780	86	0.40
QC subscale Reading	32	85	16	-1.06	206	96	15	-0.35	256	107	13	0.41
QC subscale Writing	32	89	14	-1.16	206	102	16	-0.34	256	114	13	0.42
QC subscale Speaking	32	94	17	-0.55	206	98	14	-0.29	256	107	13	0.30
QC subscale Listening	32	101	17	-0.25	206	101	15	-0.27	256	109	13	0.25
QC subscale Copying	32	90	17	-0.96	206	102	17	-0.30	256	114	16	0.36
QC subscale Dictating	32	93	15	-0.98	206	106	19	-0.29	256	118	16	0.36
QC subscale Reading Aloud	32	92	18	-0.77	206	101	17	-0.29	256	112	16	0.33
Specific Language Tests												
Dutch Dictation	32	31.3	3.4	-1.18	206	34.3	3.0	-0.20	256	35.7	2.4	0.31
English Dictation	31	18.1	1.7	-1.14	206	19.1	1.1	-0.15	256	19.5	0.8	0.25
Missing Letters	30	15.7	2.4	-1.05	202	17.1	1.8	-0.23	256	18.1	1.4	0.30
Pseudo Words	30	25.5	2.5	-0.43	205	26.1	2.1	-0.13	254	26.7	1.9	0.16
Sound Deletion	30	18.4	2.1	-0.16	204	18.6	2.4	-0.09	253	19.0	1.9	0.09
Spoonerisms	28	13.2	4.1	-0.77	185	15.5	3.9	-0.17	238	17.0	3.5	0.22
Incorrect Spelling	32	33.4	3.5	-1.06	205	35.6	2.7	-0.24	256	37.1	2.1	0.33
Dutch-English Rhyme Words	32	34.4	5.2	-0.97	205	36.8	3.3	-0.18	256	38.2	2.0	0.27
Letter Order	32	33.2	4.2	-0.70	205	35.4	3.7	-0.11	255	36.5	3.4	0.18
Counting Letters	30	10.9	2.0	-0.62	197	11.8	2.0	-0.12	255	12.4	1.7	0.17
Mirror Reading Part 1	32	17.8	2.5	-0.43	204	18.4	2.4	-0.15	256	19.1	1.8	0.18
Mirror Reading Part 2	32	19.4	1.4	-0.33	204	19.6	1.2	-0.12	256	19.8	0.5	0.14
Mirror Reading Total	32	37.2	3.5	-0.46	204	38.0	3.3	-0.16	256	39.0	2.0	0.19

Appendix B2

Differences between the three final groups (74 dyslexics, 369 nondyslexics, and 37 not identified students)

	Dyslexic (N=74)				Not identified (N=37)				Non-dyslexic (N=369)			
	N	Mean	SD	Zscore	N	Mean	SD	Zscore	N	Mean	SD	Zscore
Sample characteristics												
Age (years)		20.0	1.7			19.9	1.5			19.6	1.5	
Age (months)		240	20			239	18			236	18	
Gender (male)	21 = 28.4%				12 = 32.4%				87 = 23.6%			
Gender (female)	53 = 71.6%				25 = 67.6%				282 = 76.4%			
Handedness (right)	63 = 85.1%				30 = 81.1%				309 = 83.7%			
Handedness (left)	5 = 6.8%				6 = 16.2%				41 = 11.1%			
Handedness (ambidextrous)	4 = 5.4%				1 = 2.7%				17 = 4.6%			
Handedness (missing data)	2 = 2.7%				0 = 0%				2 = 0.5%			
Tests of Intelligence												
Vocabulary	73	10.7	3.4	-0.64	37	13.4	3.8	0.07	368	13.6	3.7	0.12
Verbal Analogies	73	19.7	4.4	-0.48	37	20.8	4.9	-0.28	367	22.7	4.8	0.12
Conclusions	69	17.6	6.9	-0.24	37	17.8	8.3	-0.21	357	19.9	7.5	0.07
Numeric Progressions	74	11.9	3.0	-0.29	37	13.0	3.1	0.04	368	13.1	3.4	0.05
Speed of Calculation	73	13.1	5.6	-0.42	37	14.1	4.4	-0.27	369	16.7	6.7	0.11
Hidden Figures	69	16.9	5.3	-0.48	37	18.7	6.3	-0.19	361	20.4	5.9	0.11
Raven Progressive Matrices	72	20.4	4.5	-0.29	35	22.1	3.5	0.11	362	21.9	4.2	0.05
School Grades												
School Grade Dutch	62	6.4	0.6	-0.38	32	6.2	1.1	-0.63	312	6.9	0.8	0.14
School Grade English	63	6.4	1.1	-0.34	32	6.6	1.2	-0.22	309	6.9	1.1	0.09
School Grade Other Languages	64	6.4	0.6	-0.66	31	6.7	0.7	-0.28	319	7.0	0.7	0.16
School Grade Mathematics	61	6.5	0.7	0.03	32	6.3	1.3	-0.20	323	6.5	0.7	0.01
School Grade Remaining Courses	66	6.8	0.5	-0.08	33	6.6	1.4	-0.38	324	6.9	0.6	0.05

Appendix B2

Differences between the three final groups (74 dyslexics, 369 nondyslexics, and 37 not identified students)

	Dyslexic (N=74)				Not identified (N=37)				Non-dyslexic (N=369)			
	N	Mean	SD	Zscore	N	Mean	SD	Zscore	N	Mean	SD	Zscore
Short-Term Memory (STM)												
STM Numbers Forwards	72	14.3	3.4	-0.85	37	15.8	3.7	-0.42	367	18.1	3.3	0.21
STM Numbers Backwards	71	16.0	4.4	-0.80	37	16.8	4.5	-0.58	368	19.8	3.1	0.21
STM Numbers Total	71	30.2	6.9	-0.93	37	32.6	7.4	-0.56	366	37.9	5.5	0.24
STM Letters Forwards	72	11.8	3.1	-0.80	37	13.1	3.9	-0.45	368	15.6	3.5	0.20
STM Letters Backwards	69	13.4	4.5	-0.72	37	14.5	5.7	-0.46	365	17.3	3.8	0.18
STM Letters Total	69	25.3	6.8	-0.84	37	27.7	8.8	-0.51	365	32.9	6.4	0.21
Questionnaire Communication (QC)												
QC total	73	649	86	-0.92	37	655	75	-0.87	369	769	90	0.27
QC subscale Reading	73	86	15	-0.97	37	90	11	-0.74	369	106	13	0.27
QC subscale Writing	73	91	13	-1.09	37	92	12	-1.02	369	113	13	0.32
QC subscale Speaking	73	94	16	-0.62	37	94	12	-0.58	369	105	14	0.18
QC subscale Listening	73	99	15	-0.38	37	98	13	-0.50	369	107	14	0.13
QC subscale Copying	73	92	16	-0.86	37	92	16	-0.84	369	112	16	0.25
QC subscale Dictating	73	94	16	-0.91	37	96	18	-0.82	369	116	17	0.26
QC subscale Reading Aloud	73	93	17	-0.75	37	93	14	-0.75	369	110	16	0.22
Specific Language Tests												
Dutch Dictation	73	31.4	3.2	-1.18	37	33.7	2.4	-0.40	369	35.7	2.3	0.27
English Dictation	72	18.2	1.5	-0.99	37	18.9	1.2	-0.36	369	19.5	0.8	0.23
Missing Letters	71	15.7	2.2	-1.02	36	16.5	1.3	-0.60	368	18.0	1.4	0.26
Pseudo Words	71	25.4	2.4	-0.53	37	25.5	1.9	-0.46	366	26.7	1.8	0.15
Sound Deletion	71	17.6	3.1	-0.55	37	18.7	2.7	-0.03	364	19.0	1.7	0.11
Spoonerisms	64	13.7	3.7	-0.67	37	15.5	4.1	-0.18	342	16.8	3.6	0.15
Incorrect Spelling	73	33.6	3.1	-1.00	37	35.3	2.8	-0.37	368	36.9	2.2	0.23
Dutch-English RhymeWords	73	34.5	4.3	-0.95	37	36.0	3.8	-0.46	368	38.1	2.2	0.23
Letter Order	73	33.7	4.1	-0.56	37	34.5	3.6	-0.35	367	36.4	3.5	0.15
Counting Letters	66	10.7	2.0	-0.73	37	12.1	2.0	0.03	364	12.3	1.7	0.13
Mirror Reading Part 1	73	17.7	2.6	-0.52	37	18.5	2.0	-0.15	369	19.0	2.0	0.12
Mirror Reading Part 2	73	19.3	1.7	-0.51	37	19.9	0.5	0.15	369	19.8	0.6	0.09
Mirror Reading Total	73	37.0	3.8	-0.59	37	38.3	2.0	-0.07	369	38.8	2.3	0.12

Appendix B3

Differences between five groups: the original criterion group of dyslexics (33), the dyslexics that were later identified (41), the original criterion group of nondyslexics (255, minus the false negative!), the nondyslexics that were later identified (114), and the not identified students (37; left out were 15 students with too many missing variables)

	Dyslexic Criterion (N=33)				Dyslexic Added (N=41)			
	N	Mean	SD	Zscore	N	Mean	SD	Zscore
Sample characteristics								
Age (years)		19.8	1.3			20.1	1.9	
Age (months)		238	16			242	23	
Gender (male)	10 = 28.4%				11 = 26.8%			
Gender (female)	23 = 71.6%				30 = 73.2%			
Handedness (right)	28 = 85.8%				35 = 85.4%			
Handedness (left)	2 = 6.1%				3 = 7.3%			
Handedness (ambidextrous)	2 = 1.1%				2 = 4.9%			
Handedness (missing data)	1 = 3.0%				0 = 0%			
Tests of Intelligence								
Vocabulary	33	10.6	3.7	-0.66	40	10.7	3.1	-0.62
Verbal Analogies	32	20.7	4.2	-0.30	41	19.0	4.5	-0.63
Conclusions	29	17.9	7.3	-0.19	40	17.3	6.6	-0.27
Numeric Progressions	33	12.0	2.8	-0.27	41	11.9	3.2	-0.30
Speed of Calculation	32	13.7	5.7	-0.35	41	12.7	5.5	-0.49
Hidden Figures	29	16.7	5.5	-0.51	40	17.1	5.2	-0.45
Raven Progressive Matrices	31	21.0	4.6	-0.15	41	20.0	4.5	-0.40
School Grades								
School Grades Dutch	27	6.3	0.5	-0.60	35	6.6	0.5	-0.21
School Grades English	26	6.3	1.3	-0.41	37	6.5	1.0	-0.29
School Grades Other Languages	27	6.4	0.6	-0.75	37	6.5	0.6	-0.59
School Grades Mathematics	23	6.7	0.7	0.28	38	6.4	0.7	-0.12
School Grades Remaining Courses	28	6.9	0.6	0.06	38	6.8	0.5	-0.17

Non-dyslexic Criterion (N=255)				Non-dyslexic Added (N=114)				Not Identified (N=37)			
N	Mean	SD	Zscore	N	Mean	SD	Zscore	N	Mean	SD	Zscore
	19.7	1.5			19.5	1.6			19.9	1.5	
	236	18			234	19			239	18	
61 = 23.9%				26 = 22.2%				12 = 32.4%			
194 = 76.1%				88 = 77.2%				25 = 67.6%			
213 = 83.5%				96 = 84.2%				30 = 81.1%			
29 = 11.4%				12 = 10.5%				6 = 16.2%			
12 = 4.7%				5 = 4.4%				1 = 2.7%			
1 = 0.4%				1 = 0.9%				0 = 0%			
255	13.6	3.7	0.13	113	13.5	3.8	0.10	37	13.4	3.8	0.07
254	22.9	4.8	0.15	113	22.4	4.8	0.06	37	20.8	4.9	-0.28
246	20.0	7.7	0.08	111	19.7	7.2	0.04	37	17.8	8.3	-0.21
255	13.3	3.2	0.13	113	12.5	3.7	-0.11	37	13.0	3.1	0.04
255	17.2	6.8	0.20	114	15.4	6.5	-0.09	37	14.1	4.4	-0.27
247	20.5	5.8	0.12	114	20.4	6.0	0.10	37	18.7	6.3	-0.19
249	22.0	4.1	0.08	113	21.6	4.4	-0.03	35	22.1	3.5	0.11
216	6.9	0.8	0.21	96	6.8	0.8	-0.01	32	6.2	1.1	-0.63
215	7.0	1.1	0.14	94	6.8	1.0	-0.02	32	6.6	1.2	-0.22
221	7.0	0.7	0.20	98	6.9	0.6	0.07	31	6.7	0.7	-0.28
222	6.5	0.7	0.06	101	6.4	0.7	-0.08	32	6.3	1.3	-0.20
222	6.9	0.6	0.09	102	6.9	0.6	-0.02	33	6.6	1.4	-0.38

Appendix B3

Differences between five groups: the original criterion group of dyslexics (33), the dyslexics that were later identified (41), the original criterion group of nondyslexics (255, minus the false negative!), the nondyslexics that were later identified (114), and the not identified students (37; left out were 15 students with too many missing variables)

	Dyslexic Criterion (N=33)				Dyslexic Added (N=41)			
	N	Mean	SD	Zscore	N	Mean	SD	Zscore
Short-Term Memory (STM)								
STM Numbers Forwards	31	14.5	3.6	-0.80	41	14.1	3.3	-0.89
STM Numbers Backwards	30	16.0	4.8	-0.80	41	16.0	4.2	-0.80
STM Numbers Total	30	30.4	7.5	-0.90	41	30.1	6.4	-0.95
STM Letters Forwards	31	11.9	3.2	-0.78	41	11.8	3.1	-0.81
STM Letters Backwards	30	13.7	4.6	-0.65	39	13.2	4.5	-0.77
STM Letters Total	30	25.6	7.0	-0.79	39	25.1	6.8	-0.87
Questionnaire Communication (QC)								
QC total	32	644	93	-0.97	41	653	81	-0.88
QC subscale Reading	32	85	16	-1.08	41	88	14	-0.88
QC subscale Writing	32	89	14	-1.20	41	92	13	-1.00
QC subscale Speaking	32	94	17	-0.57	41	93	15	-0.66
QC subscale Listening	32	101	17	-0.26	41	98	14	-0.48
QC subscale Copying	32	90	17	-0.97	41	93	15	-0.77
QC subscale Dictating	32	93	15	-1.00	41	96	17	-0.83
QC subscale Reading Aloud	32	92	18	-0.78	41	93	15	-0.72
Specific Language Tests								
Dutch Dictation	32	31.3	3.4	-1.21	41	31.5	3.1	-1.17
English Dictation	31	18.1	1.7	-1.14	41	18.3	1.3	-0.88
Missing Letters	30	15.7	2.4	-1.06	41	15.8	2.1	-0.99
Pseudo Words	30	25.5	2.5	-0.45	41	25.3	2.3	-0.59
Sound Deletion	30	18.4	2.1	-0.17	41	17.0	3.6	-0.83
Spoonerisms	28	13.2	4.1	-0.79	36	14.0	3.5	-0.58
Incorrect Spelling	32	33.4	3.5	-1.08	41	33.8	2.8	-0.94
Dutch-English Rhyme Words	32	34.4	5.2	-0.98	41	34.6	3.4	-0.92
Letter Order	32	33.2	4.2	-0.70	41	34.1	4.0	-0.46
Counting Letters	30	10.9	2.0	-0.63	36	10.5	2.1	-0.82
Mirror Reading Part 1	32	17.8	2.5	-0.46	41	17.6	2.6	-0.56
Mirror Reading Part 2	32	19.4	1.4	-0.37	41	19.2	1.9	-0.62
Mirror Reading Total	32	37.2	3.5	-0.50	41	36.8	4.1	-0.66

Non-dyslexic Criterion (N=255)				Non-dyslexic Added (N=114)				Not Identified (N=37)			
N	Mean	SD	Zscore	N	Mean	SD	Zscore	N	Mean	SD	Zscore
253	18.1	3.2	0.22	114	18.0	3.4	0.19	37	15.8	3.7	-0.42
255	19.8	3.0	0.22	113	19.7	3.3	0.19	37	16.8	4.5	-0.58
253	38.0	5.3	0.25	113	37.8	5.9	0.22	37	32.6	7.4	-0.56
255	15.7	3.5	0.24	113	15.3	3.4	0.12	37	13.1	3.9	-0.45
254	17.2	3.8	0.16	111	17.5	3.7	0.24	37	14.5	5.7	-0.46
254	32.9	6.4	0.21	111	32.9	6.3	0.21	37	27.7	8.8	-0.51
255	781	86	0.38	114	743	93	0.01	37	655	75	-0.87
255	107	13	0.39	114	101	14	-0.01	37	90	11	-0.74
255	114	12	0.40	114	110	14	0.14	37	92	12	-1.02
255	107	13	0.30	114	102	14	-0.08	37	94	12	-0.58
255	109	13	0.24	114	103	16	-0.14	37	98	13	-0.50
255	114	16	0.35	114	108	16	0.04	37	92	16	-0.84
255	118	16	0.35	114	113	17	0.07	37	96	18	-0.82
255	112	16	0.32	114	106	16	0.01	37	93	14	-0.75
255	35.8	2.4	0.30	114	35.6	2.2	0.23	37	33.7	2.4	-0.40
255	19.5	0.8	0.25	114	19.5	0.7	0.19	37	18.9	1.2	-0.36
255	18.1	1.4	0.29	113	17.8	1.5	0.17	36	16.5	1.3	-0.60
253	26.7	1.9	0.15	113	26.7	1.6	0.16	37	25.5	1.9	-0.46
252	19.0	1.9	0.08	112	19.2	1.0	0.17	37	18.7	2.7	-0.03
237	17.0	3.5	0.21	105	16.2	3.8	0.00	37	15.5	4.1	-0.18
255	37.1	2.1	0.32	113	36.4	2.3	0.04	37	35.3	2.8	-0.37
255	38.2	2.0	0.26	113	37.9	2.6	0.17	37	36.0	3.8	-0.46
254	36.5	3.4	0.17	113	36.2	3.6	0.10	37	34.5	3.6	-0.35
254	12.4	1.7	0.16	110	12.2	1.8	0.06	37	12.1	2.0	0.03
255	19.1	1.8	0.16	114	18.8	2.3	0.02	37	18.5	2.0	-0.15
255	19.8	0.5	0.12	114	19.8	0.7	0.02	37	19.9	0.5	0.15
255	39.0	2.0	0.17	114	38.6	2.9	0.02	37	38.3	2.0	-0.07

3

Five describing factors of dyslexia

Two subtypes of dyslexia (phonological, visual) have been under debate in various studies. However, the number of symptoms of dyslexia described in the literature exceeds the number of subtypes, and underlying relations remain unclear. We investigated underlying cognitive features of dyslexia with exploratory and confirmatory factor analyses. A sample of 446 students (63 with dyslexia) completed a large test battery and a large questionnaire. Five factors were found in both the test battery and the questionnaire. These 10 factors loaded on 5 latent factors (spelling, phonology, short-term memory, rhyme/confusion, and whole-word processing/complexity), which explained 60% of total variance. Three analyses supported the validity of these factors. A confirmatory factor analysis fit with a solution of five factors (RMSEA = .03). Those with dyslexia differed from those without dyslexia on all factors. A combination of five factors provided reliable predictions of dyslexia and non-dyslexia (accuracy > 90%). We also looked for factorial deficits on an individual level to construct subtypes of dyslexia, but found varying profiles. We concluded that a multiple cognitive deficit model of dyslexia is supported, whereas the existence of subtypes remains unclear. We discussed the results in relation to advanced compensation strategies of students, measures of intelligence, and various correlations within groups of those with and without dyslexia.

Keywords: Dyslexia, methodological issues, cognitive strategies

Based on: Tamboer, P., Vorst, H. C. M., & Oort, F. J. (2016). Five describing factors of dyslexia. *Journal of Learning Disabilities*, *49*(5), 466-483.

Introduction

One of the hottest topics in studies of dyslexia is the existence of subtypes, which was suggested 20 years ago by Castles and Coltheart (1993). The main question is whether dyslexia should be considered as one disorder with one cause but with different behavioural outcomes, or as a collection of different disorders with some similar symptoms. This debate is complicated by the existence of an overwhelming quantity of reported symptoms, hypotheses, and theories and by the various ways researchers have tried to incorporate all of these findings into causal explanations.

In the literature, we found many symptoms well supported by evidence: impairments of reading (Marinus & De Jong, 2010) and spelling (McCarthy, Hogan, & Catts, 2012), reversal errors (Lachmann & Geyer, 2003), impaired phonological processing (Blomert & Willems, 2010; Vaessen & Blomert, 2010; Ziegler et al., 2010; Ziegler & Goswami, 2006), auditory processing (Schulte-Körne & Bruder, 2010), procedural learning (Nicolson et al., 2010), working memory (Savage, Lavers, Pillay, 2007), rapid naming (Wolf & Bowers, 1999), morphological awareness (Siegel, 2008), visual attention span (Bosse, Tainturier, & Valdois, 2007), and multisensory spatial attention (Facoetti et al., 2010). How should subtypes of dyslexia be related to this overwhelming quantity of symptoms and theories? This question was discussed by Ramus and Ahissar (2012), who stated, 'The abundance and diversity of these new theories partly stem from the fact that the large body of data on cognitive deficits in dyslexia fails to fit into a single coherent theoretical framework' (p.105). Nevertheless, these authors found that two subtypes can be distinguished in the literature: a majority subtype characterized by a phonological deficit and one or several minority subtypes characterized by a visual deficit, whereas additional subtypes of phonological and visual dyslexia might emerge from the consideration of underlying aetiologies. However, no studies have provided evidence for distinct, reliable cognitive (or biological) profiles associated with each subtype. Nevertheless, both phonological and visual/attentional impairments have been proposed to represent the core deficit of dyslexia (for a discussion, see Vidyasagar & Pammer, 2010).

One thing that researchers agree on is that the most common symptoms of dyslexia are related to language difficulties, especially reading and spelling. For instance, a general definition of the World Health Organization (2008) states that dyslexia is a specific reading disorder characterized by a specific and significant impairment in the development of reading skills that is not solely accounted for by mental age, visual acuity problems, or inadequate schooling. Clearly, any subjective view of the causal nature of dyslexia is left out of this definition. However, because reading and spelling abilities cannot have an evolutionary basis by themselves, they must emerge from cognitive abilities that are required for reading and spelling. The discussion about which and how many cognitive abilities these might be (phonological, visual, or any other) is important for reliable diagnoses of dyslexia. Diagnoses based on reading and spelling abilities and subjective theoretical preferences of specialists might result in false positives and false negatives. Moreover, the influence of compensation strategies can then not objectively be determined.

The complex relation between reading deficits and cognitive deficits was underlined in a study by Menghini et al. (2010). These researchers investigated the idea that dyslexia might be accompanied by quite a large number of independent deficits, as proposed by the multiple cognitive deficit model of dyslexia (Pennington, 2006). They conducted a general linear model analysis, taking into account that various tasks require multiple cognitive abilities that are not limited to a single cognitive domain. In a sample of children between 8 and 17 years of age, these researchers found that in a four-step hierarchical regression analysis about half of the variance in two reading efficiency tests could be explained by cognitive deficits, such as phonological, visuospatial, attentional, and executive impairments.

In a study by Le Jan et al. (2011), an attempt was made to refine the diagnosis of dyslexia with a multivariate predictive model based on cognitive deficits only and not on deficits of reading and spelling. Using various exploratory analyses, these researchers found that eight variables from four cognitive categories (metaphonological, morphology knowledge, visuoattentional, and audition) classified 94% of children from elementary school correctly. That dyslexia can be predicted with such a high reliability without a reading and spelling assessment could mean a change of paradigms: dyslexia defined as a combination of cognitive deficits instead of a reading disorder. The predictive model of Le Jan et al. implies that these categories are relatively independent from each other. However, because the number of variables almost exceeded the number of those with dyslexia in this study, exploratory factor analyses were performed within categories and not between categories. Typically, multivariate analyses for investigating how various symptoms of dyslexia are related require large samples, or fewer variables than used in the study by Le Jan et al.

When samples are large enough, the advantage of exploratory analyses is that latent variables can be extracted and interpreted without making theoretical assumptions beforehand. To our knowledge, this has been done only a few times before, on a limited set of tests. For instance, Bosse, Tainturier and Valdois (2007) conducted a principal components analysis in two different samples of French and English children. Data from three phonological tasks and three visual attention span tasks revealed that a visual attention factor accounted for 35.8% of the variance in the French sample and for 40.6% in the English sample, and that a phonological factor accounted for 27.1% of the variance in the French sample and for 33.9% in the English sample. In addition, the researchers tried to determine whether subtypes of dyslexia could be distinguished. They defined selective deficits of each factor such that children were considered as being impaired on a factor when their factorial score fell below the 10th percentile of the control group factorial coefficient. They found that 15% of the French and 7% of the English with dyslexia had both disorders, 19% of the French and 34.5% of the English with dyslexia a selective phonological deficit, 44% of the French and 34.5% of the English with dyslexia a selective visual deficit, and 22% of the French and 24% of the English with dyslexia no deficit at all.

In another study, Di Filippo and Zoccolotti (2012) conducted a factor analysis on various rapid automatized naming tasks in a sample of Italian children. They found that a naming factor accounted for 61.43% of the variance and that a visual search factor accounted for 14.44% of the variance. The first factor specified two factors: pictorial naming and detailed orthographic analysis. Because the pictorial naming factor included both orthographic and nonorthographic stimuli, the authors argued that naming slowness 'is not specifically related to activation of the orthographic lexicon but refers more generally to the ability to retrieve name codes from the semantic lexicon. Therefore, the pictorial factor seems to require the integration of visually derived and phonological codes. As such, it might be relatively independent from purely visual or phonological processes' (p. 386).

The studies by Bosse, Tainturier and Valdois, and Di Filippo and Zoccolotti revealed that factor analysis is a useful tool for extracting independent cognitive features of dyslexia. However, a few details of these studies need some clarification. The children with dyslexia in these studies were selected based on delayed reading. However, although delayed reading is the main element and the only symptom in the definition of the World Health Organization, some caution is required. The possibility exists that some (highly intelligent) children with dyslexia are able to compensate for impaired reading already at the ages around 11 years, which can result in the exclusion of some with dyslexia from samples. On the other hand, children without dyslexia with delayed reading, caused by something else other than dyslexia (low intelligence or poor schooling), may be included in samples. For example, in the study by Le Jan et al., 5 of 13 poor readers who were never investigated before were predicted to be without dyslexia by the cognitive prediction model. The existence of poor readers without dyslexia might explain why almost a quarter of the children with dyslexia in the study of Bosse, Tainturier and Valdois did not show phonological or visual deficits. Another explanation for these results could be that these children without deficits in fact have dyslexia, but with symptoms not measured that still resulted in delayed reading.

We concluded that in most studies, cognitive features are investigated in relation to reading deficits, except for the study by Le Jan et al. (2011). In this study, however, an exploratory factor analysis of many symptoms could not be performed due to a relatively small sample. Nevertheless, the finding that dyslexia can be diagnosed on the basis of cognitive variables without assessing reading or spelling deficits highlights that a systematic exploration of cognitive features underlying dyslexia is justified.

The aim of the present study was to investigate how many independent features can be found in a large battery of tests assessing various deficits related to dyslexia. These included both cognitive abilities such as phonology, but also reading and spelling abilities. To ensure that multivariate analyses could reliably be performed, we used a large sample of 446 students. In addition, we investigated whether the same features could be extracted from a large self-report questionnaire. We took four successive steps in the analysis of the data. First, we performed exploratory factor analyses on both the tests and the questionnaire independently as well as a confirmatory analysis on the whole set. Second, we investigated differences in the latent variables between those with dyslexia and those without. Third, we investigated the validity of the factors of dyslexia by analysing the predictive power of the factors and by investigating relations between factors of dyslexia and factors of intelligence. Fourth, we investigated whether subtypes of dyslexia can be distinguished by analysing individual profiles based on cognitive features of dyslexia.

Method

Participants and procedure

We used data from 446 first-year psychology students (17 – 25 years, 114 males), collected in a previous study (Tamboer, Vorst, & Oort, 2014). This sample consisted of three groups: those with dyslexia (N = 63, 14%), those without (N = 345, 77%), and unknown (N = 38, 9%). All students were raised in the Netherlands, had no serious health problems, and had no history of any serious neurological disorder. The data were collected at the University of Amsterdam. All students were informed about the general nature of the tests and the questionnaires in advance according to a standard protocol. Afterward, the students received a more detailed debriefing. Anonymity was guaranteed by the standard protocol of the University of Amsterdam. All students had up to three weeks after their debriefing to request that their test results were not used.

Assessment of dyslexia

Dyslexia and nondyslexia were assessed with two methods. First, dyslexia and non-dyslexia were determined with six sources of biographical information: a formal diagnosis by an educational psychologist, other test results (at school or at an institution), school records of remedial training during school days, family members with dyslexia, and a general self-assessment of dyslexia. Only in the case of high consistency were students identified as being with dyslexia or not. Second, we applied various discriminant and logistic regression analyses on 10 tests and a self-report questionnaire. In this way, all students could be classified as having dyslexia or not. Finally, dyslexia and non-dyslexia were determined based on consistency between the two methods. This resulted in a group of 63 students with dyslexia, a group of 345 students without dyslexia, and a relatively small group of 38 students who could not be identified because there was too much inconsistency between the two methods of identification.

An important issue in many studies is selection bias. Different selection procedures across studies can result in exclusion of those with dyslexia in some samples, whereas in other samples of those without dyslexia are falsely included. In the present study, we prevented inclusion of those without dyslexia in the dyslexia sample by requiring consistency between two independent methods of identification. Exclusion of those with dyslexia from the dyslexia sample was prevented by the discriminant and logistic regression analyses that classify students without making assumptions about which symptoms of dyslexia are most important. For instance, some students from the sample with dyslexia had many biographical indications of dyslexia, but only no formal diagnosis of dyslexia. The regression analyses prevented those with real dyslexia without a formal diagnosis to be excluded from the dyslexia sample. Nevertheless, in the case of too much inconsistency, students were classified into the group of not identified.

In addition, we used the regression function that has acquired the best prediction of dyslexia as a measure of severity of dyslexia. Figure 1 shows the distribution of this score. This standardized score is clearly not normally distributed, but consisted of two separate distributions with all of those with dyslexia falling within the smallest group and all of those without dyslexia falling within the largest group. The means of both groups differed (*SD* = 2.45) on this standardised score. The groups of those with and without dyslexia did not differ in age, ratio of men to women, or percentage of right-handedness (see Table 1).

Figure 1

Frequency Distribution of severity of dyslexia Z-Regression-Score
(A low score represents more severe dyslexia.)

N = 408

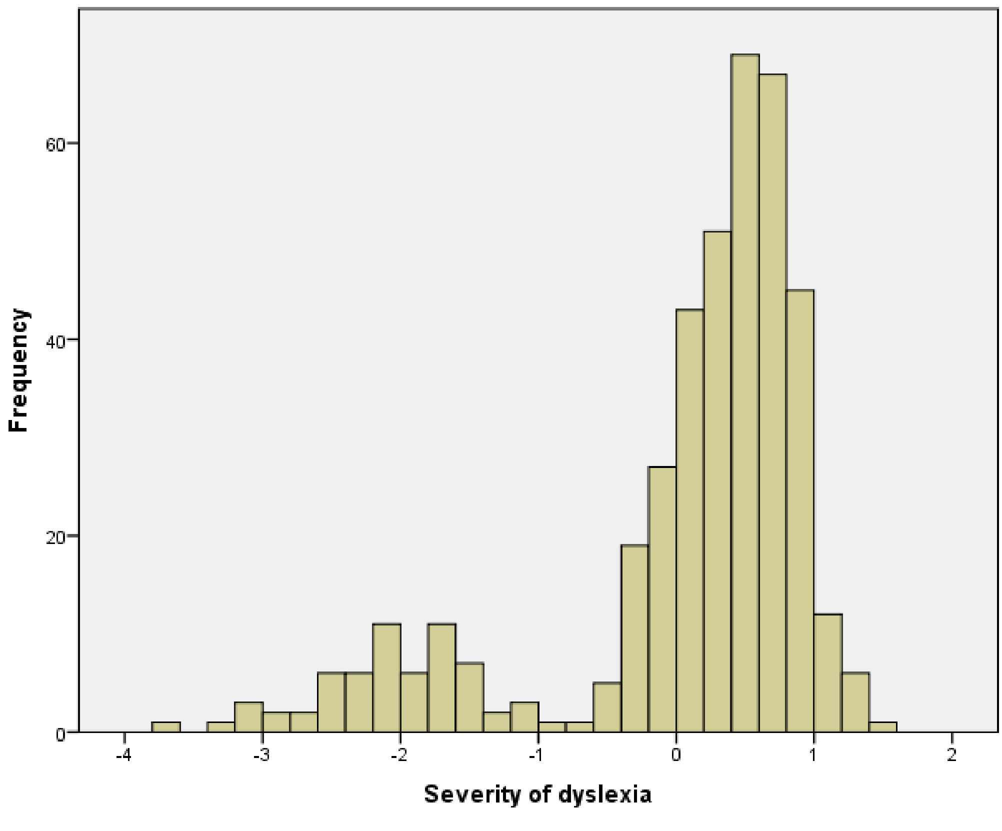


Table 1

Sample characteristics of dyslexic and non-dyslexic students: gender, age, and handedness

	Dyslexic (N=63)	Non-dyslexic (N=345)
Age (years)	20.1 (1.7)	19.7 (1.5)
Age (months)	241 (21)	236 (19)
Gender (male)	18 = 28.6%	83 = 24.1%
Gender (female)	45 = 71.4%	262 = 75.9%
Handedness (right)	53 = 84.1%	289 = 83.8%
Handedness (left)	4 = 6.3%	39 = 11.3%
Handedness (ambidextrous)	4 = 6.3%	15 = 4.3%
Handedness (missing data)	2 = 3.2%	2 = 0.6%
Severity of dyslexia (Z-score)	-2.01 (0.67)	0.44 (0.40)

Tasks

We selected nine tests that are related to cognitive aspects of language processing and a short-term memory test. We did not categorize these tests because we assumed that some tests may require various cognitive abilities. Nevertheless, we tried to incorporate various difficulties of those with dyslexia in these tests, such as phonological, visual, attentional, auditory, and spelling difficulties. Furthermore, we incorporated a large self-report questionnaire, assessing the same difficulties, because we found in our previous study (Tamboer, Vorst, & Oort, 2014) that combinations of self-report questions can reliably predict dyslexia. Further support for the predictive value of a self-report can be found in a study by Snowling et al. (2012). In this study, a factor analysis showed two factors related to dyslexia underlying a 15-item questionnaire: reading and word finding.

Self-report questions

The *Communication Questionnaire* (Kramer & Vorst, 2007) aims to acquire information about specific language difficulties. The term *dyslexia* is not mentioned in this questionnaire. The questionnaire was designed according to a 7 × 5 × 4 facet design. There are seven sub-scales representing different aspects of language, each consisting of 20 statements with seven response categories. Here, we give an example for each subscale: reading: ‘Sometimes I skip a letter, which results in reading a different word’; writing: ‘Sometimes I forget to write down a syllable’; speaking: ‘While speaking, I sometimes exchange similar words’; listening: ‘I hear a story exactly like some-one tells it’; copying: ‘When I copy out a text, I sometimes exchange letters with similar sounds’; dictating: ‘I make mistakes in dictation, because I don’t hear the correct sounds’; reading aloud: ‘When reading aloud, I sometimes repeat a part of the text.’ All statements can also be categorised into five subscales representing sounds, letters, words, sentences, and text. A leading thought during the creation of statements was that four typical mistakes might distinguish those with dyslexia from others: skipping (forgetting), adding, changing, and exchanging.

Short-Term Memory

The *Short-Term Memory Test* – designed for our previous study – aims to measure the capacity of short-term memory. We used the concept of digit span: the number of digits a person can retain and recall. There are four subtests: numbers and letters, both forward and backward. And each subtest consists of 24 series: 6 of 4, 6 of 5, 6 of 6, and 6 of 7 items for the subtests numbers and letters forward, and 6 of 3, 6 of 4, 6 of 5, and 6 of 6 items for the subtests numbers and letters backward. The numbers and letters are presented one by one, for one second each on a computer screen. The participants have to retype these numbers and letters after the last one of a series has been presented. About half of all series consist of some typical difficulties for those with dyslexia, either phono-logical, visual or both. For example, a typical phonological confusion is between the numbers 7 and 9, which resemble each other phonologically in Dutch (zeven/negen). Typical visual confusions are between the numbers 6 and 9 and the letters [m] and [w]. The letters [p], [d], and [b] resemble each other phonologically as well as visually.

Cognitive language tests

Dutch Dictation consists of 10 sentences in the Dutch language (maximum score $10 \times 4 = 40$). Each sentence was read aloud and could be heard through headphones. The whole sentence had to be typed into the computer. Each sentence consisted of a few words that are vulnerable to spelling errors.

English Dictation consists of 10 sentences in the English language (maximum score $10 \times 2 = 20$). Each sentence was read aloud and could be heard through headphones. At the same time, this sentence, except two words, could be read on the computer screen. These words had to be retyped into the computer. Although Dutch students are familiar with the English language, some English words are well known for their vulnerability to spelling errors for Dutch people. The two words had to be typed into the computer.

Missing Letters consists of 10 sentences in the Dutch language (maximum score 10 × 2 = 20). Each sentence was read aloud and could be heard through headphones. For each sentence two words were repeated, while these words were shown on the computer screen with a few letters left out of both words. The missing letters had to be typed into the computer.

Pseudowords consists of 30 pseudowords (maximum score 30), which are nonwords that sound like real words. Each pseudoword was read aloud and could be heard through headphones. At the same time this pseudoword was presented on the computer screen. It had to be decided whether the visually presented pseudoword was spelled correctly, which was the case for half of all pseudowords. Usually pseudowords are administered the other way around, with participants reading the words aloud themselves. The reason for changing this was that it would be practically impossible to get all students in private sessions for this test.

Sound Deletion consists of 20 difficult Dutch words (maximum score 20). Each word was read aloud and could be heard through headphones. Some of these words were pronounced correctly and some words incorrectly by leaving out or adding one letter. On the computer screen, each word was presented three times, each time with a slightly different spelling with one of them being spelled correctly according to what was pronounced. Participants had to decide which of the visually presented words was heard through the headphones. For example, the existing word “fietsen-stalling” – which means bicycle shed – was read out as “fiestenstalling.” The possible answers were “fiet-sentalling,” “fiestensalling,” and “fiestenstalling.”

Spoonerisms consists of 20 words (maximum score 20). A spoonerism is a word that consists of two existing smaller words and that still consists of two small existing words when the first letters of both small words are exchanged. Each word was read aloud and could be heard through headphones. For example, the word “kolen-schop” had to be altered into “scholen-kop” and typed into the computer.

Incorrect Spelling consists of 40 Dutch words (maximum score 40). All words were presented on a computer screen for 50ms each. Half of the presented words were spelled correctly and half were spelled incorrectly. The participants had to decide whether the words were spelled correctly or not.

Dutch–English Rhyme Words consists of 40 Dutch–English word pairs (maximum score 40), presented on the computer screen with the Dutch words on the right. For half of the word pairs, the nouns of the Dutch and the English word resembled each other visually but not aurally. For the other half of the word pairs, the nouns of the Dutch and the English word resembled each other aurally. The participants had to decide whether the two words of a pair rhymed or not.

Letter Order requires the ability to read words as a whole and consisted of 20 Dutch sentences (maxi-mum score 20 × 2 = 40; time limit of 5 min). Theoretical hypotheses about reading words as a whole are described in the dual route model of reading (for an extended description, see De Groot, Dannenburg, & Van Hell, 1994). The idea for this test is based on a well-known text: ‘Aoccdrnig to rscheearch at Cmabrigde uinervtisy, it deosn’t mttaer waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteres are at the rghit pclae. The rset can be a tatol mses and you can sitll raed it wouthit a porbelm. Tihs is bcuseae we do not raed ervey lteter by it slef but the wrod as a wlohe.’ We created Dutch sentences based on the same principle; the order of letters in the words was changed, apart from the first and last letters. The sentences became more difficult toward the end of the test. The sentences had to be typed in with all words correctly spelled. For example, the word ‘Aoccdrnig’ should be typed as ‘According’.

Intelligence tests

Six cognitive tests were based on Guilford’s structure of intellect model: (a) *Vocabulary* (cognition of semantic units: knowing and understanding words and concepts), (b) *Verbal Analogies* (cognition of meaningful verbal relations), (c) *Conclusions* (cognition of meaningful symbolic relations, or the ability to understand and structure difficult situations and the evaluation of semantic implications), (d) *Numeric Progressions* (cognition of symbolic systems in progressions of numbers), (e) *Speed of Calculation* (the ability to assess simple symbolic rules), and (f) *Hidden Figures* (spatial intelligence). For general (nonverbal) intelligence we used the *Advanced Progressive Matrices Set 2* (Raven, Court, & Raven, 1979).

Analyses

To find latent variables related to dyslexia, we carried out various principal components analyses (PCAs) on the nine specific language tests, the four subscales of the *Short-Term Memory Test,* and the specific self-report questions of the *Communication Questionnaire*. A confirmatory factor analysis was carried out in LISREL. All analyses were carried out with the whole sample of 446 students. This means that the issue of selection bias is not relevant for these analyses.

The validity of the factors of dyslexia was further investigated in two ways. First, group differences between students with (N = 63) and without dyslexia (N = 345) on factor scores were determined with *t*-tests. Second, the predictive power of the five factors was analysed with discriminant analysis and logistic regression analysis.

The existence of subtypes of dyslexia was investigated by assessing the nature and number of specific factorial deficits on an individual level. Factorial deficits were determined with scores lower than the 10th percentile for the group of those without dyslexia.

Possible relations between dyslexia and intelligence were investigated with correlational analyses of factors of dyslexia and factors of intelligence. These analyses were carried out for the groups of students with and without dyslexia separately.

Results

The aim was to extract factors of dyslexia from a large data set with exploratory factor analysis. We used the data from the *Communication Questionnaire*, nine specific language tests, and four subscales of the *Short-Term Memory Test*. Table 2 shows all correlations between these scores, with most of them significant at the .01 level. Most correlations are relatively low, except those between the subscales of the *Short-Term Memory Test* and between the subscales of the questionnaire, with the latter ones all greater than .60. These high correlations are problematic when performing factor analysis, because the monotrait–heteromethod correlations should be higher than the heterotrait–monomethod correlations – which appear to be the other way around in this study. Indeed, a PCA based on these scores resulted in only two factors, with all tests loading on one factor and all subscales of the questionnaire loading on the other factor. The next step and successful method was performing factor analyses for the tests and subscales of the questionnaire separately, and then on standardized sum scores of the tests and the questionnaire together.

Table 2
Correlations between sum scores of 9 language tests, 4 short-term memory tests and 7 questionnaires (N = 446)

	DD	ED	ML	PW	SD	SP	IS	DERW	LO	STM NF	STM NB	STM LF	STM LB	CQR	CQW	CQS	CQL	CQC	CQD	CQRA
DD	1																			
ED	.40	1																		
ML	.48	.29	1																	
PW	.27	.15	.24	1																
SD	.09	.08	.12	.26	1															
SP	.19	.16	.20	.17	.04	1														
IS	.44	.37	.39	.13	.02	.19	1													
DERW	.25	.23	.22	.21	.06	.19	.32	1												
LO	.22	.19	.18	.19	.06	.13	.21	.19	1											
STMNF	.25	.23	.24	.24	.17	.15	.27	.29	.21	1										
STMNB	.27	.21	.26	.24	.10	.15	.23	.30	.25	**.61**	1									
STMLF	.28	.30	.26	.22	.17	.17	.26	.30	.20	**.62**	**.53**	1								
STLB	.21	.28	.24	.20	.12	.19	.21	.29	.16	**.53**	**.65**	**.65**	1							
CQR	.24	.23	.29	.12	.08	.14	.26	.30	.18	.29	.24	.31	.23	1						
CQW	.32	.27	.31	.21	.08	.18	.29	.32	.18	.34	.30	.35	.27	**.77**	1					
CQS	.19	.14	.19	.10	.05	.12	.20	.15	.10	.24	.23	.29	.19	**.66**	**.71**	1				
CQL	.10	.10	.12	.13	.07	.15	.13	.16	.08	.22	.19	.27	.22	**.60**	**.66**	**.72**	1			
CQC	.21	.20	.24	.13	.06	.13	.26	.24	.15	.31	.27	.34	.25	**.68**	**.78**	**.68**	**.66**	1		
CQD	.29	.26	.29	.19	.05	.14	.30	.27	.09	.31	.26	.32	.24	**.69**	**.82**	**.72**	**.70**	**.83**	1	
CQRA	.20	.23	.22	.16	.02	.13	.25	.24	.13	.30	.27	.33	.26	**.73**	**.77**	**.76**	**.68**	**.80**	**.80**	1

DD = Dutch Dictation
ED = English Dictation
ML = Missing Letters
PW = Pseudo Words
SD = Sound Deletion
SP = Spoonerisms
IS = Incorrect Spelling
DERW = Dutch-English Rhyme Words
LO = Letter Order
STMNF = STM Numbers Forwards
STMNB = Numbers Backwards
STMLF = Letters Forwards
STMLB = Letters Backwards
CQR = Questionnaire Reading
CQW = Questionnaire Writing
CQS = Questionnaire Speaking
CQL = Questionnaire Listening
CQC = Questionnaire Copying
CQD = Questionnaire Dictating
CQRA = Questionnaire Reading Aloud

PCA of the Questionnaire

A plausible explanation for the high correlations between the subscales of the questionnaire is that various latent variables are balanced over the subscales as a result of the 7 × 5 × 4 facet design. To acquire better insight into the nature of the items regarding these three dimensions, we conducted PCAs on the 20 items for each of the seven original subscales separately. For all subscales, this resulted in two or three interpretable factors with eigenvalues greater than one (all together 18 factors). Next, we determined for each factor the most common feature of the items with common high factor loadings. We found that many of the 18 factors represented the same latent variable. For example, one latent variable – confusion – was represented as a factor for each subscale. Altogether, five different latent variables could be distinguished: spelling, phonology, short-term memory, confusion, and complexity. The first three of these are not surprising, but confusion and complexity need some clarification. Confusion represented difficulties of skipping (forgetting), adding, changing, and exchanging letters, words, or sentences. Complexity represented difficulties of processing complex words or complex sentences. An overview of these results is presented in Table 3. In the next step, we considered the five latent variables that were found in the seven-factor solutions as five scales, and added up single items according to these scales. This resulted in five sum scores. To be able to compare these with sum scores from the tests, we transformed these sum scores to standardized factor scores by requesting all factors in a PCA. The rotated components matrix is shown in Table 4. The factor scores were assigned names according to the sum scores with the highest factor loading.

PCA of the Tests

A PCA of the nine specific language tests and four subscales of the *Short-Term Memory Test* resulted in three factors, together explaining only 52% of all variance: spelling, phonology, and short-term memory. The rotated components matrix is shown in Table 5. The spelling factor represented high factor loadings (>0.63) for all tests that are related to spelling disabilities (*Dutch Dictation*, *English Dictation*, *Missing Letters*, and *Incorrect Spelling*). The short-term memory factor represented high factor loadings (>0.78) for the four subscales of the *Short-Term Memory Test*. The phonology factor represented high factor loadings for *Sound Deletion* (0.81) and *Pseudowords* (0.69). *Spoonerisms* and *Letter Order* both had only moderate factor loadings (0.37) on spelling. *Dutch–English Rhyme Words* had only moderate

Table 3

PCA`s on subscales of the Communication Questionnaire (N = 446)

Subscale	N items	N factors	Factors
Reading	20	2	confusion, complexity
Writing	20	3	confusion, complexity, spelling
Speaking	20	3	confusion, complexity, phonology
Listening	20	2	confusion, phonology
Copying	20	3	confusion, complexity, short-term memory
Dictation	20	3	confusion, complexity, spelling
Reading aloud	20	2	confusion, complexity

Table 4

Rotated Component Matrix: 5 factors explaining all variance of 5 sum scores of the Communication Questionnaire (N = 446)

	Factor 1 (phonology)	Factor 2 (spelling)	Factor 3 (confusion)	Factor 4 (short-term memory)	Factor 5 (complexity)
Sum score – spelling	0.25	**0.88**	0.24	0.28	0.17
Sum score – phonology	**0.85**	0.26	0.30	0.26	0.23
Sum score – short-term memory	0.28	0.33	0.30	**0.83**	0.20
Sum score – confusion	0.31	0.26	**0.84**	0.29	0.21
Sum score – complexity	0.49	0.32	0.38	0.33	**0.64**

Extraction Method: Principal Component Analysis
Rotation Method: Varimax with Kaiser Normalization

Table 5

Rotated Component Matrix: 3 factors explaining 52% of variance of 13 tests (N = 446)

	Factor 1 (short-term memory)	Factor 2 (spelling)	Factor 3 (phonology)
Dutch Dictation	0.08	**0.76**	0.11
English Dictation	0.15	**0.63**	-0.01
Missing Letters	0.11	**0.67**	0.16
Pseudo Words	0.14	0.25	**0.69**
Sound Deletion	0.07	-0.03	**0.81**
Spoonerisms	0.11	0.37	0.17
Incorrect Spelling	-0.14	**0.74**	-0.14
Dutch-English Rhyme Words	0.34	0.43	0.09
Letter Order	0.18	0.37	0.07
STM Numbers Forwards	**0.78**	0.18	0.15
STM Numbers Backwards	**0.81**	0.18	0.04
STM Letters Forwards	**0.79**	0.22	0.11
STM Letters Backwards	**0.84**	0.16	0.04

Extraction Method: Principal Component Analysis
Rotation Method: Varimax with Kaiser Normalization

factor loadings on short-term memory (0.34) and spelling (0.43), which makes the interpretation of this test difficult. However as expected, these are tests that require more than one cognitive ability related to dyslexia.

The finding of the spelling, phonology, and short-term memory factors was consistent with our expectation based on the results of the questionnaire. Unexpectedly we did not find more factors. However, the expectation that more than three latent trait variables might exist is justified for three reasons. First, the three factors from the analysis only explain half of all variance. Second, three tests (*Spoonerisms*, *Letter Order*, and *Dutch–English Rhyme Words*) are only weakly explained by these factors. Third, inspection of the scree plot revealed that four other latent variables could be distinguished with eigenvalues between 0.7 and 0.9. A possibility is to request more components from the PCA, but the number of extra requested factors would depend on purely subjective decisions. Another possibility is to create five sum scores and test the hypothesis that five latent trait variables can be distinguished according to the same trait variables acquired from the questionnaire. Thus, we created three sum scores by adding the sum scores of the tests with high factor loadings on the three factors from the analysis: spelling (*Dutch Dictation* + *English Dictation* + *Missing Letters* + *Incorrect Spelling*), phonology (*Sound Deletion* + *Pseudowords*), and short-term memory (all four subscales). A fourth factor was acquired by adding the sum scores of *Spoonerisms* and *Letter Order* and was named whole-word processing. We used these two tests together because both tests partly depend on the processing of a word as a whole instead of letter-by-letter processing and because both tests show highly similar factor loadings on all factors. The fifth sum score was in fact the sum score of the test *Dutch–English Rhyme Words* and was named rhyme. To be able to compare these sum scores with sum scores of the questionnaire, we transformed these sum scores to standardized factor scores by requesting all factors in a PCA. The factor scores were assigned with names according to the sum scores with the highest factor loading, which were greater than 0.95 for each factor.

PCA of the Whole Data Set

We determined five standardized sum scores of the *Communication Questionnaire* and five standardized sum scores of the tests, which represented all variance for the whole data set. We conducted a PCA with these sum scores and found five factors with eigenvalues greater than one, which together explained 60% of all variance. The rotated components matrix is shown in Table 6. For each of the five factors, one high factor loading was found from a test factor and one from a questionnaire factor, all falling between 0.61 and 0.83, whereas all other factor loadings were 0.28 or less. As expected, we found a spelling factor (15.7% of the variance explained), a short-term memory factor (12.1% of the variance explained), and a phonology factor (10.8% of the variance explained). We named a fourth factor rhyme/confusion (11.3% of the variance explained), which explains the test factor rhyme and the questionnaire factor confusion. A fifth factor we named whole-word processing/complexity (10.3% of the variance explained), which explains the test factor whole-word processing and the questionnaire factor complexity. This analysis supports the existence of two more factors other than spelling, phonology, and short-term memory.

Table 6

Rotated Component Matrix: 5 factors explaining 60% of variance of 10 factors of tests and questionnaire (N = 446)

	Factor 1 (spelling)	Factor 2 (short-term memory)	Factor 3 (rhyme/ confusion)	Factor 4 (phonology)	Factor 5 (whole-word processing/ complexity)
Variance explained	15.7%	12.1%	11.3%	10.8%	10.3%
Questionnaire – spelling	**0.83**	0.01	0.05	0.22	-0.01
Questionnaire – phonology	-0.10	0.28	0.09	**0.61**	0.25
Questionnaire – short-term memory	0.06	**0.71**	-0.13	-0.05	-0.12
Questionnaire – confusion	-0.07	0.21	**0.71**	-0.03	-0.16
Questionnaire – complexity	0.10	-0.03	0.21	-0.18	**0.70**
Test – spelling	**0.82**	0.09	-0.00	-0.13	0.07
Test – phonology	0.15	-0.13	-0.02	**0.77**	-0.15
Test – short-term memory	0.05	**0.72**	0.17	0.11	0.13
Test – rhyme	0.11	-0.17	**0.72**	0.08	0.15
Test – whole-word processing	-0.03	0.03	-0.23	0.20	**0.66**

Extraction Method: Principal Component Analysis
Rotation Method: Varimax with Kaiser Normalization

Validity of Five Dyslexia Factors: Confirmatory Factor Analysis

To find support for the five factors from the exploratory analyses, we conducted a confirmatory analysis in LISREL 8.50. We investigated the hypothesis that the data of the exploratory analysis fitted with a multitrait–multimethod model (MTMM model), which was developed by Campbell and Fiske (1959). We entered the covariance matrix of the five sum scores of the questionnaire and the five sum scores of the tests and distinguished five latent trait variables (the five factors that were found in the exploratory analysis) and two latent method variables (test and self-report). In the suggested model, the variance of each of the 10 sum scores was explained by one trait factor, one method factor, and error variance. Generally, various fit indices are used to quantify the ability of a model to reproduce the observed data. The root mean square error of approximation (RMSEA) is a measure of the error of approximation of the model-implied covariance matrix to the population covariance matrix and should be less than .05. We found an RMSEA of .03 with $p = 0.88$, which means that the null hypothesis of a close fit cannot be rejected. The root mean square residual (RMR) is a badness-of-fit measure based on the differences between the covariance matrix of the model and the covariance matrix of the data and should be less than .05. We found a RMR of .02, which indicates a good fit. Other fit indices are based on comparing the fit of a model with the fit of a baseline model: the nonnormed fit index (NNFI) and the comparative fit index (CFI). Both indices should be close to 1. We found an NNFI of .997 and a CFI of .999. Based on these indices, we could accept the hypothesis that five general factors are underlying features that were found in both the test battery and the questionnaire.

Validity of Five Dyslexia Factors: Group Differences

More support for the legitimacy of the factors was found with analyses of group differences. Table 7 shows that those with dyslexia had lower factor scores than those without dyslexia on all five factors of dyslexia. Effect sizes (Cohen’s d) were calculated by dividing the differences between the means by the mean of the standard deviations. The largest effect size (1.79) was found for the spelling factor. Medium effect sizes were found for the phonology, short-term memory, and whole-word processing/complexity factors. A small effect size was found for the rhyme/confusion factor. We also note that the standard deviations are higher for the group of those with dyslexia as compared to the group of those without dyslexia. In addition, the results of the group of not-identified students are presented. All of their factor scores fell between those of the students with and without dyslexia with some scores closer to those of students with dyslexia and some scores closer to those of students without dyslexia.

Table 7

Group differences on factors of dyslexia

	Dyslexic (N=63)		Non-dyslexic (N=345)					Not-identified (N=38)	
	Mean	SD	Mean	SD	T	Effect Size	*p*	Mean	SD
Factors of Dyslexia									
Spelling	-1.36	1.16	0.31	0.71	15.38	1.79	<0.0000005	-0.59	0.84
Phonology	-0.54	1.45	0.10	0.82	4.94	0.56	0.000001	-0.01	1.30
Short-term memory	-0.49	1.25	0.15	0.90	4.84	0.60	0.000002	-0.53	1.00
Rhyme/confusion	-0.22	1.31	0.12	0.81	2.75	0.32	0.006198	-0.70	1.52
Whole-word-processing/complexity	-0.58	1.16	0.11	0.94	5.10	0.66	0.000001	-0.02	0.91

Validity of Five Dyslexia Factors: Prediction Analysis

Additional support for the legitimacy of the factors was found with prediction analyses: discriminant analysis (DA) and logistic regression analysis (LRA; *stepwise method*, *cross-validated*). The groups of those with and without dyslexia were predicted with all five factors of dyslexia in the regression formula: 92% correct classification with DA and 95% correct classification with LRA. Table 8 presents an overview of these predictions. As all factors of dyslexia contributed to the prediction (were entered in the regression equation), the best prediction of dyslexia is derived when many features of dyslexia are taken into account. This not only supports the legitimacy of the factors but also emphasizes the importance of diagnosing dyslexia on the basis of many different tests.

Table 8

Prediction analyses of dyslexics and non-dyslexics with five factors of dyslexia

Discriminant Analysis

Dyslexics and non-dyslexics predicted with 5 factors of dyslexia
(spelling, phonology, short-term memory,
whole-word-processing/complexity, rhyme/confusion)

92% of cross-validated grouped cases correctly classified

	Dyslexic	Non-dyslexic
Dyslexic (63)	57	6
Non-dyslexic (345)	27	318

Logistic Regression Analysis

Dyslexics and non-dyslexics predicted with 5 factors of dyslexia
(spelling, phonology, short-term memory,
whole-word-processing/complexity, rhyme/confusion)

95% of cross-validated grouped cases correctly classified

	Dyslexic	Non-dyslexic
Dyslexic (63)	50	13
Non-dyslexic (345)	6	339

Subtypes of Dyslexia: Factorial Deficits on an Individual Level

We investigated the existence of subtypes of dyslexia based on individual profiles. These profiles were created on the basis of five factorial scores. Because all factor scores were normally distributed within the whole group, deficits of the five factors were determined whereby students whose score fell below the 10th percentile of the group of those without dyslexia were considered as being impaired on that factor, which was the same decision rule applied by Bosse, Tainturier and Valdois (2007). Table 9 shows for each factor how many students with dyslexia have a deficit. On each factor, 10% of those without dyslexia have a deficit as a result of the decision rule. Table 10 shows the number of deficits for those with dyslexia, those without dyslexia, and not-identified students. A deficit in spelling is the most common of all deficits for those with dyslexia (78%). Furthermore, a majority of those with dyslexia had more than one deficit (75%). In addition, we found that about half of the group of not-identified students had more than one deficit (53%), whereas a very small minority had no deficit (5%). In the study by Bosse, Tainturier and Valdois, subgroups of those with dyslexia could easily be determined, because they found only two factors. Thus, they found four different groups of those with dyslexia: with either one of the two deficits, with two deficits, or with no deficit at all. In the same way, we tried to categorize all possible profiles of factorial deficits, however, without finding any meaningful categorization of combinations. In our sample, the number of possible combinations of deficits was too large to distinguish meaningful and interpretable subgroups.

Table 9

Factorial deficits on an individual level for dyslexics, non-dyslexics not-identified students (deficit when lower score than 10th percentile cut-off score of non-dyslexics)

	Dyslexic (N=63)	Non-dyslexic (N=345)	Not-identified (N=38)
Spelling	77.8%	10%	47.4%
Phonology	34.9%	10%	18.4%
Short-term Memory	38.1%	10%	42.1%
Confusion	28.6%	10%	44.7%
Exchange / Complexity	31.7%	10%	10.5%

Table 10

Number of deficits in the groups of dyslexics, non-dyslexics not-identified students (deficit when lower score than 10th percentile cut-off score of non-dyslexics)

	Dyslexic (N=63)	Non-dyslexic (N=345)	Not-identified (N=38)
0 Deficits	0%	58.0%	5.3%
1 Deficit	25.4%	35.1%	42.1%
2 Deficits	44.4%	6.7%	39.5%
3 Deficits	23.8%	0.3%	10.5%
4 Deficits	6.3%	0%	2.6%
5 Deficits	0%	0%	0%

Relations With Intelligence: Dyslexia as a Multicognitive Deficit

To evaluate the relation between general intelligence and dyslexia-related abilities, we investigated correlations between factors of dyslexia and factors of intelligence. First, we conducted a PCA on seven measures of intelligence: the six cognitive tests based on Guilford's structure of intellect model and *Raven's Advanced Progressive Matrices Set 2*. The rotated components matrix is shown in Table 11. We found two factors with eigenvalues greater than one. Interpretation of the scree plot revealed that it is justified to incorporate a third latent variable in the analysis because its eigenvalue was just less than one (0.92) and much higher than the next eigenvalue (0.70). We distinguished three factors: nonverbal intelligence, speed of (numeric) processing, and vocabulary, together explaining 65% of the variance.

Table 11

Rotated Component Matrix:

3 factors explaining 65% of variance in 7 measurements of intelligence (N = 446)

	Factor 1 Nonverbal intelligence	Factor 2 Speed of (numeric) processing	Factor 3 Vocabulary
Variance explained	35.8%	15.6%	13.2%
Raven Progressive Matrices	**0.70**	0.23	-0.07
Conclusions	**0.59**	*0.41*	0.18
Hidden pictures	**0.78**	-0.06	0.15
Numeric progressions	0.11	**0.83**	0.07
Speed of calculation	0.14	**0.83**	0.06
Verbal analogies	*0.47*	0.15	*0.53*
Vocabulary	0.01	0.06	**0.92**

Extraction Method: Principal Component Analysis
Rotation Method: Varimax with Kaiser Normalization

Table 12

Differences between dyslexics and non-dyslexics on factors of intelligence and schoolgrades

	Dyslexic (N=58)			Non-dyslexic (N=324)						Not-identified (N=38)		
	N	Mean	SD	N	Mean	SD	T	Effect Size	*p*	N	Mean	SD
Factors of Intelligence												
Nonverbal Intelligence	58	-0.35	0.96	324	0.08	0.99	3.11	0.44	0.002	35	-0.18	1.00
Speed of (numeric) Processing	58	-0.28	0.77	324	0.06	1.05	2.39	0.37	0.017	35	-0.10	0.74
Vocabulary	58	-0.62	0.85	324	0.12	0.98	5.41	0.81	<0.001	35	-0.11	1.08
School Grades												
School Grade Dutch	53	6.48	0.54	291	6.86	0.79	3.32	0.57	0.001	32	6.24	1.06
School Grade English	53	6.44	1.19	287	6.93	1.13	2.86	0.42	0.005	32	6.55	1.25
School Grade Other Languages	55	6.46	0.60	297	6.99	0.65	5.65	0.85	<0.001	31	6.69	0.66
School Grade Mathematics	53	6.50	0.75	299	6.51	0.70	0.09	0.01	0.929	32	6.34	1.31
School Grade Remaining Courses	57	6.85	0.55	301	6.92	0.60	0.89	0.12	0.376	33	6.63	1.41

Table 12 shows that those with dyslexia had lower factor scores than those without dyslexia on these factors of dyslexia. The sample sizes are a bit smaller because a few students had no data for intelligence measures. Furthermore, an analysis of group differences of school grades (acquired in our previous study) showed that those with dyslexia had lower scores than those without dyslexia on school grades Dutch, English, and other languages, but not on school grades for mathematics and remaining courses. Effect sizes (Cohen's d) were calculated by dividing the differences between the means by the mean of the standard deviations. Large effect sizes were found only for the vocabulary and school grade other languages factors. We should note that most standard deviations are slightly lower for the group of those with dyslexia as compared to those without dyslexia. In addition, the results of the group of not-identified students are presented.

Correlations between severity of dyslexia, factors of dyslexia, and factors of intelligence are shown in Table 13 (with dyslexia) and Table 14 (without dyslexia). We found several remarkable correlations. First, severity of dyslexia correlated with spelling (.41) only in the group of those without dyslexia. Second, in the group of those with dyslexia, the spelling factor correlated negatively with the other factors of dyslexia, of which two were significant: phonology (.31) and whole-word processing/complexity (0.42). In the group of those without dyslexia these correlations were lower or not significant. Third, in the group of those without dyslexia, various low but significant correlations were found between factors of dyslexia and factors of intelligence, whereas no significant correlations were found in the group of those with dyslexia.

Discussion

Overview of Main Results

With PCA and a large sample of young psychology students, we found five factors of dyslexia that together explained 60% of the variance in a large battery of specific language tests and language related questions. We named these factors as follows: spelling, phonology, short-term memory, rhyme/confusion, and whole-word processing/complexity. Strong support for the legitimacy of these factors was provided by three analyses. First, five factors from two data sets (tests and questions) loaded on five common latent variables. In a confirmatory analysis we tested a fit of the MTMM model and found a solution with an RMSEA of .03.

Table 13

Correlations (Pearson) between severity of dyslexia, factors of dyslexia and factors of intelligence

Group of dyslexics (N=58)

(Severity of dyslexia ~ low score)

	Nonverbal Intelligence	Speed of (numeric) Processing	Vocabulary	Severity of Dyslexia	Spelling	Phonology	Short-term Memory	Confusion	Exchange / Complexity
Nonverbal Intelligence	1								
Speed of (numeric) Processing	0.169	1							
Vocabulary	-0.210	0.035	1						
Severity of dyslexia	-0.168	0.122	0.155	1					
Spelling	-0.224	-0.116	0.053	0.118	1				
Phonology	0.018	0.034	0.160	0.127	-0.314*	1			
Short-term Memory	0.180	-0.040	-0.083	0.176	-0.245	-0.044	1		
Confusion	0.122	0.056	-0.102	0.085	-0.206	-0.008	-0.222	1	
Whole word processing / Complexity	0.166	0.112	0.213	0.029	-0.419**	0.058	-0.020	-0.136	1

*** Correlation is significant at the 0.01 level (2-tailed)*
**   Correlation is significant at the 0.05 level (2-tailed)*

Table 14

Correlations (Pearson) between severity of dyslexia, factors of dyslexia and factors of intelligence

Group of non-dyslexics (N=324)

(Severity of dyslexia ~ low score)

	Nonverbal Intelligence	Speed of (numeric) Processing	Vocabulary	Severity of Dyslexia	Spelling	Phonology	Short-term Memory	Confusion	Exchange / Complexity
Nonverbal Intelligence	1								
Speed of (numeric) Processing	-0.058	1							
Vocabulary	-0.039	-0.047	1						
Severity of dyslexia	0.030	0.048	0.188*	1					
Spelling	-0.120*	0.106	0.086	0.414**	1				
Phonology	-0.065	0.033	0.103	0.059	-0.084	1			
Short-term Memory	0.117*	0.115*	0.057	0.062	-0.200**	-0.073	1		
Confusion	0.118*	-0.094	0.111*	0.070	-0.069	-0.009	0.026	1	
Whole word processing / Complexity	0.147**	0.068	0.143**	0.181**	-0.077	-0.091	-0.046	-0.012	1

*** Correlation is significant at the 0.01 level (2-tailed)*
** Correlation is significant at the 0.05 level (2-tailed)*

Second, those with dyslexia had significantly lower factor scores and higher standard deviations than those without dyslexia. Third, prediction analyses showed that all factors contributed to the classification of those with and without dyslexia (with DA 92% correctly classified and with LRA 95% correctly classified).

Regarding the existence of subtypes of dyslexia, we found that 78% of those with dyslexia had a deficit in spelling and that 75% had more than one deficit. However, we could not distinguish any meaningful categorization of combinations of deficits. In addition, we found that about half of the not-identified students had more than one deficit (53%), whereas a very small minority had no deficit (5%).

Furthermore, we investigated relations between factors of dyslexia, severity of dyslexia, factors of intelligence (nonverbal intelligence, speed of [numeric] processing, vocabulary), and school grades. The main findings were that the spelling factor correlated negatively with the other factors of dyslexia only in the group of those with dyslexia and that severity of dyslexia correlated positively with spelling (meaning that high ability in language correlates with high ability in spelling) only in the group of those without dyslexia. Low correlations between factors of dyslexia and factors of intelligence were found in the group of those without dyslexia but not in the group of those with dyslexia. This seems to contradict the finding that those with dyslexia had significantly lower scores than those without dyslexia on all factors of intelligence. However, as expected, those with dyslexia had lower school grades in language courses, but not in mathematics and other courses.

General Evaluation of Results

The main question in this study was how to relate subtypes of dyslexia to the overwhelming quantity of symptoms and theories that have been described in more than 20 years of research. Ramus and Ahissar (2012) concluded that two subtypes of dyslexia can be distinguished in the literature – a majority subtype characterized by a phonological deficit and one or several minority subtypes characterized by a visual deficit – whereas additional subtypes of phonological and visual dyslexia might emerge from the consideration of underlying aetiologies. Furthermore, support has been found for a multiple cognitive deficit model as pro-posed by Pennington (2006) in three studies (Bosse, Tainturier, & Valdois, 2007; Di Filippo & Zoccolotti, 2012; Menghini et al., 2010) in which various cognitive abilities were related to reading difficulties of those with dyslexia. Le Jan et al. (2011) found that dyslexia and non-dyslexia can be identified with cognitive abilities without making use of reading or spelling assessments. The importance of the present study is that five features related to dyslexia could be extracted from a very large battery of tests and self-report questions that included both language abilities and general cognitive abilities related to dyslexia. This implied that the statistical approach chosen in this study was successful, which may encourage further exploration of accompanying features of dyslexia. Based on these five underlying features of dyslexia, we could not distinguish subtypes of dyslexia. However, future research with a similar approach may clarify the issue of subtypes and the possible existence of more features underlying dyslexia.

The main conclusion from the results in this study is that a multiple cognitive model provides a better explanation of dyslexia than the distinction between subtypes. However, the interpretations of the results of this study with regard to a multiple cognitive view of dyslexia are complex. For instance, we found that a majority of those with dyslexia were characterized by various combinations of deficits, but also that a majority of those with dyslexia were characterized by a spelling deficit. This may be surprising in a student sample. On the other hand, spelling deficits of those with dyslexia are known to persist into adulthood, even after years of remedial teaching. Furthermore, we should emphasize that more cognitive abilities may be involved in dyslexia than were found in the present study. For instance, due to limited testing time we were not able to assess various other abilities such as motor abilities and visuo-attentional abilities, although these abilities may have been involved implicitly in the tests that were conducted in this study. Nevertheless, this is the first study that extracted features of dyslexia independently from reading assessment and without making theoretical assumptions beforehand.

Although we did not find subtypes of dyslexia, this may be explained by the fact that the sample was too small in relation to the number of possible combinations of factorial deficits. Alternatively, it is possible that subtypes of dyslexia do not exist, just various combinations of various specific deficits, whereas the underlying nature of and relations between these deficits remain unclear. That some deficits (e.g., phonological) are generally reported more often than other deficits may be explained by selection bias in some studies, or by the fact that unknown deficits were not investigated or were interpreted as other deficits. More support for the view that distinct subtypes of dyslexia do not exist is given by the remarkable finding that in the group of those with dyslexia the most common factor, spelling, did not correlate with severity of dyslexia, whereas it correlated negatively with the other factors of dyslexia. Apparently, a combination of factors is a better predictor of severity in dyslexia than the most common deficit on its own, which was supported by the prediction analyses. Of importance here is that in the present study – in contrast to many similar studies – no participants were excluded, resulting in a small group of students who could not be identified as with or without dyslexia. Half of them were characterized by more than one deficit and most of the rest by one deficit. Maybe a group of people exists that cannot be identified as with or without dyslexia merely as a result of the fact that dyslexia might exhibit itself in only one deficit while compensating for other deficits.

In relation to the finding of more factors than in previous studies, we should emphasize that this study was different regarding three important methodological features: We used a large sample that consisted of well-educated students, the sample was acquired differently than in most other studies, and we conducted exploratory factor analyses not only on tests but also on many specific self-report questions. The finding of a relatively large number of factors probably resulted not only from using many tests and a large sample, but also from the way those with dyslexia were selected. Those with dyslexia in the study by Bosse, Tainturier and Valdois (2007) were recruited from education authorities and dyslexia centres. However, no specific details were reported regarding how dyslexia was assessed, apart from the fact that all of those with dyslexia were extremely delayed readers. The advantage of the way those with dyslexia were selected for the present study was that two methods were applied: a selection method based on biographical information and a selection method based on prediction analyses without having any preference toward different theories of dyslexia. This resulted in the identification of those with dyslexia of varying severity and based on many symptoms. In the sample of Bosse, Tainturier and Valdois, highly intelligent children with dyslexia may have been excluded and children without dyslexia with delayed reading may have falsely been included, which could explain that 23% of the children with dyslexia were found to have no deficit at all. In contrast, we found no children with dyslexia without a selective deficit (25% had only one deficit). Even in the group of not-identified students, we found only a very small group (5%) of students without a deficit.

Evaluation of Separate Factors

The most common deficit of those with dyslexia in this study is a spelling deficit. Apparently, spelling difficulties are challenging to overcome even in a sample of highly educated students and even after extensive training at school, which is common practice today for those with dyslexia in Dutch schools. Although there is nothing revolutionary about impaired spelling being the most general feature of dyslexia, this factor is difficult to relate to previous findings.

Bosse, Tainturier and Valdois (2007) found a visual attention factor that was a strong independent predictor of reading speed, even after controlling for single letter identification skills. Also Di Filippo and Zoccolotti (2012) did not find a factor that could clearly be related to spelling abilities. Thus, finding a spelling deficit is not new, but finding this deficit existing independently from phonological and other abilities is new according to our knowledge.

Probably the mostly investigated symptom of dyslexia is the phonological deficit. However, the exact nature of this deficit is still a topic for debate. Some researchers consider a phonological deficit as a core deficit of dyslexia that originates in the phonological lexicon (e.g., Snowling & Hulme, 2005). In contrast, others have found support for a phonological processing deficit that results from impaired retrieval from the lexicon, with the phonological representations in this lexicon itself being unimpaired (Blomert, Mitterer, & Paffen, 2004; Ramus & Szenkovits, 2008). This view was confirmed in a recent functional magnetic resonance imaging study (Boets et al., 2013) that showed that adults with dyslexia have intact but less accessible neural representations of speech sounds. Although the results of the present study cannot support either of these views, it is clear that phonology is a symptom of dyslexia.

A third unsurprising factor in this study is short-term memory (STM), which is a well-known and generally accepted symptom of dyslexia (e.g., de Jong, 1998). Working memory is generally understood as a dynamic mechanism assumed to consist of four parts: the central executive, a visuospatial sketch pad, a phonological store, and an episodic buffer (Baddeley, 1986, 2000, 2002). The related STM construct is held to describe systems solely involved in temporary storage of information (e.g., Baddeley, 1990, 2003). However, because we used the digit span tasks in our study, it was not possible to tell whether we were dealing with a deficit in storage or with a deficit of memory-related executive functions.

Two factors in this study have not been described before in the literature to our knowledge. One was rhyme/confusion, which represented the confusion factor in specific language-related questions and the test *Dutch–English Rhyme Words*. The other was whole-word processing/complexity, which represented the complexity factor in specific language-related questions and the test *Spoonerisms* and *Letter Order*. First, we considered it to be surprising that these last two tests did not load on the confusion factor. However, because support for complex cognitive aetiologies in dyslexia has become stronger in recent years, it should not surprise us that we found factors that we cannot explain based on the knowledge we have so far. Apparently, there are two processing factors that independently contribute to difficulties of those with dyslexia.

Dyslexia as characterized by processing deficits is in line with the finding of phonological processing difficulties and with the finding that the visual attention factor of Bosse, Tainturier and Valdois (2007) was a predictor of reading speed. Also, it is consistent with the finding of a processing factor in the study of Di Filippo and Zoccolotti (2012) regarding their pictorial naming factor. As mentioned in the introduction, they proposed that naming slowness refers to the ability to retrieve name codes from the semantic lexicon and is not specifically related to activation of the orthographic lexicon. Although it is purely speculative, we suggest that factors of naming slowness or impaired visual attention may be related or even the same as the two factors in this study. For the test *Dutch–English Rhyme Words*, both words must be named silently to determine whether these words sound the same or not. However, the underlying aetiology of this factor remains unclear because any causal direction is unclear. Confusion, such as exchanging letters, might influence naming speed, or impaired naming speed might result in confusion.

Regarding the whole-word processing/complexity factor, we should mention two reading models that have been of special interest in relation to dyslexia. The dual route model of reading (Coltheart et al., 2001; De Groot, Dannenburg & van Hell, 1994; Ziegler et al., 2008) and the multitrace memory model of polysyllabic word reading (Ans, Carbonnel, & Valdois, 1998) both postulate that there are two different strategies in reading. The analytical strategy is used by starting readers and for reading difficult new words: Words are analysed letter by letter and/ or phoneme by phoneme. The global strategy, or direct route, is mainly used for words that are well known: the word is recognized as a whole. If those with dyslexia make more use of the analytical route than normal readers, this can explain a marked word length effect found in those with dyslexia for both words and nonwords, suggesting the use of a letter-by-letter strategy (Marinus & de Jong, 2010). Furthermore, we should mention that it has been emphasized that deficits of visual attention span especially prevent putting attention to the whole word (Bosse, Tainturier & Valdois, 2007; Bosse & Valdois, 2009). Thus, when words become longer and more complicated, those with dyslexia may still persist in a letter-by-letter strategy whereas those without dyslexia maintain a global strategy. In this study, whole-word processing and complexity probably represented the same latent variable.

In summary, it is important to realize that it is not essential here to explain all factors theoretically. The main finding is that five psychometric factors of dyslexia could be distinguished that all seem to be related to some kind of cognitive information processing. Support for each factor can be found in the literature. However, underlying aetiologies remain unclear as long as no evidence has been found for causal relations between these factors.

Dyslexia, Intelligence, and Coping Strategies

In general, a disadvantage of using a student sample is that conclusions cannot be generalized to a general population. However, the advantage of using a student sample in studies of dyslexia is that low performances on tests cannot alternatively be explained by low intelligence, something that is a complicating factor in diagnosing dyslexia in young children. Thus, student samples complicate evaluations of conclusions about the relation among intelligence, schooling, coping strategies, and dyslexia.

In the present study, we found support for dyslexia as a multiple cognitive deficit, but we were not able to distinguish subtypes on the basis of various combinations of deficits. Here, we emphasize that this also might result from differences in coping strategies, especially in a sample of students. Students are smart and well educated and can therefore be expected to have developed advanced coping strategies. However, these coping strategies might differ individually. This is supported by the finding that the variability of the factor scores was higher in the group with dyslexia than in the group without dyslexia, in contrast to the variability of the factor scores of intelligence. Higher variability for those with dyslexia is a general finding in the literature, but here the contrast with intelligence measures is striking. One possible explanation we can proffer is that, within the group of those with dyslexia, separate symptoms are robust on an individual level. During childhood those with dyslexia may try to compensate for dyslexia-related abilities that are not very impaired, whereas other dyslexia-related abilities remain impaired. In other words, what is weak stays weak and what is not so weak becomes stronger. This might enhance negative correlations between, for instance, spelling and other factors of dyslexia, and thus the existence of those with dyslexia characterized by various combinations of deficits. This view supports the idea that causal subtypes do not exist, but merely symptoms that depend on individually different coping strategies.

The discussion about underlying aetiologies versus coping strategies in dyslexia is also relevant for the findings in this study regarding intelligence measures. Those with dyslexia had poorer performances on all three factors of intelligence (without showing higher variabilities), however not on school grades mathematics and other courses that are unrelated to language. Seemingly inconsistent with this is that in the group of those with dyslexia no correlations were found between factors of dyslexia and factors of intelligence, whereas there were some weak correlations in the group of those without dyslexia. We should be cautious drawing conclusions; students with dyslexia having relatively low scores on measures of intelligence cannot be generalized to a general population, because the average intelligence of students is much higher than that of the general population. Maybe at high levels of performance standard intelligence measures are not suitable for assessing differences between those with and without dyslexia. We found four arguments in favour of this explanation. First, it is remarkable that school grades of mathematics and language-unrelated courses did not differ between the groups. Second, standard intelligence measures may be influenced by dyslexia, also when performing on high levels of intelligence. Although intelligence measures are supposed to be unrelated to dyslexia, many intelligence tests partly depend on language skills, such as vocabulary, verbal analogies, and conclusions. Third, various coping strategies might influence intelligence measures. The finding of unknown factors in this study also means that we should account for unknown and maybe unconsciously driven coping strategies starting at very early ages during childhood that might affect intelligence measures as well. Fourth, unknown factors as found in this study might also directly affect nonverbal intelligence measures to some extent. For instance, in studies of dyslexia, groups are usually matched on intelligence often using Raven’s Progressive Matrices. However, poorer performances by those with dyslexia on this kind of test are often overlooked because group differences did not reach significance in studies using small samples. Assuming that dyslexia and intelligence are both characterized by multiple cognitive abilities, it would be strange to not have mutual influence.

Conclusion

In summary, we propose that dyslexia is characterized by various known and unknown cognitive deficits, whereas it remains unclear whether these deficits arise from different aetiologies or from different coping strategies. In future studies, it should be investigated whether the same features of dyslexia can be found in samples of children and general adult populations and with different tests. For future research, we propose that the aetiologies of factors of dyslexia should be investigated in relation to aetiologies of general intelligence. We conclude that as we dig deeper and deeper, we may start to see something that is much more complicated than we could ever have imagined. Maybe we should leave the classification of dyslexia as a disorder and instead accept dyslexia as an alternative way of information processing that has evolved over thousands of years without being noticed.

Acknowledgments

The authors thank Jan Hoogeboom for his tremendous efforts in the processing of data.

4

A new self-report inventory of dyslexia for students: Criterion and construct validity

The validity of a Dutch self-report inventory of dyslexia was ascertained in two samples of students. Six biographical questions, 20 general language statements and 56 specific language statements were based on dyslexia as a multi-dimensional deficit.

Dyslexia and non-dyslexia were assessed with two criteria: identification with test results (Sample 1) and classification using biographical information (Sample 1 and Sample 2). Using discriminant analyses, these criteria were predicted with various groups of statements. Overall, 11 discriminant functions were used to estimate classification accuracy of the inventory. In Sample 1, 15 statements predicted the test criterion with classification accuracy of 98%, and 18 statements predicted the biographical criterion with classification accuracy of 97%. In Sample 2, 16 statements predicted the biographical criterion with classification accuracy of 94%. Estimations of positive and negative predictive value were 89% and 99%.

Items of various discriminant functions were factor analysed to find characteristic difficulties of students with dyslexia, resulting in a five-factor structure in Sample 1 and a four-factor structure in Sample 2. Answer bias was investigated with measures of internal consistency reliability.

Less than 20 self-report items are sufficient to accurately classify students with and without dyslexia. This supports the usefulness of self-assessment of dyslexia as a valid alternative to diagnostic test batteries.

Keywords: Dyslexia; diagnosis; adults; screening

Based on: Tamboer, P., & Vorst, H. C. M. (2015). A new self-report inventory of dyslexia for students: Criterion and construct validity. *Dyslexia, 21*(1), 1-34.

Introduction

In the Netherlands, relatively cheap and standardized methods of diagnosing dyslexia in children have been developed, which can be assessed in schools and without the supervision of specialists (e.g. Vorst, 2006). Generally, diagnosing dyslexia in children is based on assessments of reading and/or spelling, cognitive performances (e.g. phonological), and resistance to intervention, usually in combination with the exclusion of other possible explanations such as low schooling, low intelligence or Attention Deficit Hyperactivity Disorder (ADHD). Controversy about the best diagnosing methods remains, although there is general agreement about the definition by the World Health Organization (2015) which characterizes dyslexia as a specific reading disorder characterized by specific and significant impairment in the development of reading skills that is not solely accounted for by mental age, visual acuity problems, or inadequate schooling.

For an adult population, however, the diagnosing methods used for children may not be appropriate because in contrast with children, adults may have compensated to some extent for various difficulties accompanying dyslexia. Alternatively, various self-rating scales of dyslexia have recently been proposed (e.g. Snowling et al., 2012). However, the usefulness of self-report inventories has not fully been explored yet. Therefore, we investigated in this study the possibility of diagnosing adult dyslexia with a new self-report inventory as an alternative to test batteries in a large student sample.

In general, a problematic issue for diagnosing dyslexia both in children and in adults is that underlying aetiologies of dyslexia are still unknown, which makes it hard to interpret poor and normal performances on tests of dyslexia (Ramus & Ahissar, 2012). An additional issue for the interpretation of performances of especially adults with dyslexia is that it is unknown to what extent various coping strategies influence these performances. Adults with dyslexia who never received specialized training at school, or who even might not be aware of being dyslexic, were forced to deal with all kinds of language-related difficulties, which might differ to a large extent on an individual level. These individual differences might depend on level of intelligence, level of education, socioeconomic status or any combination of these. And it is unknown how these individually different coping strategies evolve over the course of years. The main question of this study is whether these issues affect self-report questionnaires of dyslexia in the same way as they may influence test performances.

For many disorders in general, self-report inventories have proven to be reliable instruments in adult samples both for clinical diagnoses and for quantitative measurements. For example, the Beck Depression Inventory (Beck et al., 1961), one of the most widely used self-rating scales for measuring depression, was shown to have a high concurrent validity with high positive and negative predictive values (Furlanetto, Mendlowicz & Bueno, 2005). Moreover, not only disorders but also common psychological traits can reliably be measured with self-rating scales, such as alexithymia (Vorst & Bermond, 2001) and the Big Five personality traits (Soto & John, 2009). Because self-report inventories have proven to be highly reliable and valid as a classification instrument or as a rating scale, we felt that a thorough investigation of the predictive power of a self-assessment of dyslexia is warranted.

In several previous studies, support was found for the reliability and validity of self-assessment of dyslexia. Table 1 provides an overview of the main findings. Schulte-Körne, Deimel and Remschmidt (1997) found a high predictive validity of 12 general questions related to reading and spelling in a sample of reading/spelling disabled German adults. Lefly and Pennington (2000) found that a 23-item self-report of reading his-tory had a high predictive and convergent validity in a sample of English-speaking adults with reading disability diagnosis. Wolff and Lundberg (2003) found that a 20-item self-report of symptoms of dyslexia was a more powerful predictor of Swedish adults with and without dyslexia than various phonological tests. This self-report questionnaire consisted of two scales, reading interests and other symptoms of dyslexia. Mortimore and Crozier (2006) created a 13-item questionnaire that provided a thorough overview of perceived difficulties (e.g. note taking, organization of essays and expressing ideas in writing) of people with dyslexia in British higher education. Snowling et al. (2012) designed a 15-item questionnaire that rates reading, ADHD and related skills. They found a four-factor structure (reading, word finding, attention and hyperactivity) with the reading factor correlating with measured skills of spelling and decoding fluency in English-speaking adults. Bjornsdottir et al. (2013) found a three-factor structure (dyslexia symptoms, current reading and memory) in a 23-item self-report of reading history in Icelandic adults. They found that these questions quite reliably identified those with and without dyslexia according to an extern specialist. The general finding of predictive value of self-report questions is what we found as well in a previous study (Tamboer, Vorst, & Oort, 2014), in which some self-report statements contributed to the identification of students with and without dyslexia.

Table 1

Overview of previous studies

Study	Number and nature of questions	Factor structure	Reliability	Sample	Criterion	Classification accuracy	Convergent validity
Schulte-Körne, Deimel & Remschmidt (1997)	12 general questions reading / spelling (two sum scores)			elderly adults (German) 24 reading disabled 55 controls	family genetic study	discriminant analysis: sensitivity 71% specificity 98% CART analysis: sensitivity 87% specificity 91%	
Lefly & Pennington (2000)	23 general questions reading / spelling (5-point Likert scale)		Cronbach`s alpha = .94 test-retest: .87	elderly adults (English) 22 reading disabled 40 controls	Familial dyslexia	discriminant analysis: sensitivity 81.8% specificity 77.5%	correlation with reading quotient *r* = .70
Wolff & Lundberg (2003)	20 statements dyslexic symptoms (4-point scale)	PCA: two components: 1. dyslexia symptoms (14 items) 2. reading interest (6 items)	Component: 1. alpha = .84 2. alpha = .81	students (Swedish) 50 dyslexic / reading disabled 67 controls	teacher rating in education centre	more powerful discriminator than phonological tests	
Mortimore & Crozier (2006)	13 general questions (yes/no) Chi-square differences			students (English) 62 dyslexics 74 controls	diagnosed by educational psychologist		
Snowling et al. (2012)	21 general questions (5-point or yes/no)	Conf. factor analysis 1. reading 2. word finding 3. attention 4. hyperactivity	Factors: 1. alpha = .81 2. alpha = .60 3. alpha = .81 4. alpha = .58	adults (English) 170 family history of dyslexia 91 families of children with language difficulties 156 typical developing		logistic regression sensitivity (poor readers): 62.5% specificity (normal readers): 95.0%	correlations between reading and tests of spelling (.60) word reading (.51) nonword reading (.66) low correlations with other factors and tests
Bjornsdottir et al. (2013)	23 questions (translated from Lefly & Pennington) (5-point Likert scale)	expl. factor analysis: 1. dyslexia symptoms 2. current reading 3. memory	Cronbach`s alpha = .92 test-retest .93 419 SRD: .88 679 controls: .89	young adults (Icelandic) 419 specific reading disorder 679 nondyslexics	diagnosed as a child (M 15.5 SD 7.5) by specialist in developmental neuropsychology	ROC curve analysis: sensitivity 95.5% specificity 88.8% ROC curve analysis: sensitivity 84.5% specificity 83.7%	

The aim of this study was to further investigate reliability and validity of self-assessment of dyslexia. Although all previous studies reported relatively high validity, findings were reported based on different criteria of dyslexia or reading difficulties, and analysed with different small sets of self-report questions. Both of these differences depend on the operational definition of dyslexia, which is used. While the definition of the World Health Organization (2015) is focused on reading difficulties, other definitions are also based on, for example, phonological difficulties (e.g. Snowling et al, 1997). Furthermore, while it is still unknown what causes dyslexia, it is also unknown to what extent subtypes of dyslexia should be distinguished (Bosse, Tainturier & Valdois, 2007). Many symptoms of dyslexia have been reported (e.g. Ramus & Ahissar, 2012), and especially in adult samples, it may be expected that coping strategies are based on various combinations of symptoms and differ on an individual level. It is unclear how this affects diagnoses of dyslexia in adult samples. With regard to these issues, we distinguished two main issues: the issue of criterion validity and the issue of construct validity.

Regarding criterion validity in samples of students, Mortimore and Crozier (2006) stated, '… the researcher is dependent on the criteria that are adopted by each higher education institution, and can expect that the students will have been diagnosed by a large number of professionals applying a range of different tests and procedures' (p. 239). In other words, when predictive validity of any diagnostic instrument is investigated, classification accuracy can only objectively be interpreted when an objective criterion is used. To circumvent the problem that different choices can be made for a criterion of dyslexia, we investigated two different criteria of dyslexia. One criterion was based on the previous study of ours, in which people with and without dyslexia were identified with a large battery of tests. A second criterion was based on a biographical self-report. Although we considered the first criterion to be more reliable, we hypothesized that both criteria should suffice for high criterion validity.

Regarding construct validity, Mortimore and Crozier (2006) acknowledged that dyslexic students' difficulties are not 'restricted to reading, spelling and writing, but may be experienced across the range of tasks that students encounter in higher education' (p. 236). However, these researchers collected various difficulties of dyslexics with only 13 items. Also in the other studies, the number of items was small, ranging from 12 to 23. The main feature of all these items is that they assess general difficulties of people with dyslexia. For instance, it is clear that the question 'Can you read quickly and easily?' (Snowling et al., 2012) assesses reading. However, reading requires several complex cognitive abilities of which it is unknown how they are related to other symptoms of dyslexia such as phonological or visual impairments.

Nevertheless, three studies (Bjornsdottir et al., 2013; Snowling et al., 2012; Wolff & Lundberg, 2003) reported factor structures in the small sets of items, although the number of factors is small (two or three) compared with the assumed number of symptoms involved in dyslexia. In order to cover as much symptoms of dyslexia as possible, we used items of the *Communication Questionnaire* (Kramer & Vorst, 2007), which assesses various language difficulties related to dyslexia according to a facet structure. In this study, we used – besides 20 general statements comparable with the statements or questions of the previous studies – 56 of these items. The main characteristic of these items is that they do not assess general language difficulties but, instead, very specific language difficulties. The idea was that the language difficulties of dyslexics are very specific and subtle, and may depend on different combinations of different symptoms of dyslexia (e.g. phonological, visual and attentional), which may be overlooked when only general statements or questions are used. We hypothesized that using a large numbers of these specific items, a factor structure might be found consisting of many factors, which would increase construct validity.

For this study, we created a new self-report inventory of dyslexia, which consists of three parts: 6 biographical questions (BQ), 20 general language statements (GLS) and 56 specific language statements (SLS). With these items, we performed prediction analyses in two samples of students and investigated classification accuracy. In the first sample, predictions were based on two criteria, one based on test results and one based on biographical information. In the second sample, only a biographical criterion was available. Furthermore, we investigated reliabilities and performed factor analyses on the sets of items that discriminated best between students with and without dyslexia.

Method

Participants

We used two samples of Dutch first-year psychology students. In both samples, we excluded students with missing variables, students who had a history of moderate to severe head trauma or severe general health problems, students with other disorders such as ADHD, students who were not ‘completely Dutch’ (we required that students were raised and went to school in the

Netherlands with Dutch as their mother language) and students of whom the data showed answer patterns. Furthermore, we selected only students with ages between 17 and 25 years.

Sample 1 consisted of 418 students (104 males, mean age 19.7 [1.5] years): 67 dyslexics (20 males, mean age 20.0 [1.7] years) and 351 non-dyslexics (84 males, mean age 19.6 [1.5] years). Sample 2 consisted of 405 students (114 males, mean age 20.6 [1.4] years). Dyslexia and non-dyslexia in this sample was investigated in this study. In Sample 1, the prevalence of dyslexia may seem relatively large (67/418 = 16%). However, prevalence of dyslexia may have been underestimated to some extent, especially in a group of psychology students. In the Netherlands, psychology is considered to be a relatively easy study as compared with most other studies. Although it has not been investigated, it seems plausible that students with dyslexia prefer a relatively easy study over difficult studies.

Criteria of dyslexia

We determined two criteria of dyslexia and non-dyslexia. A ‘test criterion’ was determined in order to validate the self-inventory with an independent and reliable criterion that is based on standard assessments of various symptoms of dyslexia. We used the data of 10 tests that were collected in a previous study (Tamboer, Vorst, & Oort, 2014). These tests are described in the next paragraph. We classified all students as dyslexic or not dyslexic in a procedure consisting of two steps. First, we determined that 27 students had a formal diagnosis of dyslexia. In the Netherlands, children with dyslexia can acquire an official document that states that they are dyslexic. This document can only be provided by a specialized educational psychologist and is based on an extended testing procedure. In the Netherlands, this document is widely accepted to be a reliable diagnosis of dyslexia. In the second step, we compared the performances of this dyslexic group with the performances of the remaining students on the whole battery of 10 tests. We applied two discriminant and two logistic regression analyses on the tests. One discriminant analysis and one logistic regression analysis were based on sum scores, and one discriminant analysis and one logistic regression analysis were based on item scores. Students were categorized as dyslexic when all analyses predicted a student to be dyslexic. The 27 students with an official document were all predicted to be dyslexic according to our test battery, which validates this battery as a reliable instrument for categorizing students with dyslexia (see also, Tamboer, Vorst, & Oort, 2014). Of the remaining students, 40 were categorized as dyslexic. Students were categorized as non-dyslexic when all analyses predicted a student to be non-dyslexic, which resulted in a group of 351 students.

A ‘biographical criterion’ was determined in order to establish whether it is possible to diagnose dyslexia with a self-inventory without making use of additional tests. In the present study, a biographical criterion was based on six biographical self-report questions. One question was ‘Are you dyslexic?’ Five other questions were self-assessments of reading, spelling and writing difficulties as a child and in the present (questions 6-10 in Appendix A). All questions had three answer categories (yes = 2, doubt = 1 and no = 0). A sum score was calculated by adding the six scores, which resulted in a score ranging from 0 to 12. Cut-off scores were determined in this study and are described in the Results section. We determined criterion groups of dyslexia and non-dyslexia using a cut-off score that was based on the lowest estimation of prevalence using the test criterion, which was 14% (see also Table 2 in the Results section). We studied the distribution of the biographical score and found that 55 students (13.2%) had a score of 5 or higher and that 75 students (17.9%) had a score of 4 or higher. We decided to use the conservative cut-off score of 5 or higher for dyslexic students (55) and 4 or lower for non-dyslexic students (363). Of the 55 dyslexic students, 51 (93%) were also identified as dyslexic with the test criterion. And of the 363 non-dyslexic students, 347 (96%) were also identified as non-dyslexic with the test criterion. Thus, 398 of 418 students (95%) were identically identified with both criteria.

Assessment of dyslexia: a test battery

The test battery, used for a criterion of dyslexia, consisted of ten tests. In the Netherlands, diagnosing test batteries usually consists of measurements of six abilities that are assumed to be impaired in people with dyslexia: reading words and/or pseudowords, spelling, phonological awareness, short-term memory, rapid naming and visual attention. A problematic issue for the present study was that some tests used in these batteries are suitable only for children but not for highly educated students. A second problematic issue was that some tests require individual testing, as is the case with rapid naming tests. We were not able to individually test more than 400 students. We therefore decided to adjust existing tests, originally designed for children, so that the tests are suitable for students and so that all tests can be assessed in large groups of students without the requirement of individual observers. In a follow-up study (Study 4, Chapter 5), we validated the test battery of the present study with a test battery that is used in clinical settings in a small sample of students (dyslexia and non-dyslexia based on official records) that performed both test batteries. We compared predictions of both batteries and found consistent classifications in about 95% of the cases in both the dyslexic and the non-dyslexic groups of students.

All tests are described below. We did not categorize these tests into the six categories, because we assumed that some tests may require various cognitive abilities. For example, pseudoword reading is assumed to require phonological abilities, but it cannot be ruled out that abilities of attention or short-term memory are required for good performances as well.

Dutch Dictation consists of 10 sentences in the Dutch language (maximum score 10 × 4 = 40) and aims to measure spelling abilities. This test was a reconstruction of various existing tests for children but with a higher level of spelling difficulties. Each sentence was read aloud and could be heard through headphones. The whole sentence had to be typed into the computer. Each sentence consisted of a few words that are vulnerable to spelling errors.

English Dictation consists of 10 sentences in the English language (maximum score 10 × 2 = 20) and aims to measure spelling abilities. This test was a reconstruction of various existing tests for children but with a higher level of spelling difficulties. Each sentence was read aloud and could be heard through headphones. At the same time, this sentence could be read on the computer screen except two words. These words had to be retyped into the computer. Although Dutch students are familiar with the English language, some English words are well known for their vulnerability to spelling errors for Dutch people.

Missing Letters consists of 10 sentences in the Dutch language (maximum score 10 × 2 = 20) and aims to measure reading and spelling abilities as well as abilities of attention. This test was a reconstruction of various existing tests for children but with a higher level of spelling difficulties. Each sentence was read aloud and could be heard through headphones. For each sentence, two words were repeated, while these words were shown on the computer screen with a few letters left out of both words. The missing letters had to be typed into the computer.

Pseudowords consists of 30 pseudowords (maximum score 30), which are non-words that sound like real words and aims to measure reading and phonological abilities and possible other abilities such as attention. Each pseudoword was read aloud and could be heard through headphones. At the same time, this pseudoword was presented on the computer screen. It had to be decided whether the visually presented pseudoword was spelled correctly, which was the case for half of all pseudowords. Usually pseudowords are administered the other way around with participants reading the words aloud themselves. The reason for changing this was that it would be practically impossible to get all students in private sessions for this test.

Sound Deletion consists of 20 difficult Dutch words (maximum score 20) and aims to measure phonological abilities. The test was a reconstruction of various existing tests for children but with a higher level of difficulties. Each word was read aloud and could be heard through headphones. Some of these words were pronounced correctly and some words incorrectly by leaving out or adding one letter. On the computer screen, each word was presented three times, each time with a slightly different spelling with one of them being spelled correctly according to what was pronounced. Participants had to decide which of the visually presented words was heard through the headphones. For example, the existing word ‘fietsenstalling’ – which means bicycle shed – was read out as ‘fiestenstalling’. The possible answers were ‘fietsentalling’, ‘fiestensalling’ and ‘fiestenstalling’.

Spoonerisms consists of 20 words (maximum score 20) and aims to measure phonological abilities. The test was a reconstruction of various existing tests for children but with a higher level of difficulties. A Spoonerism is a word that consists of two existing smaller words and that still consists of two small existing words when the first letters of both small words are exchanged. Each word was read aloud and could be heard through headphones. For example, the word ‘kolen-schop’ had to be altered into ‘scholen-kop’ and typed into the computer.

Incorrect Spelling consists of 40 Dutch words (maximum score 40) and aims to measure spelling and reading abilities and abilities of rapid attention. This test was designed for this study. All words were presented on a computer screen for 50ms each. Half of the presented words were spelled correctly, and half were spelled incorrectly. The participants had to decide whether the words were spelled correctly.

Dutch–English Rhyme Words consists of 40 Dutch–English word pairs (maximum score 40), presented on the computer screen with the Dutch words on the right, and aims to measure abilities of attention and phonology. This test was designed for this study. For half of the word pairs, the nouns of the Dutch and the English word resembled each other visually but not aurally. For the other half of the word pairs, the nouns of the Dutch and the English word resembled each other aurally.

Letter Order requires the ability to read words as a whole and consisted of 20 Dutch sentences (maximum score 20 × 2 = 40; time limit of 5 min) and aims to measure abilities of whole word reading. This test was designed for this study. Theoretical hypotheses about reading words as a whole are described in the dual route model of reading (for an extended description, see De Groot, Danneburg & van Hell, 1994). The idea for this test is based on a well-known text: ‘Aoccdrnig to rscheearch at Cmabrigde uinervtisy, it deosn’t mttaer waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteres are at the rghit pclae. The rset can be a tatol mses and you can sitll raed it wouthit a porbelm. Tihs is bcuseae we do not raed ervey lteter by it self but the wrod as a wlohe’. We created Dutch sentences based on the same principle; the order of letters in the words was changed, apart from the first and last letters. The sentences became more difficult towards the end of the test. The sentences had to be typed in with all words correctly spelled.

The *Short-Term Memory Test* aims to measure the capacity of short-term memory and was based on various existing tests of digit span capacity. The digit span is the number of digits a person can retain and recall. There are four subtests: numbers and letters, both forward and backward. And each subtest consists of 24 series: 6 of 4, 6 of 5, 6 of 6 and 6 of 7 items for the subtests numbers and letters forward, and 6 of 3, 6 of 4, 6 of 5 and 6 of 6 items for the subtests numbers and letters backward. The numbers and letters are presented one by one, for 1 s each on a computer screen. The participants have to retype these numbers and letters after the last one of a series has been presented. About half of all series consist of some typical difficulties for dyslexic students, either phonological or visual or both. For example, a typical phonological confusion is between the numbers seven and nine, which resemble each other phonologically in Dutch (zeven/negen). Typical visual confusions are between the numbers six and nine and the letters [m] and [w]. The letters [p], [d] and [b] resemble each other phonologically as well as visually.

Procedure

All data of the first sample were collected at the University of Amsterdam in 2009 and 2010 during five sessions of 3h each, in which tests and questionnaires of various studies were administered. These sessions took place on midweek evenings with one or two weeks between each session. All students were obligated to participate because these sessions were part of the first-year study programme. The students were informed about the general nature of the tests and the questionnaires in advance according to a standard protocol. Afterwards, the students received a more detailed debriefing. Anonymity was guaranteed by the standard protocol of the University of Amsterdam. The students had up to three weeks after their debriefing to request that their test results were not used. The data of the second sample were collected in 2011 in one session and under the same conditions.

Self-Report Inventory

The self-report inventory of this study consisted of three parts: 6 BQ, 20 GLS and 56 SLS. See Appendix A. There are two differences regarding the two samples. In the first sample, the three types of questions were collected in separate and larger questionnaires consisting of more items, while in the second sample, a selection of all questions was presented in one questionnaire. And the students of the first sample did not know that the questions were aiming to measure dyslexic difficulties (except for the BQ), while in the questionnaire for the second sample, it was mentioned that at least some of the questions were aiming to measure dyslexia.

Biographical Self-Report Questions

Biographical information regarding dyslexia is assessed with six questions. The first question is a direct self-report question: ‘Are you dyslexic?’ Five questions request to assess to what extent performances of spelling, reading and writing were poor in the past and in the present.

General Language Statements

In our first study (Tamboer, Vorst, & Oort, 2014), we created a questionnaire that consisted of 60 general statements about reading, speaking, writing, mental representations, memory and foreign languages. We selected the 20 best discriminating statements based on group differences. For each statement, there are seven response categories (7-point Likert scale).

Specific Language Statements

In our first study, we assessed the Questionnaire Communication (Kramer & Vorst, 2007) that consists of 140 statements about specific language-related difficulties. The questionnaire was designed according to a 7 × 5 × 4 facet design with seven response categories (7-point Likert scale). One dimension distinguishes between seven different aspects of how language is used in daily life or at school and universities: reading, writing, speaking, listening, copying, taking a dictation and reading aloud. A second dimension distinguishes between five different levels of how language can be represented: by sounds, letters, words, sentences or text. A third dimension distinguishes between four different difficulties that accompany dyslexic adults: skipping (forgetting), adding, changing and exchanging. For instance, dyslexic students may skip parts of sentences when reading, exchange letters when writing, change words when speaking or forget parts of texts when taking a dictation.

Support for impairments related to all aspects of the first and second dimensions can be found in the literature. The difficulties of the third dimension need some clarification. They are related to some kind of attention or organization skills or concentration, which were found to be impaired in people with dyslexia in higher education (Mortimore & Crozier, 2006). These researchers presented an overview of difficulties that very much resemble the difficulties represented by the three dimensions of the Questionnaire Communication. They distinguished 13 items representing reading, reading speed, spelling, note taking, organizing essays, general organization, time keeping, expressing ideas orally, expressing ideas in writing, handwriting, concentration, remembering facts and listening. All difficulties were reported significantly more often by those with dyslexia than by those without dyslexia.

For the present study, we selected 56 items with a high predictive value while preserving the original facet design as much as possible.

Tests of Intelligence

For being able to analyse convergent and divergent validity of the questionnaire, we needed additional information about performances on tests of intelligence. Available were the data of six cognitive tests that were based on Guilford's Structure of Intellect Model (*Vocabulary*, *Verbal Analogies*, *Conclusions*, *Numeric Progressions*, *Speed of Calculation* and *Hidden Figures*) and for general (non-verbal) intelligence (*Advanced Progressive Matrices Set 2*, Raven, Court, & Raven, 1979).

Characteristics of Samples 1 and 2

In Sample 1, the three parts of the self-report inventory were administered on three different days. The part consisting of the GLS and the part consisting of the SLS were both administered without the participants knowing that the statements were aiming to assess dyslexia. Furthermore, data of dyslexia tests and intelligence tests were available. For analyses of criterion validity, criterion groups were determined with two criteria of dyslexia, a test criterion and a biographical criterion.

In Sample 2, the three parts of the self-report inventory were administered together. In contrast to the participants of Sample 1, the participants of Sample 2 were aware that the statements were aiming to assess dyslexia. Data of dyslexia tests and intelligence tests were not available for Sample 2. For analyses of criterion validity, criterion groups were determined with only a biographical criterion.

Analyses

We investigated criterion and construct validity in five steps. In the first step, we looked for differences between dyslexic and non-dyslexic students on sum scores of the 20 GLS, the 56 SLS and all statements together. Furthermore, we looked for differences between students with and without dyslexia on each individual item.

In the second step, we performed predictions with five different sets of potential predictors: 6 BQ, 20 GLS, 56 SLS, the 76 GLS and SLS together, and all 82 items together. In Sample 1, we investigated classification accuracy of these predictions with a test criterion and a biographical criterion. In Sample 2, we investigated classification accuracy with a biographical criterion. Next, we validated the best discriminant functions of Sample 1 in Sample 2. All predictions were carried out with discriminant analysis (stepwise method) in SPSS. We also investigated whether results differed with or without the possibility of cross-validation, the leave-one-out procedure in SPSS.

In the third step, we compared the results of the prediction analyses with measures of internal consistency. We calculated Cronbach's alpha, the Spearman–Brown coefficient and the Guttman Split-Half coefficient of all 76 items and of the 20 GLS and the 56 SLS separately. We did this in each sample for the whole group, the dyslexic and non-dyslexic groups. We also calculated the mean of all items' means and the mean of all items' variance.

In the fourth step, we used the items that were selected for various discriminant functions for exploratory factor analyses. In another previous study (Tamboer, Vorst & Oort, 2016), we found a five-factor structure in a confirmatory factor model with five factors representing variance in the ten tests. We tested whether this five-factor structure (spelling, phonology, short-term memory, rhyme/confusion and whole word processing/complexity) could be replicated.

In the fifth step, we investigated convergent and divergent validity with correlations between sum scores of the self-report inventory and sum scores of the tests and intelligence measures.

Results and Conclusions

Group differences

We calculated group differences between the dyslexic (67) and non-dyslexic (351) groups of Sample 1 (according to a test criterion). First, we calculated group means and standard deviations of the sum scores of the 20 GLS, the 56 SLS and all items together. Students with dyslexia had lower scores than students without dyslexia on all three sets of items ($p < 0.001$). Second, we calculated group means and standard deviations of all individual items. Students with dyslexia had lower scores than students without dyslexia on all items ($p < 0.05$), except for seven items of the 20 GLS (nos. 1, 5, 7, 11, 15, 17 and 20) and two items of the 76 SLS (nos. 23 and 26). The items are presented in Appendix A.

Criterion validity

Predictions in Sample 1 with test criterion

In Sample 1, we first investigated criterion validity with a test criterion and five sets of items as potential predictors. With each set, a discriminant function (DF) was determined, which maximally separated the test-criterion groups of 67 students with dyslexia and 351 students without dyslexia. DF1 was acquired with 6 BQ as potential predictors, DF2 was acquired with 20 GLS as potential predictors, DF3 was acquired with 56 SLS as potential predictors, DF4 was acquired with 76 GLS and SLS together as potential predictors, and DF5 was acquired with all 82 items together as potential predictors. All discriminant functions are presented in Appendix B. Table 2 presents classification accuracies of these discriminant functions in five cross-tables. For each DF, specific details are presented. First, it is specified how the DF was generated: the number of potential predictors, the criterion that was used and the criterion groups. Second, the results of the analysis are specified: the number of selected predictors in the DF, the percentage correctly classified students with and without dyslexia and the estimated prevalence of dyslexia. Third, classification accuracy is presented in a cross-table. The left percentage above represents sensitivity (number of correctly classified students with dyslexia), and the right percentage below represents specificity (number of correctly classified students without dyslexia). Completely below, the total percentages of predicted students with and without dyslexia are presented.

Table 2

Predictions in Sample 1 with test-criterion

D = dyslexic
ND = non-dyslexic
BQ = biographical questions
GLS = general language statements
SLS = specific language statements

Discriminant function 1 was generated with:
Potential predictors: 6 BQ
Criterion of dyslexia: test battery
Criterion groups: 67 dyslexics & 351 non-dyslexics

Results:
Number of selected predictors in discriminant function: 5 BQ
Percentage correctly classified of 418 students: 94%
Estimated prevalence of dyslexia: 15%

Classification of 67 dyslexics and 351 non-dyslexics:

	Predicted as:		
	D	ND	Total
D	50 (78%)	17 (22%)	67
ND	6 (2%)	345 (98%)	351
	56 (15%)	362 (85%)	418

Discriminant function 2 was generated with:
Potential predictors: 20 GLS
Criterion of dyslexia: test battery
Criterion groups: 67 dyslexics & 351 non-dyslexics

Results:
Number of selected predictors in discriminant function: 11 GLS
Percentage correctly classified of 418 students: 88%
Estimated prevalence of dyslexia: 21%

Classification of 67 dyslexics and 351 non-dyslexics:

	Predicted as:		
	D	ND	Total
D	53 (79%)	14 (18%)	67
ND	35 (10%)	316 (90%)	351
	88 (21%)	316 (90%)	418

Discriminant function 3 was generated with:
Potential predictors: 56 SLS
Criterion of dyslexia: test battery
Criterion groups: 67 dyslexics & 351 non-dyslexics

Results:
Number of selected predictors in discriminant function: 12 SLS
Percentage correctly classified of 418 students: 94%
Estimated prevalence of dyslexia: 19%

Classification of 67 dyslexics and 351 non-dyslexics:

	Predicted as:		
	D	ND	Total
D	61 (91%)	6 (9%)	67
ND	17 (5%)	334 (95%)	351
	78 (19%)	340 (81%)	418

Discriminant function 4 was generated with:
Potential predictors: 20 GLS + 56 SLS
Criterion of dyslexia: test battery
Criterion groups: 67 dyslexics & 351 non-dyslexics

Results:
Number of selected predictors in discriminant function: 6 GLS + 11 SLS
Percentage correctly classified of 418 students: 97%
Estimated prevalence of dyslexia: 17%

Classification of 67 dyslexics and 351 non-dyslexics:

	Predicted as:		
	D	ND	Total
D	64 (96%)	3 (4%)	67
ND	8 (2%)	343 (98%)	351
	72 (17%)	346 (83%)	418

Discriminant function 5 was generated with:
Potential predictors: 6 BQ + 20 GLS + 56 SLS
Criterion of dyslexia: test battery
Criterion groups: 67 dyslexics & 351 non-dyslexics

Results:
Number of selected predictors in discriminant function: 4 BQ + 2 GLS + 9 SLS
Percentage correctly classified of 418 students: 98%
Estimated prevalence of dyslexia: 14%

Classification of 67 dyslexics and 351 non-dyslexics:

	Predicted as:		
	D	ND	Total
D	58 (87%)	9 (13%)	67
ND	1 (0%)	350 (100%)	351
	59 (14%)	359 (86%)	418

DF4 and DF5 are the best functions for predicting dyslexia and non-dyslexia according to a test criterion. DF4 (6 GLS + 11 SLS) resulted in classification accuracy of 97%, with sensitivity of 96% and specificity of 98%, and with an estimated prevalence of dyslexia of 17%. DF5 (4 BQ + 2 GLS + 9 SLS) resulted in classification accuracy of 98%, with sensitivity of 87% and specificity of 100%, and with an estimated prevalence of dyslexia of 14%. Lower classification accuracies were found for DF3 (12 SLS, 94%) and DF2 (11 GLS, 88%). DF1 (5 BQ) resulted in classification accuracy of 94%. The importance of this last finding is that a test criterion is relatively consistent with a biographical criterion.

We conclude that using the combination of GLS and SLS (and BQ) as predictors results in a classification of students with and without dyslexia which is highly consistent with a classification using test results. An important remark is that we only reported classification results in the tables that were acquired without using a cross-validation procedure (leave-one-out). We did this because we want to use the same discriminant formulas that were acquired in Sample 1 in Sample 2, thus validating the results of Sample 1 to Sample 2. Nevertheless, we also repeated the best predictions, DF4 and DF5, using a cross-validation procedure (leave-one-out). For DF4, we found that two more dyslexic students and two more non-dyslexic students were classified incorrectly, leading to classification accuracy of 96%, a reduction of 1%. For DF5, we found that one more dyslexic student and one more non-dyslexic student were classified incorrectly, leading to classification accuracy of 97%, a reduction of 1%. These differences are so small that we conclude that validity is not threatened when no cross-validation is applied.

Predictions in Sample 1 with biographical criterion

An extended criterion, such as the test criterion in this study, is not always available. Therefore, we wanted to investigate whether an easy to acquire biographical criterion can be used for a reliable classification of dyslexia and non-dyslexia. We investigated criterion validity with this biographical criterion in the same way as we did with the test criterion. Three sets of items were used as potential predictors. With each set, a discriminant function was determined, which maximally separated the 55 dyslexic and 363 non-dyslexic students according to the biographical criterion. DF6 was acquired with 20 GLS as potential predictors, DF7 was acquired with 56 SLS as potential predictors and DF8 was acquired with 76 GLS and SLS together as potential predictors. All discriminant functions are presented in Appendix B. We investigated classification accuracy in two different ways.

First, we examined classification accuracy in the dyslexic and non-dyslexic groups as they were selected by the biographical criterion. Table 3 presents classification accuracies in three cross-tables. For each discriminant function, specific details are presented in the same way as in Table 2. DF8 is most consistent with the classification based on a biographical criterion. DF8 (3 GLS + 15 SLS) resulted in classification accuracy of 97%, with sensitivity of 96% and specificity of 97%, and with an estimated prevalence of dyslexia of 15%. When we repeated the analysis with a cross-validation procedure (leave-one-out), we found slightly lower classification accuracy of 95%.

Second, we repeated these analyses but now using dyslexic and non-dyslexic students as they were selected by the test criterion. In other words, when we assume that the test criterion is more reliable than the biographical criterion, we should establish whether the dyslexic and non-dyslexic students according to the test criterion are correctly classified. This should be considered as a better test of validity than looking how the dyslexic and non-dyslexic students according to the biographical criterion were classified because the predictions are completely independent from the criterion. Nevertheless, that 95% of the 418 students were identified identically with both criteria means that the use of a biographical criterion for prediction analyses is probably justified. Table 4 presents classification accuracies in three cross-tables. For each discriminant function, specific details are presented in the same way as in Tables 2 and 3. DF8 is the best function for predicting dyslexia and non-dyslexia according to a test criterion. DF8 (3 GLS 15 SLS) resulted in classification accuracy of 95%, with sensitivity of 84% and specificity of 98%, and with an estimated prevalence of dyslexia of 15%.

Findings and conclusions (Sample 1)

With the analyses using a test criterion, we found that DF4 (6 GLS + 11 SLS) and DF5 (4 BQ + 2 GLS + 9 SLS) resulted in highest classification accuracy (97% and 98%) with an estimation of prevalence of dyslexia of 17% and 14%. With the biographical criterion, we found that using all items as potential predictors, thus DF8 (3 GLS + 15 SLS), resulted in the highest classification accuracy (98%) with an estimation of prevalence of dyslexia of 15%.

We considered the test criterion to represent a more reliable classification of dyslexic and non-dyslexic students than the biographical criterion. Thus, the importance of Table 4 is that it shows that – using a biographical criterion – the self-report inventory reliably identifies most of the ‘true’ dyslexic and non-dyslexic students. On the other hand, when a test criterion

Table 3

Predictions in Sample 1 with biographical criterion
(Classification of biographical criterion groups)

Discriminant function 6 was generated with:
Potential predictors: 20 GLS
Criterion of dyslexia: biographical criterion
Criterion groups: 55 dyslexics & 363 non-dyslexics

Results:
Number of selected predictors in discriminant function: 12 GLS
Percentage correctly classified of 418 students: 86%
Estimated prevalence of dyslexia: 22%

Classification of 55 dyslexics and 363 non-dyslexics:

	Predicted as:		Total
	D	ND	
D	43 (78%)	12 (22%)	55
ND	47 (13%)	316 (87%)	363
	90 (22%)	328 (78%)	418

Discriminant function 7 was generated with:
Potential predictors: 56 SLS
Criterion of dyslexia: biographical criterion
Criterion groups: 55 dyslexics & 363 non-dyslexics

Results:
Number of selected predictors in discriminant function: 16 SLS
Percentage correctly classified of 418 students: 94%
Estimated prevalence of dyslexia: 17%

Classification of 55 dyslexics and 363 non-dyslexics:

	Predicted as:		Total
	D	ND	
D	50 (91%)	5 (9%)	55
ND	20 (6%)	343 (94%)	363
	70 (17%)	348 (83%)	418

Discriminant function 8 was generated with:
Potential predictors: 20 GLS + 56 SLS
Criterion of dyslexia: biographical criterion
Criterion groups: 55 dyslexics & 363 non-dyslexics

Results:
Number of selected predictors in discriminant function: 3 GLS + 15 SLS
Percentage correctly classified of 418 students: 97%
Estimated prevalence of dyslexia: 15%

Classification of 55 dyslexics and 363 non-dyslexics:

	Predicted as:		Total
	D	ND	
D	53 (96%)	2 (4%)	55
ND	11 (3%)	352 (97%)	363
	64 (15%)	354 (85%)	418

Table 4

Predictions in Sample 1 with biographical criterion
(Classification of test-criterion groups)

Discriminant function 6 was generated with:
Potential predictors: 20 GLS
Criterion of dyslexia: biographical criterion
Criterion groups: 55 dyslexics & 363 non-dyslexics

Results:
Number of selected predictors in discriminant function: 12 GLS
Percentage correctly classified of 418 students: 89%
Estimated prevalence of dyslexia: 22%

Classification of 67 dyslexics and 351 non-dyslexics:

	Predicted as:		Total
	D	ND	
D	55 (82%)	12 (18%)	67
ND	35 (10%)	316 (90%)	351
	90 (22%)	328 (78%)	418

Discriminant function 7 was generated with:
Potential predictors: 56 SLS
Criterion of dyslexia: biographical criterion
Criterion groups: 55 dyslexics & 363 non-dyslexics

Results:
Number of selected predictors in discriminant function: 16 SLS
Percentage correctly classified of 418 students: 94%
Estimated prevalence of dyslexia: 17%

Classification of 67 dyslexics and 351 non-dyslexics:

	Predicted as:		Total
	D	ND	
D	57 (85%)	10 (15%)	67
ND	13 (4%)	338 (96%)	351
	70 (17%)	348 (83%)	418

Discriminant function 8 was generated with:
Potential predictors: 20 GLS + 56 SLS
Criterion of dyslexia: biographical criterion
Criterion groups: 55 dyslexics & 363 non-dyslexics

Results:
Number of selected predictors in discriminant function: 3 GLS + 15 SLS
Percentage correctly classified of 418 students: 95%
Estimated prevalence of dyslexia: 15%

Classification of 67 dyslexics and 351 non-dyslexics:

	Predicted as:		Total
	D	ND	
D	56 (84%)	11 (16%)	67
ND	8 (2%)	343 (98%)	351
	64 (15%)	354 (85%)	418

would not have been available, the importance of Table 3 is that it shows that the self-report inventory is highly consistent with a biographical criterion.

We found that both a biographical criterion and a test criterion resulted in high classification accuracies, especially when all items are used as potential predictors. Specific differences can be found when we compare Table 2 with Table 4. In other words, comparing these tables means comparing classification accuracy in the same dyslexic and non-dyslexic groups, but with two different criteria. Regarding the GLS as potential predictors, we should compare the results of DF2 with the results of DF6. There are hardly any differences. Regarding the SLS as potential predictors, we should compare the results of DF3 with the results of DF7. The only difference here is that sensitivity is higher when the test criterion is used. Regarding all items as potential predictors, we should compare the results of DF4 with the results of DF8. Also here, the main difference is that sensitivity is higher when the test criterion is used. The conclusion here is that validity (sensitivity) is higher when a test criterion is used. In other words, with the inventory, most of the non-dyslexic students (98%) are always correctly identified, but dyslexic students are better identified when a test criterion is available than when only a biographical criterion is available.

Another way to interpret the results is to consider positive predictive value (the proportion of true dyslexic students of the total number of predicted dyslexic students) and negative predictive value (the proportion of true non-dyslexic students of the total number of predicted non-dyslexic students). We compared DF4 with DF8 (GLS and SLS as potential predictors). DF4 (test criterion) resulted in a positive predictive value of (64 / 72 =) 89% and a negative predictive value of (343 / 346 =) 99%. DF8 (biographical criterion) resulted in a positive predictive value of (56 / 64 =) 88% and a negative predictive value of (343 / 354 =) 97%. Thus, when the inventory is assessed and only a biographical criterion is available, one can conclude that of the classified dyslexic students, 88% is really with dyslexia and that of the classified dyslexic students, 97% is really without dyslexia.

Summarizing, there are four conclusions. (1) Highest validity is acquired when all items are used as potential predictors. (2) Both a test criterion and a biographical criterion can be used for reliable classifications of people with and without dyslexia, although sensitivity is higher using a test criterion. (3) With both criteria, negative predictive value is higher than positive predictive value. (4) Based on the most reliable discriminant functions, estimations of prevalence in the sample of this study vary between 14% and 17%.

Predictions in Sample 2 with biographical criterion

In Sample 2, a test criterion was not available. A biographical criterion was determined in the same way as in Sample 1. In Sample 1, 55 students (13.2%) had a score of 5 or higher and 363 students (86.8%) had a score of 4 or lower. In Sample 2, 52 students (12.8%) had a score of 5 or higher and 353 students (87.2%) had a score of 4 or lower. The percentages are about the same in both samples. Therefore, we used these groups as biographical criterion groups.

Three sets of items were used as potential predictors. With each set, a discriminant function was determined, which maximally separated the 52 students with dyslexia and 353 students without dyslexia according to the biographical criterion. DF9 was acquired with 20 GLS as potential predictors, DF10 was acquired with 56 SLS as potential predictors and DF11 was acquired with 76 GLS and SLS together as potential predictors. We investigated classification accuracy with these discriminant functions. Table 5 presents all cross-tables. All discriminant functions are presented in Appendix B. For each discriminant function, specific details are presented in the same way as in Tables 2, 3 and 4.

DF11 is most consistent with the classification based on a biographical criterion. DF11 (5 GLS + 11 SLS) resulted in classification accuracy of 94%, with sensitivity of 92% and specificity of 94%, and with an estimated prevalence of dyslexia of 17%. When we repeated the analysis with a cross-validation procedure (leave-one-out), we found slightly lower classification accuracy of 91%. We concluded that the self-report inventory reliably predicts biographical criterion groups of dyslexia and non-dyslexia (94%, cross-validated 91%). However, without an alternative criterion such as a test criterion, we cannot know whether the false positives and false negatives represent false classifications of the inventory or false classifications of the criterion. It is therefore not meaningful to calculate positive predictive value and negative predictive value.

Predictions in Sample 2 by validating discriminant functions from Sample 1

Although a test criterion was not available in Sample 2, we could use the discriminant functions that were acquired in Sample 1 with a test criterion as an extern criterion in Sample 2. The best predictions in Sample 1 with the test criterion were derived with DF4 (20 GLS and 56 SLS as potential predictors) and DF5 (6 BQ, 20 GLS and 56 SLS as potential predictors). We validated these two discriminant functions in Sample 2. Thus, in Sample 1, we acquired two discriminant functions that will be tested on new cases, the students of Sample 2. Summarizing, we wanted to know whether the best classification possible with the inventory was consistent with an external criterion. The best classification in Sample 2 was

Table 5

Predictions in Sample 2 with biographical criterion

Discriminant function 9 was generated with:
Potential predictors: 20 GLS
Criterion of dyslexia: biographical criterion
Criterion groups: 52 dyslexics & 353 non-dyslexics

Results:
Number of selected predictors in discriminant function: 8 GLS
Percentage correctly classified of 405 students: 88%
Estimated prevalence of dyslexia: 22%

Classification of 52 dyslexics and 353 non-dyslexics:

	Predicted as:		Total
	D	ND	
D	46 (88%)	6 (12%)	52
ND	42 (12%)	311 (88%)	353
	88 (22%)	317 (78%)	405

Discriminant function 10 was generated with:
Potential predictors: 56 SLS
Criterion of dyslexia: biographical criterion
Criterion groups: 52 dyslexics & 353 non-dyslexics

Results:
Number of selected predictors in discriminant function: 10 SLS
Percentage correctly classified of 405 students: 90%
Estimated prevalence of dyslexia: 19%

Classification of 52 dyslexics and 353 non-dyslexics:

	Predicted as:		Total
	D	ND	
D	45 (87%)	7 (13%)	52
ND	33 (9%)	320 (91%)	353
	78 (19%)	327 (81%)	405

Discriminant function 11 was generated with:
Potential predictors: 20 GLS + 56 SLS
Criterion of dyslexia: biographical criterion
Criterion groups: 52 dyslexics & 353 non-dyslexics

Results:
Number of selected predictors in discriminant function: 5 GLS + 11 SLS
Percentage correctly classified of 405 students: 94%
Estimated prevalence of dyslexia: 17%

Classification of 52 dyslexics and 353 non-dyslexics:

	Predicted as:		Total
	D	ND	
D	48 (92%)	4 (8%)	52
ND	21 (6%)	332 (94%)	353
	69 (17%)	336 (83%)	405

Table 6

Predictions in Sample 2 by validating discriminant functions from Sample 1

Discriminant function 4:

Potential predictors:	20 GLS + 56 SLS
Criterion of dyslexia:	test battery in Sample 1
Criterion groups:	67 dyslexics & 351 non-dyslexics of Sample 1
Number of selected predictors in discriminant function:	6 GLS + 11 SLS

Results:

Percentage correctly classified of 405 students:	85%

Classification of 69 dyslexics and 336 non-dyslexics:

	Predicted as:		Total
	D	ND	
D	53 (77%)	16 (23%)	69
ND	45 (13%)	291 (87%)	336

Discriminant function 5:

Potential predictors:	6 BQ + 20 GLS + 56 SLS
Criterion of dyslexia:	test battery in Sample 1
Criterion groups:	67 dyslexics & 351 non-dyslexics of Sample 1
Number of selected predictors in discriminant function:	4 BQ + 2 GLS + 9 SLS

Results:

Percentage correctly classified of 405 students:	80%

Classification of 69 dyslexics and 336 non-dyslexics:

	Predicted as:		Total
	D	ND	
D	57 (83%)	12 (17%)	69
ND	70 (21%)	266 (79%)	336

the one derived with DF11 (5 GLS + 11 SLS), resulting in 69 students with dyslexia and 336 students without dyslexia.

Table 6 presents two cross-tables. For each discriminant function, specific details are presented in the same way as in the previous tables. We found a consistency of 85% between the classification with a biographical criterion and the classification with a discriminant function (DF4), which was acquired with Sample 1. There are two interpretations of this result. One is that it is shown that DF4 can be applied in a new sample and that the reliability of this discriminant function is relatively high (85%). A second interpretation is that the self-report inventory results in a relatively reliable classification of dyslexic and non-dyslexic students without the necessity of using an external criterion.

Conclusions

From the three analyses, we can summarize the following conclusions. Regarding the different criteria of dyslexia, we conclude that both a test criterion and a biographical criterion can be used for reliable prediction analyses, although validity of a test criterion was found to be higher than validity of a biographical criterion, especially regarding sensitivity. Regarding the different sets of predictors, we conclude that the combined set of GLS and SLS provided the highest classification accuracy and that not more than 20 items are needed for a reliable classification of people with and without dyslexia. Reliability of these sets was relatively successfully validated in Sample 2. However, two things are not completely clarified yet.

First, it remains unclear what the best criterion of dyslexia is. We speculated that a test criterion is more reliable than a biographical criterion. However, this cannot be proven from the results of this study. Nevertheless, in a new sample, we found that classifications with both criteria were consistent in 85% of all students. We conclude that criterion validity of the self-report inventory is relatively high, and how high depending on the interpretation of criteria but at least 85%.

Second, the relative success of the different sets of predictors might depend on the differences between the two samples. In Sample 1, the GLS and the SLS were administered on separate days and without mentioning that dyslexia was assessed. In Sample 2, the GLS and the SLS were administered in one questionnaire, while the participants were aware that the statements were aiming to assess dyslexia. This might have resulted in somewhat lower classification accuracy of 85% by validation than classification accuracies acquired within one sample. A possibility is that answer bias may have differed between the samples and between the GLS and SLS.

Internal consistency reliability

In previous studies (Table 1), mainly general statements were used. It is further-more difficult to establish whether and how answer bias may have affected self-report questions or statements of dyslexia. Four studies reported measures of internal consistencies, which were all between 0.80 and 0.94, except a factor ‘word finding’ (0.60) in the study of Snowling et al. (2012). We found in our study that the combination of general and specific statements resulted in the highest criterion validity but that specific statements have more predictive power than general statements. In the previous paragraph, we suggested that differences between the two samples may have resulted in a decrease of classification accuracy, such as different patterns of answer bias. We further investigated this issue by analysing internal consistency reliability of the separate sets of potential predictors in the two samples.

We investigated internal consistency reliability with three separate measures: Cronbach's alpha, the Spearman–Brown coefficient and Guttman's Split-Half coefficient. These measures were calculated for the three sets of potential predictors, for Samples 1 and 2 separately, and for the dyslexic and non-dyslexic students separately and together. Furthermore, we determined the means of item mean and item standard deviation. We had three expectations. First, we expected higher values resulting from answer bias in Sample 2, because the purpose of the inventory was known in this sample. Second, we expected higher values for the SLS than for the GLS, because for the SLS, we carefully explored symptoms of dyslexia, while it is unknown whether GLS represent only difficulties that are characteristic for dyslexia. (This suggestion is supported by the fact that 74 of 76 SLS showed group differences but that 7 of 20 GLS did not show group differences.) Third, we expected higher values for those without than for those with dyslexia. This was based on the idea that especially the SLS can be assumed to have a factor structure. Those with dyslexia may be characterized by different combinations of symptoms as was shown in the previous study (Tamboer, Vorst, & Oort, 2016), while those with non-dyslexia are characterized by none of the symptoms.

Table 7 provides all indices. Regarding the first expectation (higher values for Sample 2 than for Sample 1), we found higher values for the set of GLS in Sample 2 than in Sample 1, but the values for the set of SLS hardly differed between the two samples. This implies that knowledge about the purpose of the statements affects answer patterns on the GLS but not on the SLS. Regarding the second expectation (higher values for the SLS than for the GLS), we found higher values for the SLS than for the GLS in both samples, while this difference was very large in Sample 1. This implies that the SLS probably measure the same construct, while the GLS probably measure different constructs. Regarding the third expectation (higher values for non-dyslexic than for dyslexic students), we found that non-dyslexic students had higher values than dyslexic students in most cases. This implies that variability is higher in the dyslexic group compared with the non-dyslexic group, which was confirmed by the values of the mean of item variances. This might support the existence of a factor structure with those with dyslexia showing different combinations of response patterns on these factors.

All expectations were confirmed by the results. The SLS are less vulnerable than the GLS to answer patterns, while they probably better represent the same construct than the GLS. A factor structure may influence internal consistency reliability for the different

Table 7

Internal consistency reliabilities (20 GLS, 56 SLS, 76 GLS + SLS)

	Cronbach's Alpha	Spearman-Brown	Guttman Split-Half	Item Mean (Mean)	Item SD (Mean)
20 GLS					
Sample 1					
Whole group (n = 418)	0.50	0.62	0.62	4.93	2.42
D test-criterion (n = 67)	0.46	0.54	0.54	4.41	3.13
ND test-criterion (n = 351)	0.34	0.50	0.50	5.03	2.20
D biographical criterion (n = 55)	0.44	0.53	0.51	4.37	3.22
ND biographical criterion (n = 363)	0.37	0.52	0.52	5.01	2.22
Sample 2					
Whole group (n = 405)	0.83	0.81	0.80	4.91	2.65
D biographical criterion (n = 52)	0.70	0.71	0.70	3.77	3.35
ND biographical criterion (n = 353)	0.78	0.72	0.72	5.08	2.29
56 SLS					
Sample 1					
Whole group (n = 418)	0.97	0.94	0.93	5.42	1.63
D test-criterion (n = 67)	0.93	0.87	0.87	4.58	1.93
ND test-criterion (n = 351)	0.96	0.92	0.92	5.58	1.37
D biographical criterion (n = 55)	0.93	0.87	0.87	4.56	1.98
ND biographical criterion (n = 363)	0.96	0.93	0.92	5.55	1.41
Sample 2					
Whole group (n = 405)	0.97	0.93	0.92	5.51	2.15
D biographical criterion (n = 52)	0.95	0.86	0.85	4.26	2.78
ND biographical criterion (n = 353)	0.96	0.91	0.90	5.69	1.75
76 GLS + SLS					
Sample 1					
Whole group (n = 418)	0.95	0.89	0.86	5.29	1.84
D test-criterion (n = 67)	0.91	0.75	0.74	4.54	2.25
ND test-criterion (n = 351)	0.94	0.87	0.82	5.43	1.59
D biographical criterion (n = 55)	0.90	0.71	0.69	4.51	2.31
D biographical criterion (n = 363)	0.94	0.88	0.84	5.41	1.62
Sample 2					
Whole group (n = 405)	0.97	0.94	0.94	5.35	2.28
D biographical criterion (n = 52)	0.95	0.88	0.87	4.13	2.93
D biographical criterion (n = 353)	0.96	0.92	0.91	5.53	1.89

dyslexic and non-dyslexic groups. We conclude that the SLS are more reliable predictors of dyslexia than the GLS, which confirms the conclusions regarding criterion validity. There are two remarks. First, internal consistency tends to be high in general when using many items as is the case here. Second, high values generally indicate that the same construct is measured by most items. However, the high indices of the SLS that were found here do not mean that the scale is unidimensional. The indices are a bit lower for dyslexic students than for non-dyslexic students, which results from higher item variances and points to the possibility of factor structures in the set of SLS.

Construct validity

We hypothesized that dyslexia is most reliably predicted when it is taken into account that dyslexia is characterized by various symptoms. We therefore created a self-report inventory consisting of GLS and SLS. These SLS are based on a facet design with different facets representing different symptoms. We found that these SLS were better predictors of dyslexia than the GLS, that they were less vulnerable to answer bias and that analyses of internal consistency reliabilities supported the existence of a factor structure. Based on these findings, we further investigated construct validity of the self-report inventory by investigating whether the five-factor structure of the previous study (Tamboer, Vorst, & Oort, 2016) could be confirmed. In this previous study, we found that five components with eigenwaardes above 1 explained 60% of variance in a large battery of tests and questionnaires. We named these components: spelling, phonology, short-term memory, rhyme/confusion and whole word processing/complexity.

In previous studies of self-report questionnaires (Bjornsdottir et al., 2013; Snowling et al., 2012; Wolff & Lundberg, 2014), construct validity was investigated by performing exploratory factor analyses on all items that were available. In the present study, a principal components analysis of all 76 items resulted in 15 components in Sample 1 and in 14 components in Sample 2. A better solution was to perform exploratory factor analyses only on the items that were selected with various discriminant functions. When dyslexia and non-dyslexia were predicted with a small set of items, which are hypothesized to represent various symptoms of dyslexia, a factor structure should be visible even in these small sets of items. We only report the results of the functions that were acquired with the SLS and the combined set of GLS and SLS, because predictive validity and reliability of the GLS alone was relatively low. We performed exploratory factor analyses on the items that were selected for each discriminant function separately. In Table 8, all results are summarized. DF3, DF4 and DF5 were derived in Sample 1 and validated in Sample 2 so that analyses could be performed in both samples. All components have eigenwaardes above 1.

Table 8

Exploratory factor analysis: number of components and amount of variance explained

DF = discriminant function
GLS = general language statements
SLS = specific language statements

	Sample	N components	Percentage variance explained
DF3 (12 SLS, selected with test criterion):	1	2	53%
	2	2	57%
DF4 (6 GLS + 11 SLS, selected with test criterion):	1	5	60%
	2	3	54%
DF5 (4 BQ + 2 GLS + 9 SLS, selected with test criterion):	1	2	53%
	2	2	51%
DF7 (16 SLS, selected with biographical criterion):	1	3	56%
DF8 (3 GLS & 15 SLS, selected with biographical criterion):	1	4	54%
DF10 (10 SLS, selected with biographical criterion):	2	1	42%
DF11 (5 GLS & 11 SLS, selected with biographical criterion):	2	4	57%

The most important finding was that one discriminant function with high criterion validity (DF4) showed the highest construct validity. This function (acquired on the basis of a test criterion) consisted of a set of 17 items (6 GLS + 11 SLS) and exhibited a five-factor structure explaining 60% of all variance (all components with eigenvalues above 1). This result confirms the result of our previous study (Tamboer, Vorst, & Oort, 2016), in which we found a five-factor structure explaining 60% of variance as well. We studied the rotated component matrix of the five-factor structure found in the present study and decided that the following preliminary names for the five components were justified.

Component 1: whole word processing/complex words
Component 2: phonology
Component 3: confusion
Component 4: spelling/reading
Component 5: short-term memory/paying attention

There is some clear similarity between the two factor structures. Whole word processing, phonology, spelling/reading and short-term memory were found in both structures. It is however unclear how paying attention is related to short-term memory. Furthermore, rhyme and confusion are different constructs, although they may just as well represent the same construct. The main question now is why other discriminant functions resulted in less than five components. We further investigated the five-factor structure of DF4 by comparing this structure with the factor structures of other functions by correlating the factor scores of DF4 with factor scores of the other functions in Sample 1.

Correlations between the five-factor scores of DF4 and the four-factor scores of DF8 in Sample 1 could tell us something about how the use of a test criterion and a biographical criterion resulted in different factor structures. We found three high and significant (at 0.01 level) correlations: 0.74 between component 1 (whole word processing/complex words), 0.92 between component 2 (phonology) and 0.60 between component 4 (spelling/reading). Component 3 showed only a few low but significant (at 0.01 level) correlations, and component 5 of DF4 showed no correlations with components of DF8. This means that the fifth component (short-term memory/paying attention) is found when dyslexic and non-dyslexic students are selected with a test criterion but not when they are selected with a biographical criterion.

Correlations between the five-factor scores of DF4 and the two factor scores of DF3 in Sample 1 could tell us something about how the use of only SLS as predictors and the combined set of GLS and SLS as predictors resulted in different factor structures. We found two high and significant (at 0.01 level) correlations: 0.91 between component 1 (whole word processing/complex words) and 0.95 between component 2 (phonology). This means that using only the SLS as predictors resulted in a selection of items that represent whole word processing/complex words and phonology, but not confusion, spelling/reading and short-term memory/paying attention.

In Sample 2, DF4 resulted in a three-factor structure, explaining 54% of variance. An interpretation of the rotated component matrix learned that comparable items as in the other set of items are differently organized and more difficult to interpret. However, DF11 (acquired with a biographical criterion) resulted in a four-factor structure, explaining 57% of variance. An interpretation of the rotated component matrix here learned that interpretation was possible and that components of phonology, confusion, reading and short-term memory could be distinguished, while whole word processing/complex words were not clearly represented in this set of items.

Summarizing, we found support for a five-factor structure in a set of 17 predictors (6 GLS and 11 SLS) that were part of a discriminant function that was derived with a test criterion of dyslexia in Sample 1. With a biographical criterion, only four factors were found. Using only SLS as predictors resulted in only two factors. We conclude that different numbers of factors may originate from differences in answer patters, differences between GLS and SLS and differences between criteria used for the selection of dyslexic and non-dyslexic students. That the five-factor structure was found in Sample 1 makes sense because in this sample, it was unknown that dyslexia was assessed. This might have resulted in more variable answer patterns between items. That the five-factor structure was found with a test criterion of dyslexia makes sense because this criterion was considered to be more reliable, which might have contributed to the prevalence of those with dyslexia with more variable symptoms. This is consistent with the finding in the previous paragraph that the groups of non-dyslexics showed higher internal consistency values than the groups of dyslexics. This difference was most prominent for the comparison between those with and without dyslexia according to the test criterion, which implies that those with dyslexia are characterized by variable combinations of symptoms leading to variable answer patterns.

Convergent and Divergent Validity

Construct validity was further investigated with convergent validity and divergent validity. Of the six previous studies, only Snowling et al. (2012) thoroughly investigated convergent and divergent validity. As measurement of dyslexia, these researchers computed factor scale scores by summing the scores on the questions that had been identified as defining each factor. For comparisons, they furthermore used five raw scores on various measures. To be able to compare our results with those of the study of Snowling et al., we performed a similar analysis. Because test scores were only available in Sample 1, our analyses were restricted to this sample. As measurement of dyslexia, we only used the 56 SLS because the GLS showed low internal consistencies. We summed these items according to the original factors (spelling, phonology, short-term memory, rhyme/confusion and whole word processing/complexity) as were acquired in our previous study (Tamboer, Vorst, & Oort, 2016). We performed correlational analyses between these scores and various test scores. See Table 9. Language-related tests are assumed to measure one or more symptoms of dyslexia. General cognitive tests are usually assumed to measure cognitive abilities that are not related to dyslexia. All coefficients above 0.10 (and one of 0.09) are statistically significant because of the large sample size. However, Snowling et al. stated, '… it is the pattern of correlations which is important.' (p. 8).

Regarding convergent validity, Snowling et al. found high or moderate correlations between two scales (Reading Scale and Word Finding Scale) and three measures of language abilities (Spelling, Word Reading and Non-word Reading). We found mostly low and a few moderate correlations. The highest correlation was between the scale Spelling and *Dutch Dictation* (0.43). A considerable amount of correlations above 0.20 was found between all factor scales and most language-related tests, except *Sound Deletion*.

Regarding divergent validity, Snowling et al. found low correlations between the two scales and a measure of block design. We used seven measures of cognitive abilities that are often regarded to be independent from dyslexia. Nevertheless, most scales correlated with *Vocabulary*, *Verbal Analogies*, *Speed of Calculation* and *Hidden Figures*, of which the last one is most surprising. However, the scales did not correlate at all (except for three coefficients of 0.10) with *Conclusions*, *Numeric Progressions* and *Raven Progressive Matrices*.

The overall pattern of the results is that low and a few moderate significant correlations were found between the scales of the inventory and measures that represent language abilities, while no correlations were found between the scales of the inventory and measures of general intelligence. Although there are a few unexpected coefficients (correlations between *Speed of Calculation* and all scales of the inventory; very low correlations between *Sound Deletion* and the scales of the inventory), the general conclusion is that both convergent and divergent validity are relatively high.

Table 9

Correlations between sum scores of the 56 SLS and test scores (N = 418)

	Spelling	Phonology	Short-term memory	Rhyme / Confusion	Whole-word-processing / Complexity
Language related tests					
Dutch dictation	0.43**	0.22**	0.30**	0.21**	0.31**
English dictation	0.33**	0.20**	0.24**	0.22**	0.28**
Missing letters	0.36**	0.25**	0.31**	0.26**	0.32**
Pseudowords	0.19**	0.18**	0.21**	0.13**	0.17**
Sound deletion	0.11*	0.11*	0.10*	0.05	0.10*
Spoonerisms	0.21**	0.16**	0.21**	0.14**	0.18**
Incorrect spelling	0.41**	0.20**	0.30**	0.26**	0.32**
Dutch–English rhyme words	0.34**	0.21**	0.33**	0.25**	0.32**
Short-term memory numbers	0.35**	0.28**	0.33**	0.32**	0.37**
Short-term memory letters	0.35**	0.26**	0.32**	0.31**	0.36**
General cognitive tests					
Vocabulary	0.23**	0.14**	0.23**	0.15**	0.24**
Conclusions	0.05	-0.03	0.10*	0.07	0.08
Numeric Progressions	0.09	0.05	0.10*	0.04	0.10*
Speed of Calculation	0.20**	0.09*	0.16**	0.12**	0.16**
Verbal Analogies	0.19**	0.11*	0.17**	0.15**	0.18**
Hidden Figures	0.14**	0.07	0.14**	0.13**	0.17**
Raven Progressive Matrices	0.05	-0.02	0.08	0.05	0.08

*** Correlation is significant at the 0.01 level (2-tailed)*
** Correlation is significant at the 0.05 level (2-tailed)*

Discussion

The aim of this study was to examine whether a self-report inventory of dyslexia is suited to use as a diagnostic instrument for identifying dyslexia and non-dyslexia. We found that both a test criterion and a biographical criterion can be used for reliable predictions with high criterion and construct validity. The best predictions are acquired when both general and specific language self-report statements are used as predictors. Not more than 20 items were needed. Our findings support findings from previous studies (e.g. Bjornsdottir et al., 2013; Snowling et al., 2012). Some alterations compared with these studies were found to be successful. However, some methodological issues are still not completely resolved yet. We will discuss this study from three perspectives: methodological, theoretical and practical.

Methodological: predictive validity

Using two different criteria of dyslexia and using less than 20 language statements as predictors, we found classification accuracies of more than 95%, although a validation of the best discriminant function from one sample to an-other sample resulted in lower classification accuracy of 85%. We suggested that sample differences may have reduced classification accuracy. In Sample 1, the GLS and the SLS were administered on separate days and without mentioning that dyslexia was assessed. In Sample 2, the GLS and the SLS were administered in one questionnaire, while the participants were aware that the statements were aiming to assess dyslexia. We investigated possible differences between samples in answer bias with analyses of internal consistency reliability. Indeed, we found that values of internal consistency reliability of the general statements were higher in Sample 2 compared with Sample 1. This may have resulted in lower classification accuracy when predictions of one sample are validated in a second sample.

Although we found high criterion validity, interpretation only makes sense when the nature of different criteria is accounted for. In each classification study of dyslexia, the question is what to trust more: a criterion or a prediction. When dyslexia is considered to be a disorder characterized by reading and spelling difficulties and when self-report statements only cover these difficulties, high criterion validity can be expected. In this study, we used a test criterion that was based on the view of dyslexia as a multi-dimensional cognitive deficit. We furthermore created self-report statements on the basis of a facet design that reflects many different symptoms. The most compelling improvement in our study was that especially statements about specific difficulties showed high predictive value, although a combination of general and specific statements resulted in the highest classification accuracies.

Nevertheless, also in this study, it cannot be determined with absolute certainty whether the test criterion or the best prediction represents the true dyslexic and non-dyslexic students. Comparing predictions with a test criterion, we found a negative predictive value of 99%, which means that almost all predicted non-dyslexic students are non-dyslexic according to a test battery as well. We found a positive predictive value of 89%, which means that 11% of the predicted dyslexic students are not dyslexic according to a test battery. However, the dyslexic students in our sample are intelligent, are highly educated and – in most cases – received remedial teaching at school. Some students may have developed coping strategies resulting in high performances on tests. Thus, it might just as well be that the self-report inventory is more reliable than a test battery.

Theoretical: construct validity

Construct validity of the self-report inventory appears to be good. The best classifying discriminant function consisted of 17 items, which exhibited a five-factor structure explaining 60% of all variance. We named the components as whole word processing/complex words, phonology, confusion, spelling/reading and short-term memory/paying attention. The reliability of this structure is supported by the fact that in a previous study (Tamboer, Vorst, & Oort, 2016), a five-factor structure explaining 60% of variance was found as well. Construct validity was further supported by convergent and divergent validity. Sum scores that were based on the five-factor structure were correlated with measures of various tests that are known to be influenced by dyslexia and measures of various tests that are known to be unrelated to dyslexia. Although correlations supporting convergent validity were only moderate to low (but significant), the discrepancy with the very low correlations representing divergent validity was clear.

The five-factor structure was found using a test criterion. With a biographical criterion, a four-factor structure was found in both samples. Correlational analyses between the various factors of the different structures suggested that one component (short-term memory/paying attention) could not be identified with a biographical criterion. This may be explained by the fact that deficits of paying attention or short-term memory are not well recognized as symptoms of dyslexia by highly educated dyslexic students.

All together, we factor analysed the items of six different discriminant functions. Interpretations of these factors are difficult because some factors were only represented by a few items, while some items were related to more than one component of the analysis. Thus, we cannot say for sure whether the preliminary names are correctly representing the same construct. However, a theoretical conclusion here can be that both criterion validity and construct validity are good using both a test criterion and a biographical criterion and that any criterion should be based on more than spelling and reading difficulties alone.

Practical: extern validity

For purposes of diagnosing dyslexia, useful indices are the positive and negative predictive values. We found that negative predictive values were higher (97–99%) than positive predictive values (88–89%). This means that when a person is predicted as non-dyslexic, he or she is non-dyslexic with a very high probability. However, when a person is predicted as dyslexic, there is a chance of about 11–12% that this is not true. Thus, when the outcome of the self-report inventory is that a person is dyslexic, this person should be told that there is no absolute certainty and recommended to take steps for further investigation.

An important feature of this study is that the criteria did not exclude students based for instance on test performances. In other words, we did not distinguish between severely and moderately impaired dyslexic students. This supports the extern validity of the results. However, a shortcoming of this study is that we only used students who are highly intelligent, highly educated and raised by parents with, on average, high socioeconomic status. It thus remains unclear whether the same predictions can be applied on a more general population. On the other hand, using students may also have provided good insights in construct validity compared with when a general population would have been used. Students may have developed highly advanced coping strategies, which might remain undetected in a general population. In this study, we hypothesize that in the case of dyslexia, self-report statements may be less vulnerable to effects of coping strategies or answer bias than tests. The idea is that any self-assessment of perceived difficulties may be carried out by comparing these difficulties to other cognitive skills and to the skills of friends and family with the same intelligence, schooling and socioeconomic status. Self-report statements do not depend on intelligence, although the statements should be so that they are understandable for everybody.

Conclusion

We found support for the diagnostic value of self-report assessment of dyslexia. Highest classification accuracy was achieved with a criterion that was based on an extended test battery and that was consistent with biographical information. Construct validity was supported with factor analyses exhibiting a five-factor structure. For practical purposes, an important finding was that also only biographical information can provide a reliable and valid criterion of dyslexia. We found that self-report statements of dyslexia with the highest predictive value are characterized by specific descriptions of language difficulties, while more general statements suffer from answer bias.

Appendix A

Biographical questions

1. Are you dyslexic?
2. Did you, as a child, experience difficulties in reading?
3. Did you, as a child, experience difficulties in spelling?
4. Do you currently experience difficulties in spelling?
5. Do you currently experience difficulties in reading?
6. Do you currently experience difficulties in writing?

General language statements

1. At school I preferred languages.
2. Every week I read a book.
3. As a child I did not like reading.
4. I am a quick reader.
5. When pronouncing difficult words, I stutter.
6. I often mix up expressions.
7. I find it easy to explain something to someone.
8. I experience difficulties recalling words.
9. I rarely experience difficulties in finding the correct spelling.
10. When I am writing, I often exchange letters.
11. I easily understand rules of grammar.
12. I find it difficult to write in an organised manner.
13. I like to play word games.
14. I can easily remember faces.
15. At school I always paid attention.
16. I find serial learning easy.
17. I easily learn a foreign language.
18. I experience difficulties in reading in English.
19. I do not like that English sounds differently than it is in written.
20. When I have to write in English, I experience difficulties in finding the right spelling of words.

Specific language statements

Reading

1. I sometimes skip a letter and read a different word.
2. I have no problems with reading difficult Dutch words.
3. Because I exchange some things during reading, I understand something differently.
4. When I read a difficult sounding word for the first time, I do not experience any difficulties.
5. Words which look alike when written, I rarely exchange during reading.
6. When a sentence has a subordinate clause, I can easily read it.
7. I read a different word that sounds almost the same as the written word, but that has a different meaning.
8. Sometimes when I read a text, it appears that I have read more than is written.

Writing

1. When I am writing, I know when to use a ‘v’ or an ‘f’.
2. When I am writing something, I accidentally added superfluous words which made the text hard to understand for others.
3. When I am writing something, I rarely exchange letters (for instance ‘eo’ i.p.v. ‘oe’).
4. When I am writing something, I rarely forget to write certain words.
5. When I am writing down a word, I rarely think about its spelling although I know that I often write the word incorrectly.
6. Words such as ‘universitair’ and ‘principieel’ are not difficult for me to write.
7. When I am writing a sentence, I sometimes repeat parts of it.
8. When I am writing a sentence, words are in mixed order.

Speaking

1. When I pronounce a difficult sound for the first time, I make no mistakes.
2. During speaking, I exchange sounds such as ‘psi’ and ‘spi’.
3. During a conversation, it is not hard for me to pronounce new words correctly.
4. During speaking, I have a correct choice of words.
5. During a conversation, I use complex sentences correctly.
6. When I speak, I exchange some parts of sentences.
7. During a conversation I sometimes struggle with the correct pronunciation of words.

8. During a conversation, I accidentally skip sentences which make it hard to understand what I mean to say.

Listening

1. I immediately hear it when somebody pronounces a sound incorrectly.
2. I hardly hear any difference between a ‘p’ and a ‘b’ in a word.
3. During a conversation, I sometimes notice that I perceived a sentence in the wrong order.
4. During listening, I hear the words exactly as they are used by the speaker.
5. When somebody reads out aloud something to me, I have a hard time understandings parts that resemble each other.
6. I understand a story exactly as somebody told it.
7. Even when I did not hear something during listening, I understand what somebody means.
8. When I am listening to a story, I accidentally add something for myself which makes the story difficult to understand.

Copying

1. During transcribing, it is rarely necessary to verify whether I write the correct letters.
2. When I am transcribing something, I rarely forget a letter in a word.
3. When I have to transcribe a word repeatedly, I sometimes write it down differently.
4. When I am transcribing something, I exchange letters with similar sounds (for example ‘vrede’ and ‘wrede’).
5. When I transcribing a sentence, I make no mistakes.
6. I transcribe something else than is written, which makes the text difficult to understand.
7. During transcribing, I sometimes write a different word that resembles the word which is written.
8. I can easily transcribe a difficult text.

Dictation

1. During a dictation, I rarely hesitate whether a word should be written with ‘au’ or with ‘ou’.
2. When making a dictation, I almost automatically write down the words without mistakes.
3. Also when somebody uses extended words during a dictation, I immediately under-stand them while I am writing them down.

4. When I have to write down a word more than once during a dictation, I sometimes write them with different spelling.
5. When I want to write down literally what somebody says, I accidentally write down a different sentence because I think I heard it this way.
6. During a dictation, I easily write down long sentences.
7. When I make notes of a lesson, I forget some letters because I could not hear them.
8. A dictation usually goes by so quickly that I exchange parts of the text.

Reading aloud

1. When I have to read out aloud, I exchange words with similar pronunciation.
2. When I read out aloud something for the first time, I correctly pronounce difficult words.
3. When I read out aloud a text, I easily pronounce sentences in the correct order.
4. During reading out aloud, I skip parts so that others do not understand it anymore.
5. When I read aloud words with peculiar sounds, others can perfectly understand me.
6. During reading aloud, I rarely make up sentences which are not written down.
7. During reading aloud, I exchange words that look like each other.
8. When I read out aloud, I accidentally alter a sentence so that a new sentence arises.

Note that:

Litteral translations from Dutch to English are presented. The sentences represent items and examples of situations typical for the Dutch language and the Dutch aspects of dyslexia. Dyslexia research in other languages should incorporate aspects and situations that are typical for their own language.

Appendix B

Items of DF3 (12 SLS, selected with test criterion in Sample 1):

SLS: 2, 4, 12, 14, 17, 26, 35, 36, 42, 43, 44, 45

Items of DF4 (6 GLS + 11 SLS, selected with test criterion in Sample 1):

GLS: 2, 10, 13, 15, 19, 20

SLS: 2, 12, 14, 17, 26, 35, 36, 42, 43, 44, 45

Items of DF5 (4 BQ + 2 GLS + 9 SLS, selected with test criterion in Sample 1):

BQ: 1, 2, 3, 4

GLS: 10, 13

SLS: 10, 11, 14, 17, 35, 36, 38, 42, 45

Items of DF7 (16 SLS, selected with biographical criterion in Sample 1):

SLS: 2, 4, 9, 12, 14, 16, 17, 18, 20, 28, 33, 35, 39, 42, 44, 45

Items of DF8 (3 GLS + 15 SLS, selected with biographical criterion in Sample 1):

GLS: 10, 11, 14

SLS: 2, 4, 11, 14, 16, 17, 18, 20, 26, 35, 39, 41, 42, 44, 45

Items of DF10 (10 SLS, selected with biographical criterion in Sample 2):

SLS: 1, 8, 11, 20, 30, 40, 42, 47, 54, 56

Items of DF11 (5 GLS + 11 SLS, selected with biographical criterion in Sample 2):

GLS: 2, 3, 4, 14, 19

SLS: 1, 11, 18, 20, 30, 35, 40, 42, 52, 54, 56

5

Six factors of adult dyslexia assessed by cognitive tests and self-report questions: Very high predictive validity

The Multiple Diagnostic Digital Dyslexia Test for Adults (MDDDT-A) consists of 12 newly developed tests and self-report questions in the Dutch language. Predictive validity and construct validity were investigated and compared with validity of a standard test battery of dyslexia (STB) in a sample of 154 students.

There are three main results. First, various analyses of principal components showed that six or more factors of dyslexia can be distinguished (rapid naming, spelling, reading, short-term memory, confusion, phonology, attention, complexity). All factors are represented by the MDDDT-A. Second, various discriminant analyses showed good predictive validity for both the tests of the MDDDT-A (90%) and the STB (90%). However, predictive validity of the questionnaire was highest (97%). Third, we analysed the best predictors of dyslexia and found that predictive validity is higher when construct validity is high, that is when a set of predictors represents many characteristics of dyslexia.

The main conclusion is that a digital test battery can be a reliable screening instrument for dyslexia in students, especially when it is accompanied by self-report questions. A theoretical conclusion is that dyslexia is characterized by at least six cognitive impairments in a complex way. In students, this structure may be modulated by high intelligence and good schooling through various compensation strategies. It is therefore recommended to include assessments of all characteristics of dyslexia to achieve the most reliable diagnoses in variable samples and in samples of different countries.

Keywords: Dyslexia, digital assessment, self-report questions, discriminant analysis, principal components analysis

Based on: Tamboer, P., Vorst, H. C. M., & de Jong, P. F. (2017). Six factors of adult dyslexia assessed by cognitive tests and self-report questions: Very high predictive validity. *Research in Developmental Disabilities, 71*, 143-168.

Introduction

In the Netherlands, various methods for diagnosing dyslexia have been developed for children. Reliable methods for diagnosing dyslexia in adults are however sparse and expensive. Although many young adults with dyslexia have been tested during their school days, an unknown but probably substantial portion of them was never tested at all. Of older adults with dyslexia, a majority was never tested because when they were young, diagnosing dyslexia at school was not a widespread practice like it is nowadays. This resulted in the present situation where there are still many adult people who are not aware of having dyslexia. They may experience various difficulties at work while they do not know that dyslexia may be the underlying cause. But even when they come up with the idea of having themselves tested, prohibitive costs may prevent them following through on it.

Worldwide, most traditional diagnostic instruments are based on definitions of dyslexia, such as the one provided by the World Health Organization (2008). Dyslexia is generally defined as a specific reading disorder characterized by a specific and significant impairment in the development of reading skills, which is not due to problems with visual acuity, schooling, or overall mental development. It is also generally assumed that early learning delays cannot be overcome completely despite remedial teaching programs, and that these learning delays interfere with academic achievements into adulthood for most of the people with dyslexia. Prevalence estimates of dyslexia range from about 5% to about 15% of the population across countries.

Specific characterizations of dyslexia are provided by various theories which were developed in more than two decades of scientific research. Many theories have related reading difficulties of people with dyslexia to various underlying deficits (see for reviews: Elliott & Grigorenko, 2014; Peterson & Pennington, 2012). Of these, the most frequently reported are phonological deficits. The phonological deficit theory posits that dyslexia is caused by impairments in phonological information processing, probably caused by problems in the access to or fuzziness of phonological representations of spoken words (e.g. Shaywitz & Shaywitz, 2005; Snowling & Hulme, 2005; Vellutio et al., 2004). In recent years, also visual/attentional deficits have been frequently reported although their relevance as a cause of dyslexia is debated (e.g. Goswami, 2015; Lobier & Valdois, 2015). The most important hypotheses postulate deficits in visual attention span (e.g. Bosse, Tainturier, & Valdois, 2007; Lobier, Zoubrinetzky, & Valdois, 2012; Romani et al., 2011), temporal spatial attention (e.g.

Facoetti et al., 2010), noise exclusion (e.g. Sperling et al., 2006) and visual crowding (e.g. Lorusso et al., 2004; Martelli et al., 2009).

Apart from phonological and visual/attentional deficits, other symptoms of dyslexia are also widely investigated and discussed. These include deficits of short-term memory, rapid naming, speed of processing and many more. None of these deficits is regarded as a necessary or sufficient cause of dyslexia. Instead, most researchers adhere to a multiple deficit view of dyslexia, that dyslexia is caused by multiple cognitive factors which operate probabilistically (Pennington, 2006; van Bergen, van der Leij, & de Jong, 2014).

In the present study, our goal is not to evaluate different theories of dyslexia, but to find the best way to diagnose dyslexia in adults. Especially for diagnosing dyslexia, it is crucial to understand that apart from reading and spelling deficits, many other deficits are related to dyslexia. Although the literature on adult dyslexia is sparse compared to that on dyslexia as a developmental disorder in children, various symptoms of dyslexia in adults were recently investigated. As children with dyslexia, adults with dyslexia experience difficulties with phoneme awareness, rapid automatized naming, reading, spelling, written word recognition and working memory (e.g. Callens et al., 2014; Cavalli et al., 2016; Kemp, Parrila & Kirby, 2009; Nergård-Nilssen & Hulme, 2014; Vellutio et al., 2004). In addition, deficits in various executive functions have been reported (Smith-Park et al., 2016).

In summary, adults with dyslexia seem to experience most of the deficits that are also commonly reported in children with dyslexia. However, an issue to account for is that reading and spelling difficulties of adults with dyslexia may differ depending on the language. Some languages are known for their difficult spelling and related difficulties with phonological transitions (for instance English), while other languages are known for their transparency between their spelling and phonology (for instance Italian). Therefore, the best diagnosis of dyslexia in adults should comprise different measurements for different languages. The present study focuses on the Dutch language. Another issue to account for is that while in children with dyslexia reading and spelling difficulties may be the key symptom of dyslexia, in adults with dyslexia a complicating factor is that some well-educated people may have overcome difficulties such as these. Some deficits may have been compensated for, while other deficits may manifest themselves more severely when compared to children.

Which tests to be used to diagnose dyslexia in adults has hardly been studied. One example in the English language is the Dyslexia Adult Screening Test (DAST) (Nicolson and Fawcett, 1997, 1998), which includes measures of reading, spelling, writing, rapid naming, phonology, working memory, balance, and various verbal skills. In a validation study of the

DAST (Harrison & Nichols, 2005), it was found that the spelling, reading, writing, and phonology tests contributed to the best classification accuracy, which however remained below 90 percent. Another example is a study by Callens et al. (2014). In this study, it was examined how a reliable diagnosis of dyslexia could be made in students in higher education with Dutch as their first language. Predictive validity of 53 subtests was examined covering abilities of reading and spelling, phonological awareness, general intelligence, vocabulary, rapid naming, memory, morphology and syntax, math, and speed of processing. Callens et al. recommended that a test battery for adults in higher education with Dutch as their first language should include measures of word reading, word dictation, proofreading, phonological awareness, rapid naming, and calculation. This is quite an extensive list of tests for a reliable diagnosis, and it requires a lot of time to administer all these test as well as the need for test assistants. As a result, the costs of a diagnosis will be substantial and this is a hurdle for persons who must pay for themselves. For these persons, a cheap and easily available screening battery for dyslexia could be of immense help. Two workable solutions are fully digitised tests with automatic reports and digitised self-assessment through self-reports.

One such battery in the English language is the Bangor Dyslexia Test (BDT) (Reynolds & Caravolas, 2016), which is a valid screening tool for dyslexia with a classification rate of 94%. This test battery measures various abilities such as verbal/phonological short-term working memory, spatial awareness, arithmetic skills, and executive functioning. By our knowledge no fully digitised test batteries are available for adults in the Dutch language, although the ‘IDAA15 +’ (Schraeyen et al., 2009) was validated for young adults (16 years). A disadvantage of this test is that it only can be used by professional psychologists with relatively large costs for the client.

A second cheap solution for a reliable digitised diagnosis of dyslexia is a self-report questionnaire. In several previous studies, support was found for the reliability and validity of self-assessment of dyslexia (Snowling et al., 2012; Tamboer & Vorst, 2015; Willcutt et al., 2010). For example, Tamboer and Vorst (2015) reported classification performances using various criteria of dyslexia of 94% or higher. This study also concerns the further validation of the questionnaire that was used by Tamboer and Vorst.

The prime aim of the present study was to validate the Multiple Diagnostic Digital Test Battery for Adults (MDDDT-A), a fully automated computerised battery of both tests and self-report questions for diagnosing dyslexia in adults. The battery can be ad-ministered without the help of test assistants. The tests and self-report questions of the MDDDT-A were derived from tests and questions of three previous studies (Tamboer & Vorst, 2015; Tamboer, Vorst & Oort, 2014; Tamboer, Vorst & Oort, 2016). We investigated criterion and construct validity of the MDDDT-A in a new sample, by relating the MDDDT-A to a standard battery of tests. The aim was also to replicate previous findings from the three previous studies. The following three questions were asked in this study:

1. What is the relationship of the components of the MDDDT-A with a battery of standard tests (STB) covering the key features of dyslexia? (construct validity)
2. How well can the MDDDT-A battery predict dyslexia in students as compared to a standard test battery of dyslexia? (predictive validity)
3. To what extent can a self-report questionnaire on symptoms of dyslexia contribute to the prediction of dyslexia?

Regarding construct validity, we aimed to create a set of tests and questions covering as many cognitive deficits of dyslexia as possible. In the Netherlands, diagnostic test batteries usually consist of measures of spelling, reading, and some presumed underlying constructs: phonological awareness, rapid naming, and short-term memory. The STB used in this study consisted of all these measures including a measure of visual attention span. In one of three previous studies (Tamboer, Vorst & Oort, 2016), we investigated the factor structure of an extended version of the tests and questions of the MDDDT-A and found five latent variables explaining 60% of the variance. The factors were spelling, phonology, short-term memory, rhyme/confusion and whole-word-processing/complexity. It remained unclear, however, how the last two factors are related to well-known deficits of for example reading, rapid naming and visual attention span. A problematic issue is that deficits such as rapid naming and reading cannot easily be measured directly in a digitalised test battery without an assistant, because computerised speech recognition is not accurate enough. Thus, in the tests of the MDDDT-A battery some features of dyslexia cannot be included, although they may be included in the self-report questionnaire.

Regarding predictive validity, the question is whether it is necessary to measure all deficits that accompany dyslexia. The existence of many theories and symptoms does not necessarily require that all symptoms are measured to obtain a high predictive validity. For instance, Tops et al. (2012) found that the combination of only three tests (word reading, word spelling and phonological awareness) sufficed for the identification of dyslexic and non-dyslexic students in higher education with 91% accuracy. However, consistent with their recommendation to use many tests for a reliable diagnosis, it is yet unclear whether these results can be generalised to a general population sample. In one of our studies (Tamboer, Vorst & Oort, 2014), we found that predictive validity largely depends on which criterion of dyslexia is used. For instance, predictive validity increased in this study when less stringent inclusion and exclusion criteria were used for groups of students with and without dyslexia. Therefore, an optimal criterion will be investigated in the present study as well.

In a third study (Tamboer & Vorst, 2015), we investigated predictive and construct validity of self-report questions. We found that less than 20 self-report items are sufficient to predict dyslexic and non-dyslexic students with an accuracy of at least 94%. Construct validity was supported by a five-factor structure of various dyslexia-related difficulties. We suggested that self-report questions have a higher predictive validity than standard tests of dyslexia, because self-report questions may suffer less from certain types of bias than tests. For instance, in contrast to tests, it can be assumed that there is hardly any influence of general intelligence or schooling on questions because participants compare their own performances with their own social environment.

In the present study, we selected a new sample of Dutch students with and without dyslexia. They completed the MDDDT-A, the STB (administered by a test assistant), and a digitised self-report questionnaire. We assessed construct and predictive validity of the two test batteries and questionnaire in several ways. We investigated construct validity of the test battery and the questionnaire with correlational and principal components analyses. Predictive validity was determined by comparing numerous predictions using sum scores and item scores of the two test batteries and of the self-report questionnaire. Thereby, we used an independent biographical criterion of dyslexia. Finally, we examined the factor structure of the best predictors of our biographical criterion of dyslexia.

Method

Participants

A sample of 154 participants (mean age 20.9 years, 143 females) completed all tests and the questionnaire. All participants were students at the University of Amsterdam, were raised in the Netherlands, and none of them had serious health problems or a history of serious neurological disorders.

Assessment of dyslexia

For an assessment of dyslexia, a biographical criterion was used. This criterion was determined with six biographical indicators of dyslexia, which were assessed in the questionnaire:

1. An official certificate of dyslexia, acquired by an educational psychologist (dyslexia = 2, doubt = 1, no dyslexia = 0),
2. Test results at school or at an institution (dyslexia = 2, doubt = 1, no dyslexia = 0)
3. Extra lessons or remedial training during school days (yes = 2, doubt = 1, no = 0)
4. Information about dyslexic family members (biological mother, father, sisters or brothers) (at least one with dyslexia = 2, doubt = 1, no = 0),
5. A self-report of dyslexia: 'Are you dyslexic?' (yes = 2, doubt = 1, no = 0)
6. A self-report of language difficulties during school years and at present: five questions about reading, spelling, and writing (yes = 2, doubt = 1, no = 0). The score was the mean score over the five questions.

Because an official certificate can be based on tests that resemble the tests in this study, we separated the official certificate from the other biographical indicators. We summarized the scores of the remaining five biographical indicators of dyslexia. Consequently, scores could range from 0 to 10. All students on the first biographical criterion, that is with an official certificate of dyslexia (28), had a total score of 6 or higher on the sum of the other indicators of dyslexia. Only 6 participants without an official certificate had scores of 6 or higher as well. Based on this relatively high consistency, we decided to use only the participants with an official certificate as a criterion group for dyslexia. Participants without an official certificate of dyslexia and a total score of 0 or 1 (72) were categorised into the criterion group of not having dyslexia. With this procedure, 54 participants remained of whom we could not determine with high certainty whether they had dyslexia or not.

An advantage of this procedure is that the selection of the dyslexic and the non-dyslexic group was independent of the tests that we aimed to validate. A disadvantage, however, is that the number of students that did not qualify for either group is large. In most studies, tests are used for a selection of participants with and without dyslexia, although usually with strict inclusion and exclusion criteria. Participants with only moderate symptoms of dyslexia are usually excluded. Thus, also in these studies the results of studies are based on the performance of extreme groups. For analyses of reliability of predictions, this is a severe problem. The use of strict criterion groups might result in misclassifications. People with moderate symptoms of dyslexia may only be identified when their performances on tests are compared with performances of other people with moderate symptoms. Therefore, we also extended the two criterion groups to two larger groups based on consistency between the biographical information and the first prediction analyses. This resulted in groups of 37 students with dyslexia, 98 students without dyslexia and 19 students who could not be identified. For the details, see the results.

Tasks

MDDDT-A (Tests)

The MDDDT-A consists of 12 tests that are related to cognitive aspects of language processing, such as phonological, visual, attentional, auditory, and spelling difficulties.

Cronbach's Alpha was calculated for each test in the whole group of 154 students, except for Letter Order (limited testing time resulted in not all students finishing the test), and Counting Letters (only two items).

In the MDDDT-A, conventional oral answers were replaced by answers typed into the computer. Most of the tests required a choice between two or three answers. In test 1, 2, 3, 6, and 9 (see below), real words or sentences had to be typed into the computer. This might be problematic for two groups of people: elderly with little or no experience in typing and people with dyslexia who might experience some sort of confusion when typing the keys on the keyboard. We took four measures to limit typing bias as much as possible. First, there was no time limit in these tests. Second, typing was kept to a minimum, mostly requiring the typing of single words. Only in test 1 and 9 whole sentences had to be typed. Third, in these typing tests (1 and 9), only a few words were scored, so that typing errors in irrelevant words were ignored. Fourth, we registered how much time it took to complete the 12 subtests, so that very slow and very fast response times could be analysed further for possible sloppy typing. In the present research, none of these issues came up, because students are usually experienced in typing into a computer. For admitting the MDDDT-A in general population samples, we also included a control task measuring typing speed, so that serious issues with correct typing can be identified in people with little experience with typing.

For people with dyslexia, it is often difficult to read task instructions from a computer screen. Therefore, most instructions were read out aloud through headphones at the same time they were displayed on screen. All tests were also preceded by an example item. In the first six tests, all separate items were presented auditorily through the headphone set. Recordings were made of a well-trained female voice reading out aloud all items. For each item, a separate recording was made. These recordings were combined with the visual presentation on the computer screen.

All tests were scored automatically by the computer. Raw scores were registered for all single test items. Total raw scores were automatically computed. These scores were used for consecutive analyses. For future use of the MDDDT-A these analyses are used for automatic reports by the computer.

Dutch Dictation consists of 10 sentences in the Dutch language. Each sentence is presented twice through the headphones, first in normal reading speed, then word by word with small breaks between them. The complete sentences must be typed into the computer. There is no time limit. Each sentence consists two words with each two spelling difficulties, thus four spelling difficulties per sentence. Only the mistakes in these two words are counted. This results in a score of 0 (4 errors) − 4 (no errors) for each sentence and a maximum score of 40 for the whole test (Cronbach's $\alpha = 0.76$).

English Dictation consists of 10 sentences in the English language. Each sentence is presented once through the headphones. At the same time, this sentence can be read on the computer screen except for two omitted words. These words are repeated once through the headphones. The participants only had to enter these words into the computer. There was no time limit. Although Dutch students are familiar with the English language, some English words are well known for their vulnerability to spelling errors for Dutch people. Each English word represented one spelling difficulty. For instance, the word 'noise' can be misspelled as 'noice'. Each word is scored with 0 (incorrect) or 1 (correct). The maximum score is 20 for the whole test (Cronbach's $\alpha = 0.64$).

Missing Letters consists of 10 sentences in the Dutch language. Each sentence is played out once through the headphones. At the same time, this sentence is displayed on the computer screen with the difference that in two difficult words a few letters have been omitted. These words are repeated once through the headphones. The participants only must enter the missing letters of these words into the computer. There is no time limit. Each word is scored with 0 (incorrect) or 1 (correct). The maximum score is 20 for the whole test (Cronbach's $\alpha = 0.52$).

Pseudowords consists of 30 pseudowords, which are nonwords that sound like real words. Each pseudoword is played out once through the headphones. At the same time this pseudoword is displayed on the computer screen. The participants must decide whether the visually presented pseudoword is spelled correctly, which is the case for half of all pseudowords (participants should click on either ‘correct’ or ‘incorrect’, which is displayed on the computer screen). There is a time limit of about six seconds per word. The usual approach for pseudowords is to have the participants read the words aloud themselves. We decided on a different approach however, because it would have been practically impossible to get all students in individual sessions for this test. Each word is scored with 0 (incorrect) or 1 (correct). The maximum score is 30 for the whole test (Cronbach’s α = 0.51).

Sound Deletion consists of 20 difficult Dutch words. Each word is played once through the headphones. Some of these words are pronounced correctly and some words incorrectly by leaving out or adding one sound. On the computer screen, each word is presented three times, each time with a slightly different spelling with one of them being spelled accordingly to what is pronounced. Participants must decide which of the visually presented words they heard through the headphone set. For example, the existing word ‘fietsenstalling’ − which means bicycle shed − is read out as non-existent ‘fiestenstalling’. The possible answers are: ‘fietsentalling’, ‘fiestensalling’ and ‘fiestenstalling’. There is a time limit of about fifteen seconds per word. Each word is scored with 0 (incorrect) or 1 (correct). The maximum score is 20 for the whole test (Cronbach’s α = 0.84).

Spoonerisms consists of 20 words. A Spoonerism is a compound of two existing smaller words, that also allows for a compound of two other existing words when the first letters of the smaller words are interchanged. For example, the word ‘kolen-schop’ becomes ‘scholen-kop’. Each original word is read out once through the headphones and the participants must type the novel word into the computer. There is a time limit of about fifteen seconds per word. Each word is scored with 0 (incorrect) or 1 (correct). The maximum score is 20 for the whole test (Cronbach’s α = 0.87).

Incorrect Spelling consists of 40 Dutch words. All words are displayed on a computer screen for 50ms each. Half of the presented words are spelled correctly and half are spelled incorrectly. Participants must decide whether the words are spelled correctly or not (they must click on ‘correct’ or ‘incorrect’, which is displayed on the computer screen). There is no time limit for the answers. Each word is scored with 0 (incorrect) or 1 (correct). The maximum score is 40 for the whole test (Cronbach’s α = 0.64).

Dutch–English Rhyme Words consists of 40 Dutch-English word pairs. They are displayed on a computer screen for two seconds with the Dutch words on the right. For half of the word pairs, the nouns of the Dutch and the English word resemble each other visually

(English ‘deep’ and Dutch ‘reep’). For the other half of the word pairs, the nouns of the Dutch and the English word resemble each other aurally (English ‘room’ and Dutch ‘bloem’). There are 20 rhyming and 20 non-rhyming pairs. Participants must decide whether the two words of a pair rhyme or not, which is the case in half of all word pairs. There is no time limit. Each response is scored with 0 (incorrect) or 1 (correct). The maximum score is 40 for the whole test (Cronbach’s $\alpha = 0.83$).

Letter Order consists of 20 sentences in the Dutch language. Words consisting of four letters or more are displayed with all letters in random order except the initial and final letters. Each sentence contains two long words in which the letter order is mixed up in this way. The words in the sentences become more difficult towards the end of the test. The participants must enter the complete sentences into the computer with all words spelled correctly. For example, the word “Aoccdrnig” must be typed as “According”. There is a time limit of five minutes for the whole test. Each word is scored with 0 (incorrect) or 1 (correct). The maximum score is 40 for the whole test.

Counting Letters aims to measure the effects of global reading and consists of two sentences. The idea for this test is based on a well-known language puzzle aimed at counting the number of times the letter [f] appears in the following sentence: ‘Finished files are the result of years of scientific study combined with the experience of years’. Many people only see the [f] three times. It has been suggested that the [f] in [of] is overlooked because it sounds more like a [v]. Another suggestion is that [of] is overlooked completely. We created two Dutch sentences based on the same principle. In the sentence ‘Het deftige hondje van de man en de vrouw drinkt water uit de kraan’, participants must count the number of times the letter [d] appears in the sentence (6 times). In the sentence ‘Met de neus en de mond is het niet moeilijk en zelfs gemakkelijk een liedje te neuriën’, participants must count the number of times the letter [n] appears in the sentence (8 times). There is a time limit of about 12 s per sentence. The total score is determined by adding the number of correctly counted letters [d] and [n]. The maximum score is 14.

Short-term Memory Test Forward is a digit span test: the number of digits a person can retain and recall. There are 24 series: 6 with 4, 6 with 5, 6 with 6, and 6 with 7 digits. The digits are presented one by one, for one second each. After displaying the last one of a series of 6, the 6 digits must be typed into the computer in the correct order. There is no time limit for the answers. Each series of digits is scored with 0 (incorrect) or 1 (correct). The maximum score is 24 for the whole test (Cronbach’s $\alpha = 0.80$).

Short-term Memory Test Backward is also a digit span test: the number of digits a person can retain and recall, but this time in reversed order. There are 24 series: 6 with 3, 6 with 4, 6 with 5, and 6 with 6 digits. The digits are presented one by one, for one second each on a computer screen. After displaying the last one of a series of 6 the 6 digits must be typed into the computer in reversed order. The maximum score is 24 for the whole test (Cronbach's $\alpha = 0.83$).

Standard Test Battery (STB)

The STB consists of 10 tests, which must be administered by a test assistant who scores all the answers. All tests are related to cognitive aspects of language processing.

The *One-Minute-Test* (Brus & Voeten, 1979) aims to measure word reading ability. This test is widely used in primary schools to assess progress in reading. The test consists of lists of words of increasing difficulty. Participants must read the words as fast as they can. The score is the number of words read aloud correctly in one minute.

The *Klepel Test* (Van den Bos et al., 1994) aims to measure nonword reading ability. The test is widely used in dyslexia centres in diagnosing dyslexia. The test consists of lists of pseudowords of increasing difficulty. The score on the test is the number of pseudowords read aloud correctly in two minutes.

Visual Attention Span (van den Boer, van Bergen & de Jong, 2015) aims to measure visual attention span, the number of orthographic units (e.g., letters, letter clusters or syllables) that can be processed simultaneously at a glance (Bosse, Tainturier & Valdois, 2007). First, a central fixation point is shown for 1000ms. Second, a white screen is presented for 500ms. Third, a group of letters is presented for 200ms. There are 25 items: 5 items consisting of 6 letters, 5 items consisting of 7 letters and 5 items consisting of 8 letters and 10 items consisting of 10 letters. Participants must read aloud all letters. The score is determined by summarizing all letters that were read aloud correctly (maximum = 205).

The *Amsterdam Sound Deletion Test* (van Bergen, Bishop, van Zuijen & de Jong, 2015) aims to measure phonological awareness. Nonwords are presented auditory. The words are played from a tape recording by a professional speaker. On all items the participants must first repeat the present word and then repeat the word while leaving out one specific sound. The test consists of 12 items. In the first 8 items (2 groups of 4 items), one sound must be left out (for instance: SKOOM without K = SOOM). In the next 4 items, one sound must be left out twice in the same word (for instance: PORSVUST without S = PORVUT). For each item, accuracy and response time are registered. The score is determined as follows: For the each of the three groups of four items the median response time is determined. Next, the mean of these three values is determined. Finally, a fluency score is computed as the number of correct answers per minute.

Digit Span Forward is a subtest of the WAIS (Wechsler, 2008). An observer reads aloud a series of digits and the participant must repeat them in the correct order. The test starts with two series of two digits, then two series of three digits, and so on. If two series with the same number of digits are repeated incorrectly, the test is stopped. The score is then determined by the number of correctly repeated series.

Digit Span Backward is a subtest of the WAIS (Wechsler, 2008). An observer reads aloud a series of digits and the participant must repeat them in reversed order. The test starts with two series of two digits, then two series of three digits, and so on. If two series with the same number of digits are repeated incorrectly, the test is stopped. The score is then determined by the number of correctly repeated series.

Rapid Automatized Naming (RAN) is a subtest of the GL & SCHR (De Pessemier & Andries, 2009) and aims to measure speed of automatized naming. Several items are presented on a card to the participant, who must read aloud the items as fast and as precisely as possible. There are four subtests in which the items are either five different digits, letters, colours or pictures. Thus, one card consists of 50 digits and there are five different digits. Per card, the number of errors the time needed to name all items are registered. For each card the score was transformed to the number of correctly pronounced items per second.

MDDDT-A (Self-report questions)

In one of our previous studies (Tamboer & Vorst, 2015), we investigated 60 general language statements and 140 specific language statements. We found that these statements differentiated with 97% accuracy between persons with and without dyslexia (using a biographical criterion). The general language statements assessed general abilities of reading, speaking, writing, making mental representations, memory and knowledge of foreign languages. The 140 specific statements were designed by a cross classification of three dimensions (7 × 5 × 4). One dimension distinguished between seven distinct aspects of how language is used in daily life or at school and universities: reading, writing, speaking, listening, copying, taking a dictation and reading aloud. A second dimension distinguished between five distinct levels of how language can be represented: by sounds, letters, words, sentences or by text. A third dimension distinguished between four different difficulties that accompany dyslexic adults: skipping (forgetting), adding, changing and exchanging. For instance, dyslexics may skip parts of sentences when reading, exchange letters when writing, change words when speaking, or forget parts of a text when making a dictation. Below one example is given for each subscale of the first dimension. Reading: ‘Sometimes I skip a letter, which results in reading a different word’; Writing: ‘Sometimes I forget to write down a syllable’; Speaking: ‘While speaking, I sometimes exchange similar words’; Listening: ‘I hear a story exactly like someone tells it’; Copying: ‘When I copy out a text, I sometimes exchange letters with similar sounds’; Dictating: ‘I make mistakes in dictation, because I don’t hear the correct sounds’; Reading aloud: ‘When reading aloud, I sometimes repeat a part of the text’.

In the present study, we further analysed two reduced sets of 20 general language statements and 56 specific language statements, which were selected based on our previous study. We selected the best statements based on group differences and predictive power, keeping the design of the specific statements intact as much as possible. All statements were scored on a 7-point Likert scale.

Procedure

All participants were recruited by E-mail or through advertising at various departments of the University of Amsterdam. They were informed about the general nature of the tests and the questionnaires in advance in accordance with a standard protocol. All tests were assessed in one session in a special testing room where noise from outside the room could not be heard. Only one participant and one assistant were present in the room, and during the digital testing the participant was alone. The order of the STB, the MDDDT-A and the questionnaire was counterbalanced. The total testing time of all tests and questions was between two and three hours. Afterwards, the students received a more detailed debriefing. Anonymity was guaranteed by the standard protocol of the University of Amsterdam.

Design and analyses

Construct validity was investigated by analysing convergent validity between the MDDDT-A with the STB, a diagnostic battery that consists of tests that are commonly used for diagnosing purposes in the Netherlands. When convergent validity between the MDDDT-A and the STB is high, this means that the same underlying constructs are measured. This means that the MDDDT-A is indeed measuring difficulties in dyslexia and not something else. Convergent validity will be investigated with correlational analyses and with a comparison between factor structures of the two test batteries. Both batteries were factor analysed with principal components analysis, and correlations between the resulting components are presented. First, convergent validity of only the tests of the MDDDT-A was investigated. Second, convergent validity of the self-report questions was investigated.

Criterion validity was investigated with discriminant analyses on biographical criterion groups. Two sets of analyses were done. First with biographical criterion groups with strict inclusion and exclusion criteria and then in adjusted criterion groups with less stringent inclusion and exclusion criteria. In each set of analyses, we used six sets of predictors: 1) all test scores of the MDDDT-A, 2) all test scores of the STB, 3) all test scores together, 4) single item scores of the questionnaire, 5) single item scores of the tests of the MDDDT-A, and 6) these item scores together.

Based on predictions with the best classification accuracy, a final classification of students with and without dyslexia was determined. These groups were used to investigate validity of factors of dyslexia. Three analyses were performed: group differences, discriminant analysis with eight factor scores as predictors, and analyses of factor structures in various discriminant functions.

Results

Construct validity

Correlations (Table 1)

To get a first impression of construct validity of the MDDDT-A, we performed correlational analyses with all tests from the MDDDT-A and the STB. High correlations between two tests or test batteries of which one is validated and another is not, indicate high convergent validity, and thus high construct validity. Relatively high and significant correlations were found between the spelling tests of the MDDDT-A and the reading tests *One-Minute-Test* and *Klepel Test*. High and significant correlations were also found between all short-term memory tests. Apart from these correlations, which were expected, many more moderate and significant correlations were found.

Table 1

Correlations between 22 tests (Pearson) (N = 154)

MDDDT-A:		**STB:**	
DD	= Dutch Dictation	OMT	= One-Minute-Test (OMT)
ED	= English Dictation	KT	= Klepel Test (KT)
ML	= Missing Letters	VAS	= Visual Attention Span (VAS)
PW	= Pseudowords	ASDT	= Amsterdam Sound Deletion Test (ASDT)
SD	= Sound Deletion	DSF	= Digit Span Forward (DSF)
SP	= Spoonerisms	DSB	= Digit Span Backward (DSB)
IS	= Incorrect Spelling	RANL	= RAN Letters (RANL)
DER	= Dutch-English Rhyme Words	RANN	= RAN Numbers (RANN)
LO	= Letter Order	RANC	= RAN Colours (RANC)
CL	= Counting Letters	RANP	= RAN Pictures (RANP)
STMF	= Short-term Memory Test Forward		
STMB	= Short-term Memory Test Backward		

	DD	ED	ML	PW	SD	SP	IS	DER	LO	CL	STMF	STMB	OMT	KT	VAS	ASDT	DSF	DSB	RANL	RANN	RANC	RANP
MDDDT-A:																						
DD	1																					
ED	.49**	1																				
ML	.40**	.33**	1																			
PW	.36**	.30**	.41**	1																		
SD	.13	.04	.09	.27**	1																	
SP	.41**	.28**	.29**	.36**	.23**	1																
IS	**.51****	.40**	.35**	.38**	.24**	.43**	1															
DER	.27**	.16*	.09	.34**	.35**	.23**	.24**	1														
LO	.28**	.19*	.14	.33**	.17*	.16	.24**	.37**	1													
CL	.19*	.17*	.21**	.20*	.18*	.02	.33**	.19*	.24**	1												
STMB	.36**	.24**	.36**	.42**	.20*	.24**	.27**	.22**	.17*	.07	1											
STMF	.27**	.15	.36**	.41**	.25**	.31**	.33**	.23**	.36**	.21*	**.67****	1										
STB:																						
OMT	.39**	.36**	.31**	.28**	.08	.38**	.38**	.18*	.10	.18*	.21*	.23**	1									
KT	**.53****	.43**	.36**	.38**	.09	.39**	.43**	.26**	.07	.14	.38**	.32**	**.69****	1								
VAS	.35**	.23**	.24**	.18*	.20*	.27**	.38**	.16*	.18*	.21**	.28**	.23**	.30**	.40**	1							
ASDT	.43**	.27**	.27**	.27**	.18*	.38**	.32**	.24**	.04	.13	.30**	.27**	.38**	**.51****	.33**	1						
DSF	.16	.10	.10	.23**	.17*	.13	.17*	.15	.17*	.06	.42**	.38**	.20*	.26**	.16	.12	1					
DSB	.24**	.07	.38**	.28**	.10	.10	.17*	.10	.14	.10	.45**	.46**	.17*	.35**	.33**	.17*	.34**	1				
RANL	.26	.22**	.30**	.23**	.19*	.27**	.23**	.25**	.09	.06	.28**	.31**	**.72****	**.57****	.30**	.33**	.09	.20*	1			
RANN	.28	.23**	.36**	.29**	.15	.35**	.27**	.28**	.08	.05	.31**	.36**	**.72****	**.67****	.30**	.38**	.17*	.27**	**.89****	1		
RANC	.29	.16	.30**	.27**	.13	.23**	.23**	.33**	.20*	.18*	.21**	.30**	.42**	.40**	.17*	.31**	.08	.28**	**.58****	**.54****	1	
RANP	.26	.08	.26**	.23**	.09	.24**	.20*	.28**	.20*	.10	.27**	.30**	.43**	.42**	.19*	.23**	.21**	.34**	**.53****	**.57****	**.65****	1

* Correlation is significant at the 0.05 level (two-talied)
** Correlation is significant at the 0.01 level (two-talied)
Bold Correlations > 0.5

Table 2 shows the rotated component matrix (varimax rotation) of the twelve tests of the MDDDT-A. Four components with eigenvalues larger than 1 explained 63% of all variance. Four tests showed high factor loadings on the first component: *Dutch Dictation*, *English Dictation*, *Spoonerisms*, and *Incorrect Spelling*. Except for *Spoonerisms*, these tests largely depended on spelling abilities. Thus, we interpret this component as a spelling factor. The second component showed high factor loadings for the memory tests, thus pointing to a memory factor. The third component showed high loadings for *Sound Deletion* and *Dutch-English Rhyme Words*. Seeing that there were some smaller loadings of other tests on this component, a phonological factor seems logical. The interpretation of the fourth component is difficult. This factor could represent exchanging, confusion or attention. Only *Counting Letters* showed a high loading. Thus, most likely this component represents some form of attention. We repeated the analysis with oblique rotation, which allows factors to correlate. The patterns of loadings were the same and the correlations between factors low. In summary, we distinguished the following factors:

1. Spelling
2. Short-term memory
3. Phonology
4. Attention (exchanging, confusion)

Construct validity: STB (Table 3)

Table 3 shows the rotated component matrix of the ten tests of the STB. Three components with eigenvalues larger than 1 explained 69% of all variance. The first component can be interpreted as a factor rapid naming. It comes as no surprise that the *One-Minute-Test* and *Klepel Test* also load on this factor, because in these tests the words had to be read aloud. The third component can also be interpreted easily as a factor memory because only the memory tests show high factor loadings. The second component is difficult to interpret. High loadings are found for the *Klepel Test*, *Visual Attention Span*, and the *Amsterdam Sound Deletion Test*. We decided to interpret this as a productive phonology factor. An analysis with oblique rotation resulted in similar patterns of loadings and low correlations between factors. In summary, we distinguished the following factors:

1. Rapid naming
2. Productive phonology
3. Short-term memory

Table 2

Rotated Component Matrix: 4 factors explaining 63% of all variance of 12 tests of MDDDT-A (N = 154)

	Factor 1 (Spelling)	Factor 2 (Short-term memory)	Factor 3 (Phonology)	Factor 4 (Attention)
Explained variance:	21.0%	17.5%	14.6%	10.2%
Dutch Dictation	**0.77**	0.19	0.14	0.08
English Dictation	**0.77**	0.05	-0.04	0.14
Missing Letters	**0.52**	**0.50**	-0.15	0.16
Pseudowords	**0.39**	**0.47**	**0.36**	0.10
Sound Deletion	0.01	0.14	**0.74**	0.01
Spoonerisms	**0.60**	0.18	**0.37**	**-0.36**
Incorrect Spelling	**0.68**	0.16	0.24	0.21
Dutch-English rhyme words	0.16	0.06	**0.76**	0.15
Letter Order	0.14	0.20	**0.45**	**0.47**
Counting Letters	0.17	0.06	0.10	**0.85**
Short-term Memory Test Forward	0.17	**0.86**	0.11	-0.05
Short-term Memory Test Backward	0.09	**0.85**	0.23	0.14

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Bold: Factor loadings > 0.3

Table 3

Rotated Component Matrix: 3 factors explaining 69% of all variance of 10 tests of STB (N = 154)

	Factor 1 (Rapid naming)	Factor 2 (Productive phonology)	Factor 3 (Short-term memory)
Explained variance:	31.8%	22.7%	14.7%
One-Minute-Test	**0.60**	**0.59**	-0.02
Klepel Test	**0.46**	**0.70**	0.20
Visual Attention Span	-0.01	**0.69**	**0.32**
Amsterdam Sound Deletion Test	0.16	**0.72**	0.04
Digit Span Forward	0.04	0.13	**0.75**
Digit Span Backward	0.21	0.14	**0.79**
RAN Letters	**0.80**	**0.43**	-0.06
RAN Numbers	**0.78**	**0.48**	0.04
RAN Colours	**0.82**	0.05	0.15
RAN Pictures	**0.80**	0.00	**0.32**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Bold: Factor loadings > 0.3

Correlations between test factors of MDDDT-A (tests) and STB (Table 4)

Table 4 shows the correlations between the factor scores of the two analyses. One quite high and significant correlation was found between the two short-term memory factors (*r* = 0.49). Another quite high and significant correlation (*r* = 0.52) was found between the spelling factor of MDDDT-A and the productive phonology factor of STB. No correlations were found between phonology and attention of the MDDDT-A and the factors of the VTB. Thus, the construct validity of the MDDDT-A is only partly supported by the STB.

Table 4

Correlations between 7 factor scores of MDDDT-A and STB (Pearson) (N = 154)

	MDDDT-A Spelling	MDDDT-A Short-term memory	MDDDT-A Phonology	MDDDT-A Attention	STB Rapid naming	STB Productive phonology	STB Short-term memory
MDDDT-A							
Spelling	1						
Short-term memory	0	1					
Phonology	0	0	1				
Attention	0	0	0	1			
STB							
Rapid naming	.18*	.23**	.15	.02	1		
Productive phonology	**.52****	.14	.11	-.03	0	1	
Short-term memory	.01	**.49****	.09	.10	0	0	1

* Correlation is significant at the 0.05 level (two-talied).
** Correlation is significant at the 0.01 level (two-talied).
Bold Correlations > 0.4

Construct validity: MDDDT-A and STB (Table 5)

Another way to investigate construct validity of the MDDDT-A is to find latent variables of the full set of tests of both the MDDDT-A and the STB, and then to investigate relations between the separate tests and the latent variables. We performed principal components analyses on the 22 tests. We found 6 components with eigenvalues larger than 1 which explained 65% of variance, see Table 5 for the rotated component matrix. Alternatively, we repeated this analysis with oblique rotation instead of varimax rotation which allows factors to correlate. However, the patterns of loadings were the same and the correlations between factors low.

Comparable to the previous analysis of the STB, here the first component also showed high loadings for the subtests of *Rapid Naming* and the *One-Minute-Test*, and it can therefore be interpreted as a rapid naming factor. The second component is determined by the spelling factor in the MDDDT-A analysis and the productive phonology factor of the STB analysis. We interpreted this as a productive phonology factor. The loadings of the STB tests are lower than in the previous analysis. The reason for this remains somewhat unclear. The third component can clearly be considered a short-term memory factor with high loadings from the memory tests of both batteries. The fourth component seems impossible to interpret. The loadings of *Letter Order* and *Dutch-English Rhyme Words* may point to some sort of confusion. This is our preliminary interpretation. The fifth component has only one high factor loading from Sound Deletion. Because the tests *Sound Deletion* and *Dutch-English Rhyme Words* required responses instead of phono-logical production, we interpreted this factor as a receptive phonology factor. The sixth component is the attention factor of the MDDDT-A analysis with a high loading for *Counting Letters*. A moderately high loading of *Visual Attention Span* seems to confirm this. In summary, we distinguished the following factors:

1. Rapid naming
2. Productive phonology
3. Short-term memory
4. Confusion (?)
5. Receptive phonology
6. Attention

Table 5

Rotated Component Matrix: 6 factors explaining 65% of all variance of 22 tests of MDDDT-A and STB

(N = 154)

	Factor 1 (Rapid naming)	Factor 2 (productive phonology)	Factor 3 (Short-term memory)	Factor 4 (Confusion?)	Factor 5 (Receptive phonology)	Factor 6 (Attention)
Explained variance:	17.6%	15.4%	12.3%	7.4%	6.3%	6.1%
MDDDT-A:						
Dutch Dictation	.16	**.73**	.16	.17	-.01	.15
English Dictation	.07	**.75**	-.01	.15	-.16	.07
Missing Letters	.23	**.45**	**.35**	.14	-.29	.21
Pseudowords	.12	**.45**	**.36**	**.43**	.12	-.02
Sound Deletion	.04	.04	.17	.17	**.78**	.17
Spoonerisms	.19	**.61**	.12	.07	**.33**	-.18
Incorrect Spelling	.10	**.63**	.12	.14	.18	**.35**
Dutch-English Rhyme Words	.25	.14	.04	**.55**	**.49**	.01
Letter Order	.01	.14	.15	**.73**	.08	.18
Counting Letters	.04	.10	-.01	.29	.04	**.79**
Short-term Memory Test Forward	.13	.27	**.77**	.11	.07	-.08
Short-term Memory Test Backward	.19	.16	**.73**	.27	.12	.07
STB:						
One-Minute-Test	**.70**	**.45**	.03	-.11	.05	.10
Klepel Test	**.57**	**.57**	.24	-.14	.05	.09
Visual Attention Span	.19	**.32**	.23	-.23	.26	**.56**
Amsterdam Sound Deletion Test	**.30**	**.51**	.15	-.16	.29	.08
Digit Span Forward	.04	.05	**.66**	-.01	.20	-.03
Digit Span Backward	.22	.01	**.73**	-.02	-.15	.27
RAN Letters	**.88**	.18	.09	-.03	.13	.02
RAN Numbers	**.87**	.24	.17	-.05	.11	-.02
RAN Colours	**.74**	.06	.09	**.34**	-.03	.11
RAN Pictures	**.74**	-.01	.23	.26	-.04	.05

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Bold: Factor loadings > 0.3

Comparing these results with the previous analyses, it becomes clear that the six components are not merely the summation of the components found in the separate test batteries. The rotated component matrix does show that latent variables reveal relations between the two test batteries. Although some interpretations of components are preliminary, it is apparent that both batteries represent most components or latent factors. The MDDDT-A does not support the rapid naming factor, which makes sense, because rapid naming could not be assessed digitally. The confusion factor is not well represented by the tests of the STB.

Construct validity: questionnaire (No table)

With the aim of finding confirmation and possibly better interpretations of the six latent variables which were found in the previous paragraph, we investigated the existence of latent variables in the questionnaire. We performed a principal components analysis on the 76 self-report statements. We found eight components with eigenvalues larger than 1.6 which explained 62% of variance (varimax rotation). Many components with eigenvalues lower than 1.3 (and higher than 1) were found as well, but they hardly added interpretable variance. We did not include the table here because it would have taken up too much space.

1. Spelling/Grammatical rules
2. Confusion/Exchanging
3. Rapid naming
4. Complexity in language
5. Learning English
6. Reading
7. Phonology
8. Attention/Short-term memory

Comparing these factors with six of the tests, we conclude that there are two extra factors and that interpretation is not very straight-forward for all components. There is no clear memory component. The component complexity, which was found in our previous study (Tamboer, Vorst & Oort, 2016), was not found in the tests of the present study. The component learning English is new. The fact that more components were found in the present study is the result of two differences between the set of questions of the two studies. In our previous study, we analysed 140 specific language statements whereas in the present study, we analysed a reduced set of 56 specific language statements. In the present study, we also factor analysed general language statements.

Construct validity: MDDDT-A (tests), STB and questionnaire (Table 6)

A principal components analysis on a combination of tests and questions is complicated by method variance. In our previous study (Tamboer, Vorst & Oort, 2016), we avoided method variance by transforming sum scores to factor scores. Here, we tried to replicate this.

For the questionnaire, we categorised all self-report statements into the eight components, depending on their highest factor loading. Then we summarized and standardized them to eight Z-scores. We entered these scores in a factor analysis and requested eight components, thus transforming the sum scores to standardized factor scores and preserving all variance.

The set of 22 tests was reduced to 13 standardized sum scores. Tests that showed factor loadings higher than 0.6 on one component but lower than 0.4 on all other components were combined. The *One-Minute-Test* and the *Klepel Test* were combined because they had a high and significant correlation (0.7). This resulted in five sum scores (rapid naming: four subtests; short-term memory: four subtests; spelling: *Dutch Dictation* and *English Dictation*; reading: *One-Minute-Test* and *Klepel Test*; attention: *Counting Letters* and *Visual Attention Span*). Together with eight single test scores, these five sum scores were standardized to 13 Z-scores. We entered these scores in a factor analysis and requested 13 components, thus transforming the sum scores to standardized factor scores and pre-serving all variance.

The structure within the set of all 21 (i.e. 13 + 8) transformed factor scores was analysed in a principal components analysis. This resulted in eight components with eigenvalues larger than 1, explaining 52% of variance. Table 6 shows the rotated component matrix. As an alternative, we repeated this analysis with oblique rotation instead of varimax rotation which allows factors to correlate. However, the patterns of loadings were the same and the correlations between factors low. We named the components based on their highest loading. In summary, we distinguished the following factors:

1. Spelling
2. Rapid naming
3. Reading
4. Attention
5. Short-term memory/Confusion
6. English (?)
7. Phonology
8. Complexity (?)

Table 6
Rotated Component Matrix: 8 factors explaining 52% of 13 factor scores of tests and 8 factor scores of a questionnaire
(N = 154)

	Factor 1 (Spelling)	Factor 2 (Rapid naming)	Factor 3 (Reading)	Factor 4 (Attention)	Factor 5 (Confusion/ Short-term memory)	Factor 6 (English?)	Factor 7 (Phonology)	Factor 8 (Complexity?)
Explained variance:	7.4%	7.0%	6.7%	6.4%	6.4%	6.2%	6.0%	5.9%
Questionnaire:								
Spelling / Grammatical rules	**.88**	-.01	.01	.04	.07	.14	.00	-.01
Rapid Naming	-.01	**.80**	-.09	-.23	.07	.03	.09	-.04
Reading	.10	.15	**.80**	.09	-.11	-.12	-.07	.02
Attention / Short-term memory	.02	.23	-.10	**.75**	-.13	.12	.00	.05
Phonology	.00	-.05	.07	.02	.03	.03	**.78**	.03
Confusion / Exchanging	.02	.05	.09	.19	**.79**	.01	-.04	.12
Complexity in language	.01	.06	-.02	-.11	-.03	.04	-.02	**.78**
Learning English	-.04	-.06	.17	-.09	.07	**.77**	-.03	-.03
MDDDT-A:								
Spelling (Dutch+English Dictation)	**.74**	-.05	.16	-.03	.01	-.06	-.01	.05
Sound Deletion	.10	.06	-.15	.24	.02	-.05	**.43**	**-.40**
Spoonerisms	-.04	-.24	-.02	.09	**.34**	**.37**	**-.40**	-.08
Incorrect Spelling	**.32**	.08	-.20	.16	-.18	**.58**	.06	.04
Missing Letters	.22	.27	.18	.01	.22	.02	-.27	**-.32**
Pseudowords	-.07	.17	.02	.05	.19	.27	.23	.27
Letter Order	.13	-.28	-.03	**.38**	.04	-.18	-.11	.05
Dutch-English Rhyme Words	-.08	-.01	.14	-.15	.24	.09	.27	.01
STB:								
Rapid Naming (4 subtests)	-.04	**.67**	.14	.28	.04	-.09	-.13	.12
Reading (OMT+Klepel)	.10	-.08	**.65**	-.06	.07	.19	.16	-.01
Amsterdam Sound Deletion Test	.12	-.01	.01	.16	.21	-.09	.02	**.50**
MDDDT-A/STB:								
Short-term Memory (4 subtests)	.13	.10	**-.32**	-.16	**.56**	-.03	.13	.04
Attention (Counting Letters+VAS)	-.06	-.14	.15	**.53**	.16	.01	.09	-.10

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Bold: Factor loadings > 0.3

The most important finding is that many factors of the tests and questionnaire with the same interpretation load on the same component in this analysis. The first four components (spelling, rapid naming, reading and attention) show one high factor loading of a test and one high factor loading of the questionnaire. The same is true for phonology, only with moderate loadings on the tests. The fifth factor, short-term memory/confusion is a combination of two factors in our previous study, in which we concluded that these are separate factors. It is also remarkable that the sum score of the *One-Minute-Test* and the *Klepel Test* (reading) is only loading on a separate factor reading, while these tests loaded on rapid naming and attention in the factor-structure of the tests. The sixth and eighth factor are difficult to interpret. In summary, the relations between subtests and the latent variables of this analysis are not very clear. Therefore, conclusions can only be given with the greatest caution. Nevertheless, the MDDDT-A does seem to represent many aspects of dyslexia, although reading and rapid naming was only represented by the self-report questionnaire and not by the tests of the MDDDT-A.

Criterion validity

Details of analyses

All classifications were derived from discriminant analysis carried out in SPSS. Alternatively, logistic regression analyses could have been used. Pohar, Blas and Turk (2004) evaluated both analyses in various situations, and we found that their recommendations provided a reliable guideline for evaluating various predictions in our previous study (Tamboer, Vorst & Oort, 2014). As the above-mentioned researchers indicated, discriminant analysis should be preferred when violations of normality of predictor variables are not too bad (skewness within the interval [−0.2, 0.2]) and when the independent variables of predictors have four or more answer categories. Based on these recommendations we decided to use discriminant analysis. This analysis assumes normal score distributions of predictors which was the case for all variables used in our analysis.

We used the stepwise method of discriminant analysis to predict group membership (dyslexic or non-dyslexic), because we wanted to acquire a reduced set of predictors while we did not want to assign some predictors higher priority than others. We set prior probabilities to *all groups equal*. A cross-validation procedure was chosen by selecting the option *leave-*

one-out classification, to prevent that a sole case would be classified based partly on itself, and to minimize the effects of outliers.

We used only a selection of items from our previous study (Tamboer, Vorst & Oort, 2014) to prevent overfitting. It is generally recommended that the number of predictors used in a prediction analysis does not exceed the number of participants in the smallest group because this could lead to overfitting, meaning that the results become sample-specific and do not generalise to other samples (Tabachnik & Fidell, 2007).

We present total classification accuracy, sensitivity (correctly identified students with dyslexia) and specificity (correctly identified students without dyslexia). It is generally recommended that the latter should be above 90%. Also for the MDDDT-A, high specificity is more important than high sensitivity. For instance, a sensitivity of 90% would imply that 10% of the students with dyslexia are not identified. For them, there is always the possibility to take further steps, although these might be expensive. On the other hand, a specificity of 90% would imply that 10% of students without dyslexia incorrectly obtain a diagnosis of dyslexia. These students may start to worry whereas in fact, they have nothing to worry about. This situation should be avoided as much as possible.

Discriminant analyses on strict criterion groups (Table 7)

Two biographical criterion groups were determined with the questionnaire: 28 students with dyslexia and 72 students without dyslexia as a criterion. The remaining 54 students could not be identified as either having dyslexia or not having dyslexia with reasonable reliability. The two criterion groups were entered in a discriminant analysis. Next, all 154 students were classified. Thus, also the students that were not included in the criterion groups were classified. These were not included in the tables.

We analysed six different sets of potential predictors: 12 sum scores of the tests of the MDDDT-A, 10 sum scores of the tests of the STB, 22 sum scores of both test batteries together, a selection of 17 items of the questionnaire, a selection of 22 test items of the MDDDT-A, and these 39 items together. Six discriminant functions were calculated that maximally distinguished participants with and without dyslexia.

Table 7 shows six classification tables. Also presented are: overall classification accuracy of the two criterion groups combined (100 participants), sensitivity (percentage of 28 participants with dyslexia that is correctly classified), and specificity (percentage of 72 participants without dyslexia that is correctly classified).

Table 7

Discriminant analyses (cross-validated) on two biographical criterion groups:
28 participants with dyslexia (D)
72 participants without dyslexia (ND)

Discriminant analysis 1: Sum scores of MDDDT-A
Potential predictors: 12 sum scores of 12 tests
Selected predictors in discriminant function:
1. Dutch Dictation
2. Missing Letters
3. Sound Deletion
4. Incorrect Spelling
5. Counting Letters

Classification accuracy (100 participants): 88%
Sensitivity (28 participants with dyslexia): 82%
Specificity (72 participants without dyslexia): 90%

Classification of 28 dyslexics and 72 non-dyslexics:

	Predicted as:	
	D	ND
28 D	23	5
72 ND	7	65

Discriminant analysis 2: Sum scores of STB
Potential predictors: 10 sum scores of 10 tests
Selected predictors in discriminant function:
1. Klepel Test
2. Amsterdam Sound Deletion Test

Classification accuracy (100 participants): 84%
Sensitivity (28 participants with dyslexia): 82%
Specificity (72 participants without dyslexia): 85%

Classification of 28 dyslexics and 72 non-dyslexics:

	Predicted as:	
	D	ND
28 D	23	5
72 ND	11	61

Discriminant analysis 3: Sum scores of all tests of both batteries
Potential predictors: 22 sum scores of 22 tests
Selected predictors in discriminant function:
1. Dutch Dictation
2. Missing Letters
3. Klepel Test
4. Amsterdam Sound Deletion Test

Classification accuracy (100 participants): 90%
Sensitivity (28 participants with dyslexia): 89%
Specificity (72 participants without dyslexia): 90%

Classification of 28 dyslexics and 72 non-dyslexics:

	Predicted as:	
	D	ND
28 D	25	3
72 ND	7	65

Discriminant analysis 4: Items of questionnaire
Potential predictors: 17 statements
Selected predictors in discriminant function:
1. When I'm writing, I often swap letters.
2. Reading difficult Dutch words does not cause a lot of problems for me.
3. When making a dictation I almost automatically write down the words without mistakes.

Classification accuracy (100 participants): 90%
Sensitivity (28 participants with dyslexia): 86%
Specificity (72 participants without dyslexia): 92%

Classification of 28 dyslexics and 72 non-dyslexics:

	Predicted as:	
	D	ND
28 D	24	4
72 ND	6	66

Discriminant analysis 5: Items of MDDDT-A
Potential predictors: 22 items
Selected predictors in discriminant function:
1. Dutch Dictation: De hachelijke onderneming leidde tot twijfel.
2. English Dictation: That woman is my teacher, she is very strict.
3. Short-term Memory Forward: 382986

Classification accuracy (100 participants): 78%
Sensitivity (28 participants with dyslexia): 61%
Specificity (72 participants without dyslexia): 85%

Classification of 28 dyslexics and 72 non-dyslexics:

	Predicted as:	
	D	ND
28 D	17	11
72 ND	11	61

Discriminant analysis 6: Items of MDDDT-A and questionnaire
Potential predictors: 39 items
Selected predictors in discriminant function:
1. English Dictation: That woman is my teacher, she is very strict.
2. Statement: When I'm writing, I often swap letters.
3. Statement: When making a dictation I almost automatically write down the words without mistakes.

Classification accuracy (100 participants): 92%
Sensitivity (28 participants with dyslexia): 89%
Specificity (72 participants without dyslexia): 93%

Classification of 28 dyslexics and 72 non-dyslexics:

	Predicted as:	
	D	ND
28 D	25	3
72 ND	5	67

In Discriminant Analysis 1, five tests of the MDDDT-A classified 88% of 100 participants correctly. In Discriminant Analysis 2, two tests of the validated battery classified 84% of 100 participants correctly. In Discriminant Analysis 3, four tests of all tests together classified 90% of 100 participants correctly. In Discriminant Analysis 4, three statements of the self-report questionnaire classified 90% of 100 participants correctly. In Discriminant Analysis 5, three items of all tests of the MDDDT-A classified 78% of 100 participants correctly. In Discriminant Analysis 6, one item of the MDDDT-A and two statements of the self-report questionnaire classified 92% of 100 participants correctly. Comparing the classification results, we can draw several conclusions:

1. A comparison of the sets of sum scores of the two test batteries shows that highest classification accuracy is reached when a combination of spelling/reading and phonological abilities is selected in the discriminant function.
2. Highest classification accuracy based on sum scores is reached when all 22 tests are used as potential predictors.
3. When we compare the predictions with single items, we see that self-report statements are better predictors than test items. It is remarkable that only three self-report statements suffice for a good classification.
4. All items together resulted in highest classification accuracy, but overfitting may be at play here.

Adjusting the criterion groups (No table)

Based on the results of the six discriminant analyses, we reordered the two criterion groups and included the remaining group of 54 students that could not be assigned earlier to the dyslexic and non-dyslexic group. Participants could be assigned to a group if the following demands were met. Participants are considered to have dyslexia when they are categorised as such in at least four out of six classifications. They are considered to have no dyslexia when they are categorised as such in at least five out of six classifications. The remaining participants are not categorised as having dyslexia or not.

Of the 28 participants with dyslexia according to the biographical criterion, 26 were categorised as having dyslexia and 2 could not be categorised. Of the 72 participants with no dyslexia on the biographical criterion, 64 were categorised as having no dyslexia and 8 could not be categorised. Within the group of 54 remaining participants, 11 were now categorised as having dyslexia and 34 as having no dyslexia, while 9 participants still could not be categorised. As such, this procedure resulted in 37 participants with dyslexia, 98 participants without dyslexia, and 19 with no clear classification. This resulted in larger groups of students with and without dyslexia then before, while consistency with the criterion groups based on only biographical information remained high.

Discriminant analyses on adjusted criterion groups (Table 8)

Next, we repeated all analyses with the new criterion groups: 37 students with dyslexia and 98 students without dyslexia. Based on the previous analyses, the remaining 19 students could not be reliably assigned to a criterion group. Therefore, the main difference with the previous analyses is that a less stringent diagnostic criterion for dyslexia and no dyslexia is used. The two new criterion groups were entered in a discriminant analysis and then all 154 students were classified. Thus, also the students that were not included in the criterion groups were classified. These were not included in the tables.

We analysed again six different sets of potential predictors: 12 sum scores of the tests of the MDDDT-A, 10 sum scores of the tests of the STB, 22 sum scores of both test batteries together, a selection of 17 items of the questionnaire, a selection of 22 test items of the MDDDT-A, and these 39 items together. Six discriminant functions were calculated that maximally distinguished between participants with and without dyslexia.

Table 8 shows six classification tables indicating overall classification accuracy of the two criterion groups together (135 participants), sensitivity (percentage of 37 participants with dyslexia that is correctly classified), and specificity (percentage of 98 participants without dyslexia that is correctly classified).

In Discriminant Analysis 1, five tests of the MDDDT-A classified 90% of 135 participants correctly. In Discriminant Analysis 2, two tests of the validated battery classified 90% of 135 participants correctly. In Discriminant Analysis 3, four tests of all tests together classified 96% of 135 participants correctly. In Discriminant Analysis 4, three statements of the self-report questionnaire classified 97% of 135 participants correctly. In Discriminant Analysis 5, three items of all tests of the MDDDT-A classified 84% of 135 participants correctly. In Discriminant Analysis 6, one item of the MDDDT-A and two statements of the self-report questionnaire classified 98% of 135 participants correctly. Comparing the classification results with each other and with the six classification results on the biographical criterion, we can draw several conclusions:

1. A comparison of these six analyses with the previous six ones shows that these second series resulted in higher classification accuracy for all sets of potential predictors.
2. Compared to the previous analyses, more predictors were selected in the discriminant functions. This is remarkable, but it also makes sense because the classification of students with less severe dyslexia probably required more predictors in the prediction formulas.
3. All conclusions of the previous analyses can be drawn here as well. It is again remarkable that small sets of single items result in almost perfect classification accuracy in analysis 4 and analysis 6.

Table 8
Discriminant analyses (cross-validated) on two adjusted criterion groups:
37 participants with dyslexia (D)
98 participants without dyslexia (ND)

Discriminant analysis 1: Sum scores of the MDDDT-A
Potential predictors: 12 sum scores of 12 tests
Selected predictors in discriminant function:
1. Dutch Dictation
2. English Dictation
3. Missing Letters
4. Sound Deletion
5. Incorrect Spelling

Classification accuracy (135 participants): 90%
Sensitivity (37 participants with dyslexia): 76%
Specificity (98 participants without dyslexia): 95%

Classification of 28 dyslexics and 72 non-dyslexics:

	Predicted as:	
	D	ND
37 D	28	9
98 ND	5	93

Discriminant analysis 2: Sum scores of the STB
Potential predictors: 10 sum scores of 10 tests
Selected predictors in discriminant function:
1. Klepel Test
2. Amsterdam Sound Deletion Test

Classification accuracy (135 participants): 90%
Sensitivity (37 participants with dyslexia): 86%
Specificity (98 participants without dyslexia): 91%

Classification of 28 dyslexics and 72 non-dyslexics:

	Predicted as:	
	D	ND
37 D	32	5
98 ND	9	89

Discriminant analysis 3: Sum scores of all tests of both batteries
Potential predictors: 22 sum scores of 22 tests
Selected predictors in discriminant function:
1. Dutch Dictation
2. Missing Letters
3. Sound Deletion
4. Klepel Test
5. Digit Span Backward
6. Amsterdam Sound Deletion Test

Classification accuracy (135 participants): 96%
Sensitivity (37 participants with dyslexia): 89%
Specificity (98 participants without dyslexia): 98%

Classification of 28 dyslexics and 72 non-dyslexics:

	Predicted as:	
	D	ND
37 D	33	4
98 ND	2	96

Discriminant analysis 4: Items of questionnaire
Potential predictors: 17 statements
Selected predictors in discriminant function:
1. When I'm writing, I often swap letters.
2. I like to play word games.
3. When I have to transcribe a word repeatedly, it happens that I write it down differently every time.
4. When making a dictation I almost automatically write down the words without mistakes.

Classification accuracy (135 participants): 97%
Sensitivity (37 participants with dyslexia): 92%
Specificity (98 participants without dyslexia): 99%

Classification of 28 dyslexics and 72 non-dyslexics:

	Predicted as:	
	D	ND
37 D	34	3
98 ND	1	97

Discriminant analysis 5: Items of MDDDT-A
Potential predictors: 22 items
Selected predictors in discriminant function:
1. Dutch Dictation: Het nieuwe apparaat werd ijverig bestudeerd.
2. Dutch Dictation: De hachelijke onderneming leidde tot twijfel.
3. English Dictation: That woman is my teacher, she is very strict.
4. English Dictation: I write my homework in my notebook.
5. Missing Letters: De goo … aar gaf een mooie voorste … . (goochelaar, voorstelling)
6. Sound Deletion: Schermusteling
7. Short-term Memory Backward: 3952

Classification accuracy (135 participants): 84%
Sensitivity (37 participants with dyslexia): 65%
Specificity (98 participants without dyslexia): 92%

Classification of 28 dyslexics and 72 non-dyslexics:

	Predicted as:	
	D	ND
37 D	24	13
98 ND	8	90

Discriminant analysis 6:	Items of MDDDT-A and questionnaire
Potential predictors:	39 items

Selected predictors in discriminant function:
1. When I'm writing, I often swap letters.
2. I like to play word games.
3. When I have to transcribe a word repeatedly, it happens that I write it down differently every time.
4. When making a dictation I almost automatically write down the words without mistakes.
5. English Dictation: That woman is my teacher, she is very strict.
6. English Dictation: I write my homework in my notebook.
7. Missing Letters: De goo … aar gaf een mooie voorste … . (goochelaar, voorstelling)
8. Sound Deletion: Schermusteling

Classification accuracy (135 participants):	98%
Sensitivity (37 participants with dyslexia):	95%
Specificity (98 participants without dyslexia):	99%

Classification of 28 dyslexics and 72 non-dyslexics:

	Predicted as:	
	D	ND
37 D	35	2
98 ND	1	97

Final classification of people with and without dyslexia (No table)

In the next section, we will describe the results of analyses of group differences and predictive validity of factors of dyslexia. Therefore, we made a final classification groups of students by categorizing as many students as possible into groups with and without dyslexia.

We reordered the groups with the same demands as before. Participants are considered to have dyslexia when they are categorised as such in at least four out of six predictions. Participants are considered not to have dyslexia when they are categorised as such in at least five out of six predictions. The remaining participants are not categorised in either group. This resulted in 40 participants with dyslexia, 96 participants without dyslexia, and 18 without categorisation.

For a final classification, we required consistency between the two classifications (Tables 7 and 8). This resulted in 34 participants with dyslexia, 95 participants without dyslexia and 25 participants who could not be classified with high reliability.

Next, we compared this final classification with the classification based only on biographical information. Of the 28 participants with dyslexia according the biographical information, 25 were categorised as having dyslexia and 3 could not be categorised. Of the 72 participants not having dyslexia according biographical information, 62 were categorised as not having dyslexia and 10 could not be categorised. Of the group of 54 participants that could not be categorised with biographical information, 9 were now categorised as having dyslexia and 33 as not having dyslexia, while 12 participants still could not be categorised. We conclude that the classification based on two sets of six predictions was highly consistent with the biographical information.

Validity of eight dyslexia factors: three analyses

Introduction

We conducted three additional analyses to find support for the legitimacy of the factors of the different analyses described above: group differences, predictive accuracy, and factor analyses of the predictors of dyslexia in prediction formulas.

Group differences (Tables 9 and 10)

We distinguished eight latent variables in the data set of this study. However, if people with and without dyslexia do not differ on these variables, these variables may not be involved in dyslexia. Therefore, group differences on both tests and factors were investigated with the final classification from the predictive analyses (34 people with dyslexia and 95 people without dyslexia). Effect sizes (Cohen's *d*) were calculated by dividing the mean difference by the mean standard deviations over the two groups.

On all tests (see Table 9), students without dyslexia performed significantly better than students with dyslexia ($p < 0.05$). Mean effect size for the tests of the MDDDT-A was 1.07. Mean effect size for the tests of the STB was 1.16. Such effect sizes are generally considered as very high. We also note that most standard deviations are higher for the group of people with dyslexia as compared to the group of people without dyslexia.

Furthermore, we analysed the differences in factor scores of the latent variables which were found in the full set of tests and questions (see Table 10). On six of eight factors, people with dyslexia had significantly lower factor scores than people without dyslexia (not on phonology and complexity). Mean effect size of these six factors is 0.87. This is generally considered as high. We also note that all standard deviations are higher for the group of people with dyslexia as compared to the group of people without dyslexia.

Table 9

Differences between people with and without dyslexia on 22 tests (Z-scores)

	No dyslexia (N=95)		Dyslexia (N=34)			
	Mean	SD	Mean	SD	Effect size	*p*
MDDDT-A						
Dutch Dictation	0.49	0.60	-1.12	1.09	1.91	<0.0001
English Dictation	0.37	0.44	-0.90	1.48	1.32	<0.0001
Missing Letters	0.44	0.61	-0.95	1.21	1.53	<0.0001
Pseudowords	0.33	0.80	-0.74	1.14	1.10	<0.0001
Sound Deletion	0.16	0.80	-0.58	1.51	0.64	0.001
Spoonerisms	0.25	0.86	-0.85	1.02	1.17	<0.0001
Incorrect Spelling	0.39	0.74	-1.03	1.08	1.56	<0.0001
Dutch-English Rhyme Words	0.15	0.84	-0.41	1.23	0.54	0.004
Letter Order	0.15	0.89	-0.28	1.10	0.43	0.024
Counting Letters	0.22	0.85	-0.34	1.13	0.57	0.003
Short-term Memory Test Forward	0.31	0.94	-0.69	0.98	1.04	<0.0001
Short-term Memory Test Backward	0.31	0.85	-0.54	0.83	1.01	<0.0001
STD						
One-Minute-Test	0.33	0.82	-0.95	0.96	1.44	<0.0001
Klepel Test	0.48	0.71	-1.19	0.79	2.23	<0.0001
Visual Attention Span	0.27	0.82	-0.74	1.12	1.04	<0.0001
Amsterdam Sound Deletion Test	0.36	0.93	-0.94	0.59	1.71	<0.0001
Digit Span Forward	0.13	1.00	-0.32	1.02	0.45	0.027
Digit Span Backward	0.16	0.93	-0.36	0.99	0.54	0.007
Rapid Automized Naming (RAN):						
RAN Letters	0.28	0.89	-0.76	1.04	1.08	<0.0001
RAN Numbers	0.31	0.86	-0.93	0.91	1.40	<0.0001
RAN Colours	0.21	0.91	-0.61	1.15	0.80	<0.0001
RAN Pictures	0.23	0.91	-0.59	0.99	0.86	<0.0001

Table 10

Differences between people with and without dyslexia on 8 factors

	No dyslexia (N=95)		Dyslexia (N=34)			
Factors of Dyslexia	Mean	SD	Mean	SD	Effect size	*p*
Spelling	0.44	0.60	-0.94	1.21	1.52	<0.0001
Rapid naming	0.21	0.80	-0.44	1.23	0.64	0.001
Reading	0.16	0.81	-0.55	1.18	0.71	<0.0001
Attention	0.19	0.87	-0.50	1.19	0.67	0.001
Short-term memory/Confusion	0.25	0.88	-0.65	0.92	1.00	<0.0001
English (?)	0.15	0.82	-0.54	1.31	0.65	0.001
Phonology	0.01	0.86	-0.10	1.47	0.09	0.598
Complexity (?)	0.04	0.84	-0.02	1.37	0.05	0.763

Predictive validity of factors (Table 11)

Predictive validity of the factors was investigated with discriminant analysis (stepwise method, cross-validated) (see Table 11). We entered the eight factor scores in a discriminant analysis as potential predictor. Six factors were used in the regression equation: spelling, rapid naming, reading, attention, short-term memory/confusion and English. Classification accuracy of the whole group was 98% (cross-validated) with a sensitivity of 94% and a specificity of 99%. Apparently, the best prediction of dyslexia is achieved by using many aspects of dyslexia. Consistent with analyses of group differences, phonology and complexity are in this study not necessary for distinguishing between students with and without dyslexia.

Table 11
Discriminant analyses (cross-validated) of people with dyslexia (34) and without dyslexia (95)

Potential predictors:	8 factor scores
Selected predictors in discriminant function:	
1. Spelling	
2. Rapid naming	
3. Reading	
4. Attention	
5. Confusion / Short-term memory	
6. English (?)	
Classification accuracy (129 participants):	97.7%
Sensitivity (34 participants with dyslexia):	94.1%
Specificity (95 participants without dyslexia):	98.9%

Classification of 28 dyslexics and 72 non-dyslexics:

	Predicted as:	
	D	ND
34 D	32	2
95 ND	1	94

Factor structures in discriminant functions (Tables 12–14)

We performed six different predictions twice, once with biographical criterion groups and once with adjusted criterion groups. For convenience, we will call the two different sets of criterion groups Sample 1 and Sample 2. The sixth prediction (all items of the tests of the MDDDT-A and all questions) was most successful, but this was perhaps partly so because the number of potential predictors in the discriminant analysis was higher than the size of the smallest group of participants. We will analyse construct validity of all discriminant functions one by one.

Discriminant function 1 (MDDDT-A) consisted of five tests in both samples. In Sample 1, spelling, phonology and attention were represented. In Sample 2, the test *Counting Letters* was replaced by *English Dictation*, eliminating attention.

Discriminant function 2 (STB) consisted of two tests in both samples. Phonology was represented by the *AKT*. The *Klepel Test* loads on two components: rapid naming and attention.

Discriminant function 3 (all tests) consisted of four tests in Sample 1, but of six tests in Sample 2. The difference between discriminant functions 1 and 2 was spelling and attention. In discriminant function 3, both were represented. In Sample 2, short-term memory was added as well. In summary, a discriminant function acquired with all tests as potential predictor resulted in highest classification accuracy in Sample 2 with high construct validity: only the factor confusion was not necessary for the best prediction.

Discriminant function 4 (items of questionnaire) consisted of three items in Sample 1 and of four items in Sample 2. Table 12 shows the correlations between the five different items and the eight factors found previously. Only spelling and short-term memory/ confusion show high correlations. But the main conclusion is that variance of all statements is explained by more than one factor. This probably explains their selection in the discriminant function: these items somehow just hit the essence of dyslexia. Another main conclusion is that the first six factors – the factors which were supported by analyses of group differences and predictive analyses – are all well supported by more than one statement. Thus, both sets of statements (in both samples) represent six factors of dyslexia.

Discriminant function 5 (items of MDDDT-A) consisted of three items in Sample 1 and of seven items in Sample 2. Table 13 shows the correlations between the eight different items and the eight factors found previously. Compared to the self-report statements, the test items performed less well in representing the factors of dyslexia. This is consistent with much lower predictive validity: classification accuracy of 84% (self-report statements: classification accuracy of 97%). Apparently, relatively low construct validity in test items results in relatively low predictive validity.

Discriminant function 6 (all items) consisted of three items in Sample 1 and of eight items in Sample 2. No novel items were selected for the discriminant function as compared to the previous discriminant functions. Table 14 shows that – as in discriminant function 4 – variance of all statements is explained by more than one factor and that the first six factors all are well supported by more than one statement. Test items do contribute to construct validity (and thus to predictive validity), but only in combination with self-report statements.

Table 12

Correlations between 8 factors of dyslexia and self-report statements of discriminant function 4

(N = 154)

	Factor Spelling	Factor Rapid Naming	Factor Reading	Factor Attention	Factor Short-term memory/ Confusion	Factor English	Factor Phonology	Factor Complexity
When I'm writing, I often swap letters. (Sample 1 & 2)	.32**	.24**	.13	.03	.54**	.23**	.06	.00
I like to play word games. (Sample 2)	.48**	.14	.21*	.26**	.17*	.07	-.01	.07
Reading difficult Dutch words does not cause a lot of problems for me. (Sample 1)	.51**	.18*	.19*	.06	.32**	.24**	.12	.17*
When I have to transcribe a word repeatedly, it happens that I write it down differently every time. (Sample 2)	.38**	.09	.17*	.27**	.51**	.19*	.04	.07
When making a dictation, I almost automatically write down the words without mistakes. (Sample 1 & 2)	.68**	.22**	.18*	.18*	.28**	.29**	.06	.17*

* Correlation is significant at the 0.05 level (two-talied)
** Correlation is significant at the 0.01 level (two-talied)

Table 13

Correlations between 8 factors of dyslexia and test items (MDDDT-A) of discriminant function 5

(N = 154)

	Factor Spelling	Factor Rapid Naming	Factor Reading	Factor Attention	Factor Short-term memory/ Confusion	Factor English	Factor Phonology	Factor Complexity
Dutch Dictation: Het nieuwe apparaat werd ijverig bestudeerd. (Sample 2)	.45**	-.04	.11	.17*	.18*	.04	-.15	.02
Dutch Dictation: De hachelijke onderneming leidde tot twijfel. (Sample 1 & 2)	.40**	.13	.10	.08	.05	.05	-.02	.07
English Dictation: That woman is my teacher, she is very strict. (Sample 1 & 2)	.42**	.09	.31**	.03	.10	.21**	-.07	.11
English Dictation: I write my homework in my notebook. (Sample 2)	.35**	-.17*	.13	-.03	.13	.08	-.15	-.11
Missing Letters: De goo … aar gaf een mooie voorste … . (goochelaar, voorstelling) (Sample 2)	.19*	.08	.15	.12	.24**	.03	-.14	-.22**
Sound Deletion: Schermusteling (Sample 2)	.26**	.04	-.04	.11	.11	.19*	.15	-.12
Short-term Memory Forward: 382986 (Sample 1)	.18*	.04	.02	-.08	.25**	-.02	.05	.08
Short-term Memory Backward: 3952 (Sample 2)	.26**	-.01	.02	.14	.24**	-.06	-.01	.07

* Correlation is significant at the 0.05 level (two-talied)
** Correlation is significant at the 0.01 level (two-talied)

Table 14

Correlations between 8 factors of dyslexia and self-report statements and test items (MDDDT-A) of discriminant function 6
(N = 154)

	Factor Spelling	Factor Rapid Naming	Factor Reading	Factor Attention	Factor Short-term memory/ Confusion	Factor English	Factor Phonology	Factor Complexity
When I'm writing, I often swap letters. (Sample 1 & 2)	.32**	.24**	.13	.03	.54**	.23**	.06	.00
I like to play word games. (Sample 2)	.48**	.14	.21*	.26**	.17*	.07	-.01	.07
When I have to transcribe a word repeatedly, it happens that I write it down differently every time. (Sample 2)	.38**	.09	.17*	.27**	.51**	.19*	.04	.07
When making a dictation, I almost automatically write down the words without mistakes. (Sample 1 & 2)	.68**	.22**	.18*	.18*	.28**	.29**	.06	.17*
English Dictation: That woman is my teacher, she is very strict. (Sample 1 & 2)	.42**	.09	.31**	.03	.10	.21**	-.07	.11
English Dictation: I write my homework in my notebook. (Sample 2)	.35**	-.17*	.13	-.03	.13	.08	-.15	-.11
Missing Letters: De goo … aar gaf een mooie voorste … . (goochelaar, voorstelling) (Sample 2)	.19*	.08	.15	.12	.24**	.03	-.14	-.22**
Sound Deletion: Schermusteling (Sample 2)	.26**	.04	-.04	.11	.11	.19*	.15	-.12

*   Correlation is significant at the 0.05 level (two-talied)
** Correlation is significant at the 0.01 level (two-talied)

Summary

Overall, the legitimacy of at least six factors was found (spelling, rapid naming, reading, attention, short-term memory/confusion and English). These six factors showed significant group differences and were used in a discriminant function. Remarkably, phonology is not very well supported by these analyses as being a key feature of dyslexia. However, some relations between the tests and these factors or between the factors of tests and questions remain unclear. Furthermore, construct validity is best in the discriminant functions of Sample 2, which justifies repeated predictions. The most important finding is that predictive validity is higher when construct validity is high.

Discussion

Construct validity

The first question of this study was how components of the newly developed MDDDT-A are related to components of the STB, a battery of standard tests covering the key features of dyslexia. We investigated construct validity of both test batteries with various principal components analyses and correlations and found that many well-known deficits of dyslexia are represented by the MDDDTA. The tests of the MDDDT-A did not include measures of rapid naming and reading. However, the self-report questions of the MDDDT-A did identify all deficits of dyslexia that were investigated in this study. Thus, construct validity of the self-report questions was higher than both test batteries in the present study. The conclusion is that construct validity of the MDDDT-A is high.

In one of our previous studies (Tamboer, Vorst & Oort, 2016), we found that five factors of dyslexia could be distinguished: spelling, phonology, short-term memory, rhyme/confusion and whole-word-processing/complexity. In the present study, we identified four factors in the tests of the MDDDT-A, three factors in the STB, six factors in the full set of tests, eight factors in the questionnaire, and eight factors when we combined the tests and the questionnaire. We also found three areas with inconsistencies. First, not all factors showed group differences. Second, some factors were separate factors in one analysis and taken together as one factor in another analysis. And third, some tests showed loadings on distinct factors in different analyses. For instance, the *One Minute Test* and the *Klepel Test* of the STB loaded on rapid naming and productive phonology in one analysis, but on reading in another analysis.

Our main aim was to find a single analysis that clearly shows which factors characterize dyslexia best. The fact that the number of selected components varies between analyses is not only a result of the number of latent variables that are present, but also of the choices that had to be made during the analyses. For instance, choices had to be made about the lowest eigenvalue or the amount of variance that required explanation. Additionally, we were confronted with various difficulties such as losing variance when using factor scores, or incorporating more variables in the analysis than could be justified based on a relatively small sample size. Another difference between our previous and present study is found in the samples used. Previously, we performed principal components analyses on a general sample of students, in which the number of students with dyslexia was in accordance with the general prevalence rate, whereas in the present sample the percentage of students with dyslexia was relatively high.

The best way to interpret all analyses is to have an overview and detect the latent variables that seem to be at play in more than one analysis. Then, we can distinguish nine factors in total: Spelling, reading, rapid naming, attention, short-term memory, confusion, phonology, complexity, and learning English. Recent literature showed that adult dyslexia can be characterized by various symptoms (e.g. Callens et al., 2014; Cavalli et al., 2016; Kemp, Parrila & Kirby, 2009; Nergård-Nilssen & Hulme, 2014; Smith-Spark et al., 2016; Vellutio et al., 2004). Some of the factors of the present study can easily be related to one or more of these symptoms, but others are more difficult to interpret. We believe that specific conclusions can only be made with the greatest caution.

Regarding spelling, we found that spelling difficulties are well supported by the MDDDT-A. Spelling difficulties were found to be a separate factor in most analyses and apparently constitute a key symptom of dyslexia in adults. This is consistent with previous findings that spelling impairments are relatively resistant to improvement as compared to other impairments (Nergård-Nilssen & Hulme, 2014). However, the exact nature of the relations between spelling and some other symptoms (reading, rapid naming, and attention) remain unresolved in the present study.

A remarkable high correlation was found between the spelling factor of the MDDDT-A and the productive phonology factor of the STB. This productive phonology factor was determined by various tests in the STB: two reading tasks (*One Minute Test* and *Klepel Test*), one phonological task (*Amsterdam Sound Deletion Task*), two rapid naming tasks (*RAN*

Letters and *RAN Numbers*) and one visual attention task (*Visual Attention Span*). The loading of visual attention span on this phonological factor seems strange, but the result is in line with previous findings (van den Boer, van Bergen & de Jong, 2015) suggesting that phonology is involved in this task. The composition of this productive phonology factor clearly points to a complex nature of both spelling abilities and phonological abilities. Note that in the combined analysis of the tests and self-report questions, these factors were separated again.

Regarding short-term memory, the main conclusion from the present study is that it is problematic to find relations between this type of difficulty and other difficulties. Often, the role of working or short-term memory in dyslexia is related to phonological difficulties. However, the relations between working or short-term memory and the typical reading difficulties that accompany dyslexia are harder to understand. For a clear description of the relations between working memory and language, see Gathercole and Baddeley (2014). A remarkable finding in the present study was that short-term memory of both test batteries highly correlated with each other. In the STB, short-term memory was assessed auditory, whereas in the MDDDT-A this was assessed visually. Apparently, this difference does not interfere with what the tests in both batteries aim to measure.

Regarding phonology, we distinguished productive and receptive phonology in different analyses. A phonological factor was found in the final principal components analysis, however, without showing a group difference. The phonological task the *Amsterdam Sound Deletion Test* of the STB loaded in the final analysis on a factor named complexity. Possibly, phonology is at play in various domains of dyslexic's difficulties. This is consistent with the distinction between productive and receptive phonology in the analysis where only the tests were involved. The tests related to phonology in the MDDDT-A were not completely auditory. They had to be filled out visually on a computer screen after listening to audio through headphones. In contrast, the tests related to phonology in the STB required the production of speech. Unresolved also remains that phonological tests showed group differences while the final phonology factor did not. Clearly, future analyses are required to clarify these issues of phonology in relation to the MDDDT-A.

Three factors were hard to interpret: confusion, complexity and learning English. Confusion and complexity were also found in our previous study, but learning English was not. However, of these three, only learning English showed significant group differences. We have tried to interpret them as best as we can, but, we cannot rule out that some tests do not actually measure what they aim to measure and in fact measure something else related to language and dyslexia. For example, *Spoonerisms* is only supposed to measure phonology, but it also appeared to be related to spelling and attention.

The factor learning English first came forward from the questionnaire, but it appeared to correlate with *Spoonerisms* and *Incorrect Spelling*. An explanation for why this factor plays a significant role in adult dyslexia in the Netherlands may be that high performing students in the Netherlands could have overcome some of their difficulties with the Dutch language, but not with foreign languages such as English. In the Netherlands, it is a well-known effect for high performing students who have been diagnosed with dyslexia in primary school to experience difficulties with foreign languages all over again when attending secondary school. This is especially true for English because of the lack of transparency between its spelling and its pronunciation and thus higher demands on phonological abilities. The factor learning English did not load on the spelling factor of the tests which included both Dutch and English spelling. Maybe the test *English Dictation* did not represent the severe difficulties that people with dyslexia experience when for instance reading a scientific English textbook. In the questionnaire items refer to these severe difficulties. Some of the spelling items of *English Dictation* may be relatively easy for high performing students. Whether this is the case in other samples remains unknown for now. Therefore, we decided to include *English Dictation* in the MDDDT-A for further analyses in other samples. In a general population, it may be expected that *English Dictation* is relatively difficult to perform as compared to *Dutch Dictation*.

Although confusion and complexity were significant factors in our previous study, they were hard to interpret in the present study. Both came forward as such in the questionnaire. Self-reported confusion correlated to some extent with short-term memory in tests. An interpretation is that short-term memory difficulties are not exclusively memory difficulties but arise from more general working memory deficits such as quick processing of single digits or letters. On the same factor confusion, various tests showed low loadings, such as *Missing Letters*, *Pseudowords*, and *Dutch-English Rhyme Words*. Processing pseudowords strongly depends on the ability to process the sequences of letters or the ability to retrieve from memory the right place of each letter. The factor confusion in the questionnaire is partly determined by items that measure difficulties of for instance exchanging letters or words. Maybe, during quick processing of single digits or letters there is a problem with the quick organization of sounds and letters resulting in for instance exchanging letters and words. This may point to the existence of more general cognitive deficits that accompany dyslexia which are not yet fully understood. This is consistent with the finding of deficits of various executive functions (inhibition, set shifting) in a recent study (Smith-Spark et al., 2016).

The same may be the case with complexity. In the present study, a self-report of complexity in language correlated with sound deletion in the STB. In our previous study, complexity correlated with the processing of whole words, represented by the tests *Spoonerisms* and *Letter Order* in this study. These two tests are recent developments and require that participants perform something that they never did before. Thus, for these tests something new had to be learned and executed. Although speculative, this points to a general difficulty accompanying dyslexia which would allow for dyslexia to be described as a learning disorder instead of solely a language disorder. The idea of a general learning deficit underlying dyslexia is not new. According to the procedural learning hypothesis (Nicolson & Fawcett, 2007; Nicolson et al., 2010) dyslexia is the result of impairments in functional networks in the brain including the cerebellum, which permits various other secondary deficits outside the literacy domain while declarative learning remains relatively intact. However, this hypothesis has also been criticized (West et al., 2018).

In summary, we conclude that this study supports the idea that dyslexia is characterized by various impairments in adults, but that the exact nature and relations among these impairments remain unclear. In the present study, some relations or a lack of relations can be the result of differences between the nature of assessments: the digitised MDDDT-A, the STB requiring oral answers and the self-report of difficulties. Nevertheless, many correlations that came forward in this study point towards the existence of general cognitive aspects that accompany adult dyslexia, of which some are better understood and described in the literature than others.

Predictive validity

The second question of this study was how well the MDDDT-A can predict dyslexia in students as compared to the STB. The results of the tests of the MDDDT-A and the STB were about the same, both with a strict criterion as well as with a more flexible criterion. This came as no surprise because we found a high construct validity of the tests of the MDDDT-A as well as the STB, even though there were some differences. With a flexible criterion, the tests of the MDDDT-A provided a classification accuracy of 90% (five tests in the discriminant function), and the tests of the STB also provided a classification accuracy of 90% (but with only two tests in the discriminant function). This is slightly lower than the classification accuracy (91%) with three tests (word reading, word spelling and phonological awareness) in a study of Tops et al. (2012). The same classification accuracy of 90% was reported in a study for the Norwegian language (Nergård-Nilssen & Hulme, 2014) using six tests in logistic regression analysis (text reading fluency, nonword reading, word spelling, phoneme awareness, working memory, and rapid automatized naming). When the tests of the MDDDT-A and the STB were taken together, this resulted in an even higher classification accuracy of 96% (six tests in the regression equation). The similarity between the Norwegian study and the present study is striking. In both studies, the best combination of predictors of dyslexia were nonword reading (*Klepel Test*), phoneme awareness (*Sound Deletion*, *Amsterdam Sound Deletion Test*), working memory (*Digit Span Backward*), and word spelling (*Dutch Dictation*). A remarkable finding of the present study was that factor scores showed an even higher classification accuracy of 98% (six factor scores in the regression equation). These results point to the general hypothesis that higher classification accuracy is achieved when more separate difficulties of dyslexia are incorporated in the form of factor scores. In factor scores, task and method variance is probably eliminated, thereby resulting in a higher construct validity.

Another finding of the present study regarding predictive validity is that it is higher when a flexible criterion is used than when a strict criterion is used. This was found in all six predictions, meaning that the diagnosis of dyslexia depends on the nature and/or severity of the dyslexia as found in the adults used in the criterion groups. A strict biographical criterion probably only selects the adults with severe dyslexia and adults who are high performing in general. A less stringent biographical criterion probably also includes adults with less severe symptoms of dyslexia or adults who only suffer from a few symptoms. It therefore makes sense to assume that most people with dyslexia are to be found in the middle areas of a normal distribution of severity of dyslexia. Comparing this large group with only people on the far left and right end of the distribution may lead to false positives and false negatives. People on the extreme ends of the distribution cannot only be characterised by having dyslexia or not, but also by having or lacking the ability to compensate for dyslexia.

We also emphasize that higher predictive validity with a flexible criterion was based on more predictors in the regression equation than when a strict criterion was used. An interpretation is that strict criterion groups are relatively homogeneous. This might explain that in some studies only a few difficulties of people with dyslexia need to be assessed for a reliable diagnosis. People with dyslexia in one specific and homogeneous sample may resemble each other in applying compensation strategies for some difficulties while retaining other difficulties. Thus, homogeneous samples can be reliably diagnosed with only a few tests, while heterogeneous samples need more tests for a reliable prediction. Based on the results of the present study, it cannot be established how many and which difficulties should be assessed for high predictive validity in general. This depends on the characteristics of a specific sample which may vary to a large extent across countries. Moreover, also cut-off scores may depend on the characteristics of a specific sample. Therefore, it is recommended to use regression analyses such as logistic regression or discriminant analysis. These analyses determine which tests or questions discriminate best and determine weights for each test or item, which is a better solution than using cut-off scores. Taken one step further, it might even be concluded that not the quality of the different test batteries determines predictive validity, but that predictive validity depends on the method used for analysing the test or item scores.

The third question of this study was to what extent a self-report questionnaire can contribute to the prediction of dyslexia. Self-report questions can contribute to construct validity, especially in a digitised battery that cannot include the standard oral measures of reading and rapid naming. This is an advantage mainly because predictive validity seems to depend on construct validity, which was concluded above, and aligns with the conclusion of Nergård-Nilssen & Hulme (2014), and is recommended by Callens et al. (2014). Furthermore, this was confirmed by predictive validity of the questionnaire itself. Construct validity of the questionnaire was high with eight latent variables, and predictive validity was high with classification accuracy of 97% (with a flexible criterion) being even higher than classification accuracies of the tests, although factor scores provided 98% accuracy. When single test items were combined with self-report statements, classification accuracy was 98%. This result is questionable though, because more potential predictors were used than participants in the smallest group of the sample. Nevertheless, this raises the question whether tests should be used at all, if self-report questions by itself are able to predict dyslexia so well. However, it should be noted that this result came from a high achieving sample of students. We cannot yet be sure whether only a questionnaire would suffice for reliable diagnoses of dyslexia in other samples as well. Finally, seeing that the questionnaire resulted in high classification accuracy, it is remarkable that only four self-report statements were used in the discriminant function. This seems to contradict high construct validity, but the correlational analysis between these statements and the eight factors indicated that these four statements together represented almost all factors. For instance, the statement ‘When making a dictation, I almost automatically write down the words without mistakes’ showed a significant correlation with seven factors. Apparently, some single self-report statements capture the exact nature of certain difficulties of adults with dyslexia, while single tests do not. The finding that single statements predict better than test scores can be explained by the fact that the results of test scores not only depend on the typical difficulties of adults with dyslexia, but also on intelligence and schooling.

Conclusion and limitations

We validated the MDDDT-A and found that construct validity and predictive validity are very good. We also found that highest predictive validity is achieved when construct validity is high, and when a flexible criterion is used. A remarkable result was that self-report questions have high construct validity and an even higher predictive validity than tests. These results show that the MDDDT-A can be used online as a reliable screening instrument without requiring much time. Theoretically, we found support for dyslexia as a multiple cognitive deficit, although it remains unclear how many independent impairments can be distinguished. A limitation of this study was that sample size was too small for conducting confirmatory analyses. A second limitation was that we only investigated students in our sample. A third limitation was that reading and rapid naming could not be assessed with digitised tests, although this was compensated by self-report assessment of these abilities. Screening instruments should account for differences between samples, but we expect that when construct validity is kept as high as possible (for instance by assessing both tests and self-report questions), predictive validity will be high in samples within and across countries. We recommend using regression analyses to determine which tests and items and which weights are most appropriate for high predictive validity in a specific sample. One symptom may be important in one sample, but less important in another sample because the severity of any symptom of dyslexia may vary between languages and between samples of any kind.

6

Dyslexia and Voxel-Based Morphometry: Correlations between five behavioural measures of dyslexia and gray and white matter volumes

In voxel-based morphometry studies of dyslexia, the relation between causal theories of dyslexia and gray matter (GM) and white matter (WM) volume alterations is still under debate. Some alterations are consistently reported, but others failed to reach significance. We investigated GM alterations in a large sample of Dutch students (37 dyslexics and 57 non-dyslexics) with two analyses: group differences in local GM and total GM and WM volume and correlations between GM and WM volumes and five behavioural measures. We found no significant group differences after corrections for multiple comparisons although total WM volume was lower in the group of dyslexics when age was partialled out. We presented an overview of uncorrected clusters of voxels ($p < 0.05$, cluster size $k > 200$) with reduced or increased GM volume. We found four significant correlations between factors of dyslexia representing various behavioural measures and the clusters found in the first analysis. In the whole sample, a factor related to performances in spelling correlated negatively with GM volume in the left posterior cerebellum. Within the group of dyslexics, a factor related to performances in Dutch–English rhyme words correlated positively with GM volume in the left and right caudate nucleus and negatively with increased total WM volume. Most of our findings were in accordance with previous reports. A relatively new finding was the involvement of the caudate nucleus. We confirmed the multiple cognitive nature of dyslexia and suggested that experience greatly influences anatomical alterations depending on various subtypes of dyslexia, especially in a student sample.

Keywords: Dyslexia, MRI, VBM, white matter, gray matter, cerebellum, caudate nucleus

Based on: Tamboer, P., Scholte, H. S., & Vorst, H. C. M. (2015). Dyslexia and voxel-based morphometry: Correlations between five behavioural measures of dyslexia and gray and white matter volumes. *Annals of Dyslexia*, *65*(3), 121-141.

Introduction

Dyslexia has been described as a neurological disorder with a genetic origin characterised by poor reading and spelling abilities despite adequate intelligence, motivation and schooling. Dyslexia is persistent into adulthood, often regardless of remedial teaching during school days or other childhood interventions. Estimates of prevalence vary widely between 3 and 18 %. After decades of investigating the cognitive impairments of dyslexic people (e.g. Ramus & Ahissar, 2012), an important question in recent years has been whether structural and functional abnormalities in the brain can be identified in relation to dyslexia.

In this study, we address the issue of structural alterations in the brain in terms of anatomical brain morphology. A much-applied technique for analysing anatomical structures in the brain is voxel-based morphometry (VBM) (Ashburner & Friston, 2000; Wright et al., 1995), which specifies gray matter (GM) and white matter (WM) densities of separate voxels. Unfortunately, in VBM studies of dyslexia, many findings failed to be replicated or were rendered insignif-icant (statistically) by corrections for multiple comparisons. As a result, there is much discussion about the generalisability of findings.

Besides this discussion, some findings appear to be consistent across studies and much has already been learned. Two meta-analytical studies were reported in 2012, examining local GM alterations in relatively small samples of dyslexic adults. A coordinate-based meta-analysis (Richlan, Kronbichler & Wimmer, 2013) of nine VBM studies reporting 43 foci of GM reduction and 2 foci of GM increase in dyslexic readers (total sample sizes, 134 dyslexic and 132 non-impaired mostly adult readers, 11–41 years) resulted in the convergence of GM reductions in only two relatively small areas: one in the right superior temporal gyrus and one in the left superior temporal sulcus. No significant differences in whole brain GM or WM volume were reported. An activation likelihood estimation meta-analysis (Linkersdörfer et al., 2012) of nine VBM studies reporting 62 foci of GM reduction in dyslexic readers (total sample sizes, 139 dyslexic and 138 non-impaired mostly adult readers) resulted in the convergence of six clusters in bilateral temporo-parietal and left occipito-temporal cortical regions and in the cerebellum bilaterally. Again, no significant differences in whole brain GM or WM volume were reported.

Seven studies were incorporated in both meta-analyses (Brambati et al., 2004; Brown et al., 2001; Eckert et al., 2005; Hoeft et al. 2007; Kronbichler et al., 2008; Steinbrink et al., 2008; Vinkenbosch, Robichon & Eliez, 2005). In the analysis by Richlan et al., a study by

Raschle, Chang & Gaab (2011) was excluded because the participants were prereading kindergarteners with a family history of dyslexia but without diagnosis of dyslexia, and a study by Pernet et al. (2009a) was excluded because they failed to find direct group differences. In the analysis by Linkersdörfer et al., a study by Silani et al. (2005) and a study by Menghini et al. (2008) were not included. The reported coordinates of the areas of convergence were not exactly the same which may be the result of slightly different inclusion criteria of the studies. The largest cluster in the study by Linkersdörfer et al. was found in the left fusiform gyrus extending into the left inferior temporal gyrus, while Richlan et al. found a cluster in the left superior temporal sulcus. Both studies reported a cluster in the right superior temporal gyrus, but Linkersdörfer et al. reported four additional areas in the left and right supramarginal gyrus and in the left and right cerebellum, which failed to reach meta-analytical significance in the study of Richlan et al.

Besides these meta-analyses, findings from functional imaging studies are also relevant for interpreting brain anatomy of dyslexics. Two meta-analytic studies of functional overactivation and underactivations in dyslexics were performed in recent years (Richlan, Kronbichler & Wimmer, 2009, 2011). In Linkersdörfer et al., the results of these studies were used to analyse overlap between structural and functional deviations with additional activation likelihood estimation meta-analyses of imaging studies. Conjunction analyses of the meta-analyses revealed an overlap in the left cerebellum and left fusiform gyrus. Summarising all meta-analytic results, it became clear that some areas are involved in dyslexia with a high degree of certainty. However, the number and size of the areas which survived meta-analytical significance thresholds are small compared to the number and size of all areas reported in the smaller samples of the separate studies. That many areas did not survive significance thresholds does not automatically imply that these are irrelevant for dyslexia.

Support for the significance of some areas that did not survive significance thresholds can be found in the study of Pernet et al. (2009a). In this study, no significant group differences were found in a large sample of 38 dyslexics and 39 non-dyslexics. However, most of the uncorrected p values pointed to areas in accord with previous findings. Moreover, this study reported various significant correlations between GM volumes (41 loci) and behavioural measures (phoneme deletion, irregular word spelling, pseudoword reading), across groups and/or between groups. The 41 loci were found in three main territories: the cerebellum; the ventral visual cortex; and several parts of (mainly) left and dorsal hemispheric brain areas such as superior frontal, medial parietal and superior temporal areas. Thus, in a relatively large sample using correlational analyses, many more areas could be significantly related to dyslexia than in the meta-analyses. In a second study by Pernet et al. (2009b), two predictors of dyslexia were found using a classification approach: the right lentiform nucleus and the right cerebellar declive with dyslexics falling either above or below the control group's 95 % confidence interval boundaries.

In summary, much has been learned about brain anatomy in dyslexia, but two main questions remain under debate: why are more significant alterations identified in studies with smaller samples than in studies with larger samples or in meta-analytical studies, and why are correlational analyses more efficient in identifying anatomical alterations than group analyses? First of all, no explanations can be derived from gender differences. Both in the study by Pernet et al. and in most meta-analytic studies, a large majority of the participants were male. Thus, possible GM differences between dyslexic males and dyslexic females as were observed by Evans et al. (2014) could not explain the differences. On the other hand, the inclusion of a relatively small portion of female dyslexics in the samples might have had an effect on the power of group differences between dyslexics and controls.

A plausible explanation for the fact that most of the direct group differences in separate studies failed to be significant seems to be that the samples in these studies were relatively small (on average, 15 dyslexics and 15 non-dyslexics), resulting in a lack of power. However, it is then strange that other findings in separate studies were significant, while in meta-analytical studies or in a large sample (Pernet et al., 2009a), most findings failed to reach significance. Pernet et al. pointed out that, given the common use of small samples in MRI studies, 'individual variability may have been under- or overestimated … biasing results towards significance in some studies and not in others' (p. 2288). Relatively high individual variability in groups of dyslexics is well established, and the common explanation is that dyslexia is a multiple cognitive deficit. VBM results can then easily be affected by individual variation in perception and cognition in dyslexia (Eckert et al., 2005). Analyses of individual profiles has also resulted in the suggestion that subtypes of dyslexia should be distinguished (e.g. Ramus et al., 2003). Although there is an ongoing discussion about subtypes, there is strong support and consistency for at least two subtypes of dyslexia (Bosse, Tainturier & Valdois, 2007; Di Filippo & Zoccolotti, 2012; Heim et al., 2008): a phonological subtype and a visuo-attentional subtype. Furthermore, different samples of different ages may induce different results because it was found that many GM density differences between dyslexics and controls, in general, result from differences in reading experience (Krafnick et al., 2014; Clark et al., 2014).

In a recent VBM study by Jednoróg et al. (2014) with 46 dyslexic and 35 control children (mean age 10 years), the dyslexic children were split into three subtypes based on various cognitive deficits (phonological, rapid naming, magnocellular/dorsal, auditory attention shifting). Comparing VBM contrasts for the whole group revealed significantly reduced local GM volume for dyslexics compared to controls in the left inferior frontal gyrus – an area that failed to reach significance in the meta-analyses. Most importantly, comparing VBM contrasts of separate subtypes with all other groups revealed different loci of reduced and/or increased GM volume in various areas. This underlines that in VBM studies with small samples, some subtypes of dyslexia may have been overrepresented or underrepresented. It explains why more significant alterations are identified in studies with small samples than in studies with large samples or in meta-analytical studies and why correlational analyses between behavioural measures and local GM volumes are more efficient in identifying anatomical alterations than analyses of group differences. However, it is highly important to realise that the anatomical differences between subtypes of dyslexia in the study by Jednoróg et al. were found in young children while the participants in the meta-analyses were mainly adults. In the meta-analysis of functional MRI studies (Richlan, Kronbichler, & Wimmer, 2011), various differences exist between dyslexic children and adult dyslexics. It is unknown how these differences influence anatomical differences with age. In general, it is also unknown whether age differences have influenced results in the adult samples used for the meta-analyses which varied between 11 and 41 years old.

The aim of the present study was to enlarge the probability of finding significant anatomical differences between dyslexics and non-dyslexics. We used a large sample of 37 dyslexic and 57 non-dyslexic students. This sample was used in a previous study (Tamboer, Vorst & Oort, 2014) in which dyslexics and non-dyslexics were identified on the basis of various cognitive measures. In a second previous study (Tamboer, Vorst & Oort, 2016), the same sample was used for the identification of five cognitive factors related to dyslexia (phonology, spelling, short-term memory, rhyme words/confusion and whole-word reading/complexity). Thus, we took into account that dyslexia is characterised by various cognitive deficits as much as possible. Compared to the studies by Pernet et al. (2009a) who used three behavioural measures and Heim et al. (2008) who used three cognitive factors, we were able to specify relations between cognition and anatomical alterations further. First, we applied analyses of group differences using whole-brain VBM instead of analysing only a priori-determined areas in a region-of-interest (ROI) analysis, because in previous studies, brain abnormalities were reported in various brain areas, and we did not want to run the risk of missing relevant regions by limiting our analyses to ROIs. Second, we applied correlational analyses between loci of GM volume alterations and five cognitive factors. Third, to account for effects of gender, age and handedness, we exploratively performed various additional analyses with these variables as covariates.

Methods

Subjects and procedure

In this study, 37 dyslexic students (six men; six left-handed; mean age 20.61 years, SD 1.53 years) and 57 non-dyslexic students (seven men; eight left-handed; mean age 20.33 years, SD 1.14 years) participated. All participants were first-year psychology students at the University of Amsterdam, and most of them were female as most psychology students in the Netherlands are female. All students were native Dutch speakers, were raised in the Netherlands and had 12–13 years of school education at a school in the Netherlands. All students were free from medical or psychiatric diseases and had no history of sensory deficits or head trauma. ADHD was assessed with a short self-report questionnaire, which included 46 questions about attention, concentration and hyperactivity. A mean score and standard deviation on this questionnaire were calculated in a larger group of more than 1000 students. The groups of dyslexics and non-dyslexics did not differ on ADHD symptoms, and no student of the present sample had a score higher than one standard deviation above the average of the total sample of more than 1000 students. Handedness was assessed with a short self-report questionnaire, which included questions about writing hand, general hand preference and 20 specific questions. There were no students with inconsistent reports which could indicate being ambidextrous. All students who participated in this study gave informed written consent and were debriefed afterwards. All participants had the option to choose between acquiring participation points required for the first year of study or a financial reward. This study was approved by the ethics committee at the University of Amsterdam.

Neuropsychological assessment of dyslexia

This sample was acquired from a sample of students who participated in a previous study (Tamboer, Vorst & Oort, 2014). In that study, dyslexia and non-dyslexia were assessed using three sources of information: a history of language difficulties, a self-report of language difficulties and a test battery.

A history of language difficulties consisted information about persistent language difficulties at school, dyslexic family members and various test results which were assessed during school days. Some students had a formal diagnosis of dyslexia. In the Netherlands, a formal diagnosis of dyslexia can be acquired only from official institutes of dyslexia by specialists in diagnosing dyslexia and is considered to be very reliable.

A self-report of dyslexia was assessed with an extended questionnaire that consisted of two parts. The first part consisted of 30 general questions or statements, such as ‘Are you dyslexic?’, ‘Did you experience difficulties with learning to read and/or to spell during school days?’, ‘I do not like that English sounds differently than it is written’. The second part consisted of 140 specific statements which aim to acquire information about specific language difficulties. The questionnaire was designed according to a 7×5×4 facet design. There are seven subscales representing different aspects of language, each consisting of 20 statements with seven response categories. Here, we give examples for the seven subscales: reading: ‘Sometimes I skip a letter, which results in reading a different word’; writing: ‘Sometimes I forget to write down a syllable’; speaking: ‘While speaking, I sometimes exchange similar words’; listening: ‘I hear a story exactly like someone tells it’; copying: ‘When I copy out a text, I sometimes exchange letters with similar sounds’; dictating: ‘I make mistakes in dictation, because I do not hear the correct sounds’; and reading aloud: ‘When reading aloud, I sometimes repeat a part of the text’. All statements can also be categorised into five subscales representing sounds, letters, words, sentences and text. A leading thought during the creation of these statements was that four typical mistakes might distinguish dyslexics from others: skipping (forgetting), adding, changing and exchanging. These typical mistakes represent various kinds of visual, attentional, auditory or phonological confusion. Furthermore, we accounted for the possibility that these typical mistakes become more prominent when complexity increases, as is the case with pseudowords, for instance.

The test battery consisted of nine language-related tests and a short-term memory test. Together, these ten tests covered all of the known symptoms of dyslexia, such as phonological awareness, rapid naming, attentional/visual processing and short-term memory.

Dutch Dictation (auditory) aims to measure spelling abilities in the Dutch language (ten sentences, maximum score 10×4=40). Each sentence consisted of at least two difficult words.

English Dictation (auditory) aims to measure spelling abilities in the English language (ten sentences, maximum score 10×2=20). It can be assumed that Dutch students are familiar with ordinary English words that we used. Each sentence consisted of at least two difficult words.

Missing Letters (auditory) also aim to measure spelling abilities in the Dutch language (ten sentences, maximum score 10×2=20), but in a slightly different way. For each sentence, two words are repeated while these words are shown on the computer screen with a few letters left out of the word.

Pseudowords (auditory) aim to measure spelling abilities of pseudowords – non-words that sound like real words (30 words, maximum score 30). Participants have to decide whether the non-words that they hear are spelled correctly on the computer screen. Usually, pseudowords are administered the other way around by participants reading the words aloud themselves. We changed this because it would be practically impossible to have all students in private sessions for this way of testing.

Sound Deletion (auditory) aims to measure phonological abilities (20 words, maximum score 20). Participants have to decide whether the difficult Dutch words that they hear are pronounced correctly, and if not, which letter is missing or has been added (there is a choice between three words). For example, the word ‘fietsenstalling’, which means bicycle shed, is read out as ‘fiestenstalling’. The possible answers are as follows: ‘fietsentalling’, ‘fiestensalling’ and ‘fiestenstalling’.

Spoonerisms (auditory) also aim to measure phonological abilities (20 words, maximum score 20). A spoonerism is a word that consists of two existing smaller words and still consists of two small existing words when the first letters of both small words are exchanged. For example, participants hear the word ‘kolen-schop’ which has to be altered to ‘scholen-kop’.

Incorrect Spelling (visual) is the third test in our study that aims to measure spelling abilities in the Dutch language, again in a different way (40 words, maximum score 40). All words are flashed on a computer screen for 50ms. Participants have to decide whether the words are spelled correctly or not.

Dutch–English Rhyme Words (visual) aim to measure the ability to recognise similar-sounding nouns in Dutch and English (40 words, maximum score 40). Dutch–English word pairs are shown on a computer screen with the Dutch words on the right. Participants have to decide whether the words rhyme with each other or not. Typical confusion may arise in this test because the non-rhyming items have the same vowels, such as ‘Deep-Reep’.

Letter Order (visual) aims to measure the ability to read words as a whole (20 sentences, maximum score 20×2=40; time limit of 5 min). We created 20 sentences based on the same principle: the order of the letters of the words was changed, apart from the first and last letters. The words in the sentences are more difficult towards the end of the test. The sentences have to be typed in with all words correctly spelled. There are no words that consist of typical dyslexic spelling difficulties.

The *Short-Term Memory Test* aims to measure the capacity of short-term memory. We used the concept of digit span: the number of digits that a person can retain and recall. There are four subtests: numbers and letters, both forward and backward. And, each subtest consists of 24 series: 6 of 4, 6 of 5, 6 of 6 and 6 of 7 items for the subtest numbers and letters forward and 6 of 3, 6 of 4, 6 of 5 and 6 of 6 items for the subtest numbers and letters backward. The numbers and letters are presented one by one, for 1 s each on a computer screen. The participants have to retype these numbers and letters after the last one of a series has been presented. About half of all series consist of some typical difficulties for dyslexics, either phonological, visual or both. For example, a typical phonological confusion is between the numbers 7 and 9 which resemble each other phonologically in Dutch (zeven/negen). Typical visual confusions are between the numbers 6 and 9 and the letters [m] and [w]. The letters [p], [d] and [b] resemble each other phonologically as well as visually.

For the selection of non-dyslexics in the present study, we required that they had no history of language difficulties, that they had no self-report of dyslexia or related difficulties and that they had no more than three of 12 test scores below average and, if so, not lower than a half standard deviation below average.

For the selection of dyslexics in the present study, we required that dyslexics were identified as dyslexic with all sources of information. Thus, the dyslexics in the present study had a history of language difficulties, provided a self-report of dyslexia and were identified as dyslexic with the battery of 12 tests. The identification process of the test battery was based on various discriminant and logistic regression analyses using both sum scores and single test items as predictors. The use of single test items as predictors was based on the idea that sometimes, only specific test items are representative of typical difficulties for dyslexics.

Table 1 provides group differences between dyslexics and non-dyslexics on all tests and questionnaires related to dyslexia. Non-dyslexics performed significantly better than dyslexics on all tests and questionnaires of dyslexia (most $p < 0.001$), except on sound deletion, although there was a significant difference in the larger original sample.

Table 1

Behavioral measures (Z-scores) (tests and questionnaires of dyslexia; factors of dyslexia; intelligence and school grades)

	Dyslexics (N=37)		Non-dyslexics (N=57)			
	M	SD	M	SD	T	*p*
Tests of Dyslexia						
Dutch dictation	-1.04	(0.95)	0.38	(0.64)	8.64	<0.001
English dictation	-0.91	(1.42)	0.27	(0.72)	5.30	<0.001
Missing letters	-0.58	(0.95)	0.40	(0.66)	5.85	<0.001
Pseudowords	-0.55	(1.07)	0.44	(0.64)	5.55	<0.001
Sound deletion	-0.30	(1.24)	0.08	(0.74)	1.85	0.067
Spoonerisms	-0.73	(1.12)	0.36	(0.75)	5.37	<0.001
Incorrect spelling test	-1.06	(1.22)	0.33	(0.88)	6.38	<0.001
Dutch–English rhyme words	-0.91	(1.30)	0.23	(0.68)	5.57	<0.001
Letters exchanging	-0.63	(1.03)	0.33	(0.81)	5.02	<0.001
Digit Span	-0.94	(1.15)	0.40	(0.88)	6.23	<0.001
Self-report questions						
Self-report reading	-1.16	(0.62)	0.48	(0.89)	9.76	<0.001
Self-report writing	-1.19	(0.71)	0.45	(0.79)	10.16	<0.001
Self-report speaking	-0.80	(0.84)	0.17	(0.98)	4.99	<0.001
Self-report listening	-0.46	(0.81)	0.18	(1.06)	3.08	0.003
Self-report copying	-0.96	(0.67)	0.36	(0.86)	7.89	<0.001
Self-report dictating	-0.98	(0.71)	0.43	(0.90)	8.05	<0.001
Self-report reading aloud	-0.95	(0.74)	0.35	(0.97)	6.94	<0.001
Factors of Dyslexia						
Spelling	-1.10	(0.81)	0.37	(0.59)	10.23	<0.001
Phonology	-0.51	(1.46)	0.12	(0.78)	2.74	0.007
Short-term memory	-0.60	(1.07)	0.23	(1.02)	3.78	<0.000
Rhyme/confusion	-0.53	(1.37)	0.12	(0.71)	3.02	0.003
Whole-word-processing/ complexity	-0.55	(0.97)	0.42	(0.77)	5.38	<0.001

	Dyslexics (N=37)		Non-dyslexics (N=57)			
	M	SD	M	SD	T	*p*
Intelligence						
Vocabulary	-0.52	(0.87)	0.14	(1.08)	3.07	0.003
Verbal Analogies	-0.47	(0.90)	0.18	(1.04)	3.14	0.002
Numeric Progressions	-0.38	(0.78)	0.03	(1.01)	2.10	0.038
Conclusions	-0.05	(0.90)	0.15	(1.03)	0.92	0.360
Speed of Calculation	-0.53	(0.62)	0.27	(1.16)	4.34	<0.001
Hidden Figures	-0.19	(0.90)	0.13	(1.06)	1.46	0.148
Raven Progressive Matrices	-0.26	(1.14)	0.06	(1.10)	1.31	0.195
Final Course Grades from School						
Dutch	6.45	(0.86)	7.11	(0.79)	3.56	0.001
English	6.47	(1.14)	7.09	(1.25)	2.35	0.021
Other Languages	6.58	(0.62)	7.13	(0.70)	3.61	0.001
Mathematics	6.52	(1.39)	6.53	(0.80)	0.06	0.949
Other Courses	6.97	(0.57)	7.06	(0.67)	0.65	0.515

Neuropsychological assessment:

Assessment of five factors of dyslexia

In a second previous study (Tamboer, Vorst & Oort, 2016) we used factor analysis on the whole dataset to identify cognitive measures of dyslexia. With principal component analyses, five independent factors of dyslexia could be distinguished. These factors explained 60% of the variance of the tests and questionnaires which were also used for the identification study. In a confirmatory factor analysis, we found a root-mean-square error of approximation (RMSEA, which is a measure of the error of approximation of the model-implied covariance matrix to the population covariance matrix and should be lower than 0.05) of 0.03 with p = 0.88 which means that the null hypothesis of a close fit cannot be rejected.

1. Factor spelling (15.7 % of variance)
High factor loadings for this factor were found for the tests *Dutch Dictation*, *English Dictation*, *Missing Letters* and *Incorrect Spelling* and for questions related to spelling.

2. Factor phonology (10.8 % of variance)
High factor loadings for this factor were found for the tests *Pseudowords* and *Sound Deletion* and for questions related to phonology.

3. Factor short-term memory (12.1 % of variance)
High factor loadings for this factor were found for the test *Short-Term Memory* and for questions related to memory.

4. Factor rhyme/confusion (11.3 % of variance)
High factor loadings for this factor were found for the test *Dutch–English Rhyme Words* and for questions related to phonological, visual, attentional and/or auditory confusion in language.

5. Factor whole-word processing/complexity (10.3 % of variance)
High factor loadings for this factor were found for the test *Letter Order* and for questions related to complexity.

Table 1 provides group differences between dyslexics and non-dyslexics for the five factors of dyslexia. The mean scores are presented as Z-scores which were acquired in the previous study from a total sample of 495 students. T-tests were performed for the groups of the present study. Non-dyslexics performed significantly better than dyslexics on all factors of dyslexia (all $p<0.01$). The difference on the factor spelling was much larger than the differences on the other factors, which implies that spelling difficulties are the most prominent difficulties for this group of dyslexics. Furthermore, we should note that dyslexics showed higher variances for all factors, especially on the phonology and rhyme/confusion factors.

Neuropsychological assessment: Assessment of intelligence

We assumed that the intelligence of all participants was within the normal range, because all were students who had finished the highest level of school education. Nevertheless, we investigated differences in intelligence between the 37 dyslexics and 57 controls with various tests of intelligence and with final course grades from school. These grades can vary between 1 and 10. A score of six or higher means that a student has passed the course. Compared to the non-dyslexic group, the dyslexic group had lower final school grades for *Dutch Language* (p=0.001), *English Language* (p=0.021) and *Other Languages* (p=0.001), but not for *Mathematics* and *Other Courses*. Six subtests of a cognitive battery based on the structure-of-intellect model of Guilford and *Raven's Progressive Matrices* were used for measures of intelligence. The mean performance of the whole group of about 1000 students on all tests was about one standard deviation above average compared to the normative standard in a total population. In this sample, dyslexics performed worse than non-dyslexics in four subtests of a cognitive battery based on the structure-of-intellect model of Guilford: *Vocabulary* (p=0.003), *Verbal Analogies* (p=0.002), *Speed of Calculation* (p<0.0005) and *Numeric Progressions* (p=0.038), but no differences were found on the subtests *Conclusions*, *Hidden Figures* and on *Raven's Progressive Matrices* (see also Table 1).

Voxel-based morphometry

We performed a voxel-based morphometry (VBM) analysis to find differences in GM volume in brain areas over subjects. For this, we acquired a structural scan for each of the subjects. From the first 45 subjects (15 dyslexics), we obtained one T1 recording per subject (3D T1, Turbo Field Echo, voxel size=1mm^3, field of view (FOV)=256^2, 160 slices, flip angle (FA)=8, echo time (TE)=3.78, repetition time (TR)=8.24), using a 3.0-T Philips Achieva scanner. From the last 49 subjects (22 dyslexics), we obtained three T1 recordings per subject (3D T1, Turbo Field Echo sequences, voxel size=1mm^3, FOV=256^2, 160 slices, FA=8, TE=3.81, TR=8.24), using a 3.0-T Philips Achieva scanner. We used the average image. After conducting t tests, we found no differences in head coils or noise between the two samples.

Data were analysed with FSL-VBM (Good et al., 2002), using FSL (Smith et al., 2004). First, structural images were brain-extracted. Next, tissue-type segmentation was carried out using FAST4 (Zhang, Brady, & Smith, 2001). The resulting GM and WM partial volume images were then aligned to MNI152 standard space using the affine registration. The resulting images were averaged to create a study-specific template, to which the native GM images were then non-linearly re-registered with a method that uses a B-spline representation of the registration warp field (Andersson, Jenkinson & Smith, 2007a, 2007b; Rueckert et al., 1999). The registered partial volume images were then modulated (to correct for local expansion or contraction) by dividing the Jacobian of the warp field. The modulated segmented images were then smoothed with an isotropic Gaussian kernel with a kernel of 4mm. The segmented WM and GM volumes were used to determine the total amount of GM and WM per subject.

Statistical analyses

Group differences in GM volume were calculated with permutation-based non-parametric testing (using a gap test) to find voxels that differed between subjects with and without dyslexia. The resulting clusters of differences in GM volume were corrected for multiple comparisons using random field theory with a cluster threshold of $t > 2.3$ and a reliability of $p < 0.05$ for extend of the cluster. Alternatively, we thresholded the resulting contrast looking for clusters of 200 connected voxels with a *p*-value lower than 0.05 in the VBM analysis (identical to Rouw & Scholte, 2010). The choice of 200 connected voxels is large, but choosing a smaller threshold would lead to tendencies with a decreasing reliability to be relevant. This second analysis was performed with two purposes. First, in the case of finding only a few or even no significant results, we wanted to determine clusters which could be considered tendencies and which could then be compared with previous findings. Second, we wanted to explore to what degree the GM volume in these clusters could be related to behavioural constructs.

The local demeaned GM volumes and the total GM and WM volumes were correlated (Pearson) in SPSS with five demeaned factor scores representing five cognitive aspects of dyslexia. These correlations were computed for all subjects and within groups, thus resulting in 15 comparisons for each cluster of GM. The correlations were corrected for multiple comparisons using the false discovery rate (FDR).

To account for the effects of age, gender and handedness, we recalculated group differences using an ANOVA analysis with these variables as fixed factors and recalculated the correlations partialling out these variables.

Results

Dyslexic versus non-dyslexic subjects

No differences were observed between dyslexics and non-dyslexics in total GM volume (0.66 vs. 0.67; T(92) = 1.24, p=0.22) and total WM volume (0.57 vs. 0.59; T(92) =1.18, p=0.24). Voxel-by-voxel GM volume comparisons revealed no significant differences in local GM volumes between dyslexics and non-dyslexics after correcting for multiple comparisons. Uncorrected clusters (p<0.05, cluster size k>200) are presented in Table 2. Three clusters of increased GM volume for dyslexics were found in the left posterior cerebellum (and a small part of the occipital fusiform gyrus), the left inferior parietal lobe (parts of angular and posterior supramarginal gyrus) and in the right superior temporal gyrus. Eight clusters of reduced GM volume for dyslexics were found in the left and right caudate nucleus, the right inferior temporal gyrus, the right angular gyrus, the left parietal operculum (insula), the right frontal lobe and in the left and right middle frontal gyrus.

Correlations between behavioural constructs and local gray matter volumes

Five factors of dyslexia were correlated (Pearson) with total GM, total WM and 11 clusters of local GM volumes. All correlations were calculated for all subjects and within groups and corrected for multiple comparisons (5× 13× 3=195 comparisons) using FDR, which resulted in four significant correlations. The relevant brain areas are presented in Figs. 1, 2 and 3. Scatterplots are presented in Figs. 4, 5, 6 and 7. A negative correlation (r=−0.34, p=0.034) for all subjects was observed between the factor spelling and GM volume in the left posterior cerebellum (and a small part of the occipital fusiform gyrus). This means that poor performances on spelling tasks correlated with increased GM volume in this area. Within the group of dyslexics, a negative correlation (r=−0.64, p=0.003) was observed between the factor rhyme/confusion and total WM volume. This means that poor performances on tasks related to rhyme/confusion correlated with increased total WM volume. Two positive correlations within the group of dyslexics were observed between the factor rhyme/confusion and GM volume in the left caudate nucleus (r=0.55, p=0.026) and in the right caudate nucleus (r=0.56, p=0.027). This means that poor performances on tasks related to confusion correlated with reduced GM volume in the left and right caudate nucleus. A few other correlations within the group of dyslexics were found but did not reach significance after FDR correction, while no uncorrected correlations larger than 0.3 were found within the group of non-dyslexics.

Table 2

Brain areas that represent tendencies (statistical trends) of GM alterations (uncorrected for multiple comparisons, not significant, $p<0.05$, $k>200$ voxels)

Region	MNI coordinates (centre of gravity)			Voxels
	x	y	z	
dyslexics > non-dyslexics				
L posterior cerebellum (occipital fusiform gyrus)	-32	-76	-23	749
L inferior parietal lobe (parts of angular and posterior supramarginal gyrus)	-57	-53	42	215
R superior temporal gyrus	50	-7	-12	306
non-dyslexics > dyslexics				
L caudate nucleus	-14	15	7	402
R caudate nucleus	10	14	8	377
R inferior temporal gyrus	58	-53	-21	434
R angular gyrus	53	-52	22	1516
L parietal operculum (insula)	-47	-27	23	261
R frontal lobe	20	39	40	1705
R middle frontal gyrus	38	47	-12	320
L middle frontal gyrus	-41	49	-8	538

Figure 1

Left posterior cerebellum (occipital fusiform gyrus)

X = -32 Y = -76 Z = -23

Coronel Horizontal Saggital

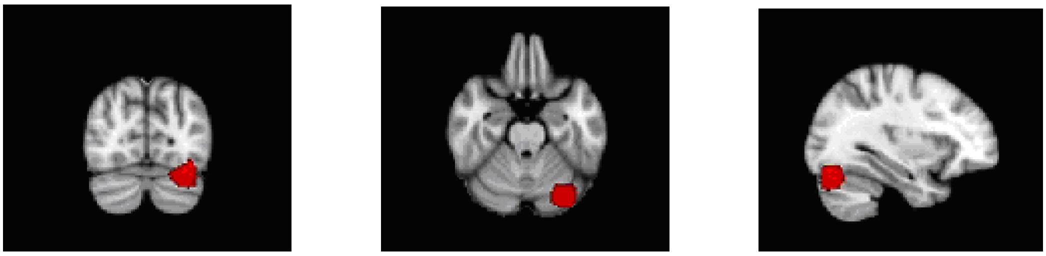


Figure 2

Left Caudate Nucleus

X = -14 Y = 15 Z = 7

Coronel Horizontal Saggital

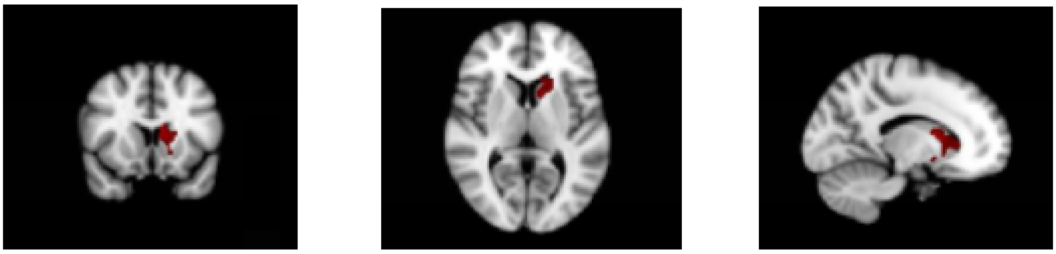


Figure 3

Right Caudate Nucleus

X = 10 Y = 14 Z = 8

Coronel Horizontal Saggital

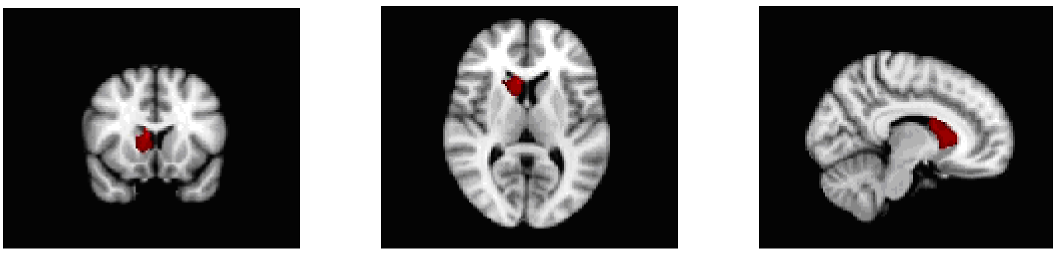


Figure 4

Correlation between Spelling and GM volume in the left posterior cerebellum

(all subjects: N = 94)

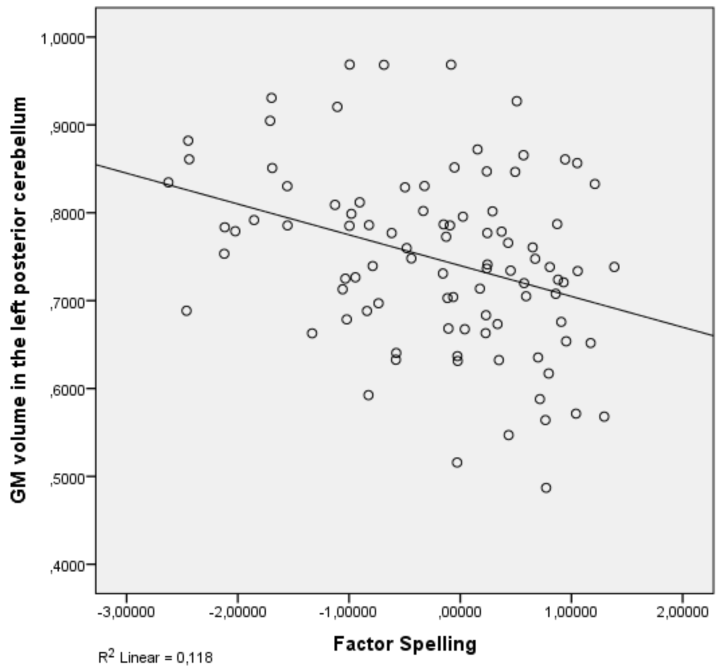


Figure 5

Correlation between Rhyme/confusion and total WM volume

(subjects with dyslexia: N = 37)

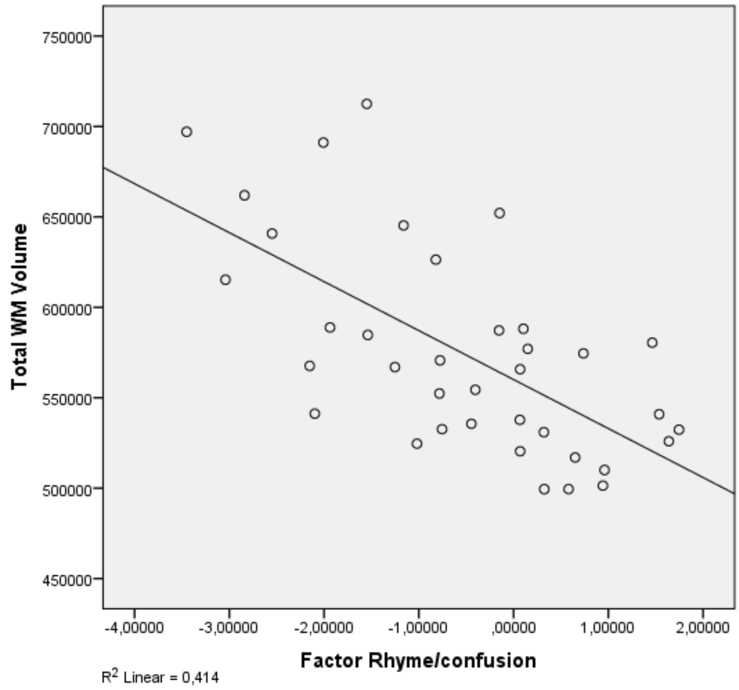


Figure 6

Correlation between Rhyme/confusion and the left caudate nucleus

(subjects with dyslexia: N = 37)

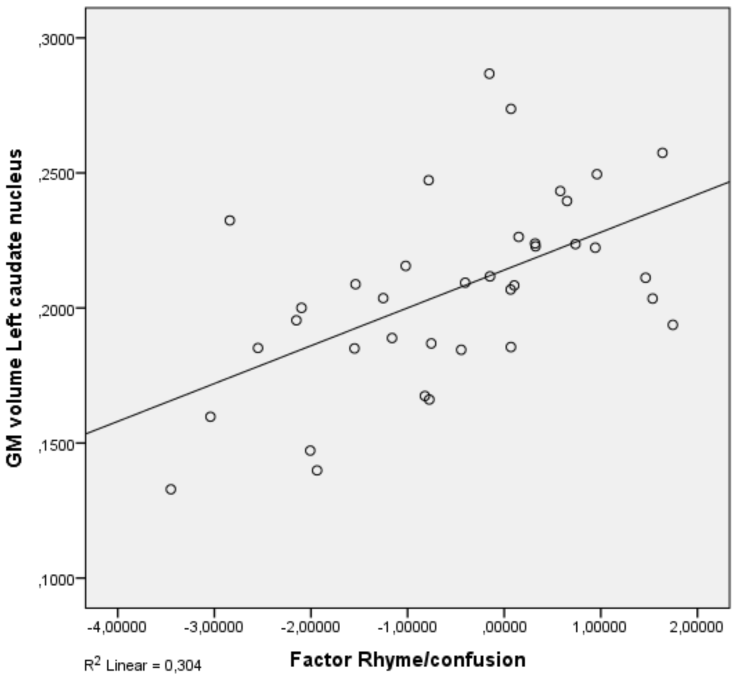
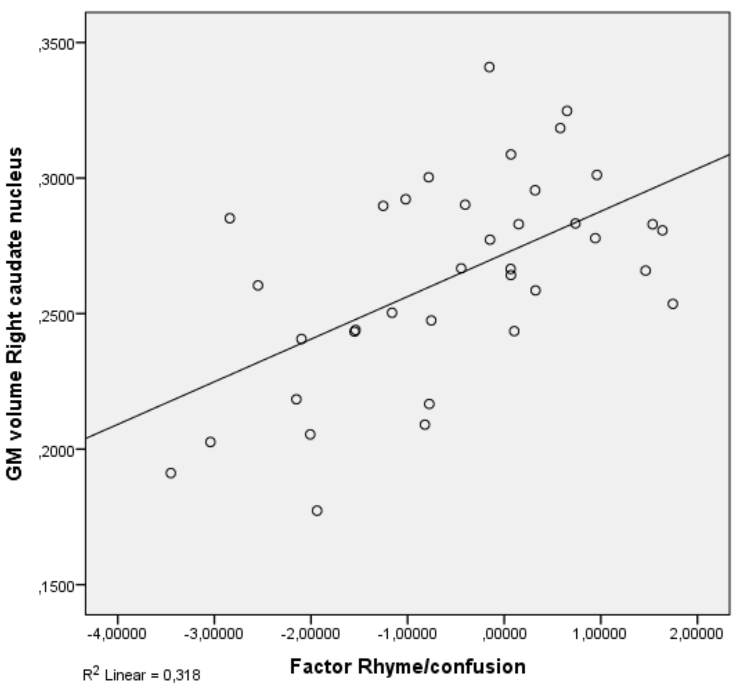


Figure 7

Correlation between Rhyme/confusion and the right caudate nucleus

(subjects with dyslexia: N = 37)



Age, gender and handedness

To account for the effects of age, gender and handedness, we performed various analyses. Chi-square tests revealed that the groups of dyslexics and non-dyslexics did not differ regarding proportion of men versus women and proportion of left- versus right-handedness. A *t*-test revealed no group differences for age. Two-way ANOVAs for group differences on total GM and WM volume revealed two main effects for gender with men having larger GM and WM volumes (both $p<0.001$) and a main effect of dyslexia when age was partialled out (with age explaining 2% of variance of WM and older students having increased WM volume) with dyslexics having reduced WM volume compared to non-dyslexics ($p=0.015$). However, we found no interaction effects between dyslexia and age, gender or handedness. Correlational analyses partialling out age, gender and handedness revealed the same significant correlations (FDR corrected) as in the analyses without covariates. A negative correlation ($r=-0.34$, $p=0.045$) for all subjects was observed between the factor spelling and GM volume in the left posterior cerebellum. Within the group of dyslexics, a negative correlation ($r=-0.55$, $p=0.049$) was observed between the factor rhyme/confusion and total WM volume. Two positive correlations within the group of dyslexics were observed between the factor rhyme/confusion and GM volume in the head of the left caudate nucleus ($r=0.59$, $p=0.026$) and in the head of the right caudate nucleus ($r=0.61$, $p=0.033$). No uncorrected correlations larger than 0.3 were found within the group of non-dyslexics.

Discussion

This study confirmed two main conclusions from previous studies. First, a large sample did not result in significant group differences between dyslexics and non-dyslexics in local GM volumes. Second, correlations between cognitive measures and local GM volumes provide significant findings. This study supported the view of dyslexia as a multiple cognitive deficit and related various cognitive measures to local alterations in GM volume. Areas of GM alterations were considered tendencies if not significant. However, most findings were in accordance with previous reports supporting the view that all anatomical brain findings in dyslexia contribute to the discussion in an accumulating way.

In summary, we found no significant differences in total GM volume, total WM volume and local GM volumes between the groups of dyslexics and non-dyslexics after corrections for multiple comparisons. However, dyslexics showed significant lower total WM volume than non-dyslexics when effects of age were partialled out. We found three areas of increased GM volume for dyslexics and eight areas of reduced GM volume for dyslexics. These areas were thresholded by 200 connected voxels with a p value lower than 0.05 and were considered tendencies. Most of these areas were reported in previous studies. In contrast with the absence of significant differences of local GM volumes, we found significant correlations between behavioural measures and alterations in GM volume, also when effects of age, gender and handedness were partialled out. An important finding was that no correlations were observed within the group of non-dyslexics, while three correlations were observed within the group of dyslexics and one in the whole group. We should also note that the present sample mainly consisted of females while in most previous studies, samples mostly consisted of males. The largest areas found in this study were observed in the right hemisphere consistent with a recent finding (Evans et al., 2014) that anatomical differences between female dyslexics and female non-dyslexics are more prominent in the right hemisphere compared to differences between male dyslexics and non-dyslexics.

Rhyme/confusion and the caudate nucleus

Within the group of dyslexics, a factor rhyme/confusion representing performances on the test ‘Dutch–English rhyme words’ and questions related to phonological, visual, attentional and/or auditory confusion in language correlated positively with GM volume in both the left and right caudate nucleus. Reductions of GM volume in the caudate nucleus have been reported only once in one of the first VBM studies of dyslexia (Brown et al., 2001) also finding these reductions bilaterally. Remarkably, hyperactivations of dyslexics in the left caudate were reported in an fMRI study, however without anatomical alterations (Hoeft et al., 2007). Numerous studies showed that the caudate nucleus is implicated in several executive control functions (Brovelli et al., 2011; Grahn, Parkinson & Owen, 2008; Lewis et al., 2004; Provost, Petrides & Monchi, 2010; White, 2009), language functions such as language switching in bilinguals (Luk et al., 2012), second language learning (Hosoda et al., 2013; Tan et al., 2011) and suppression of irrelevant words (Ali et al., 2010). In relation to these findings, the relation between the caudate nucleus and the test *Dutch–English Rhyme Words*, which requires fast switching between languages, makes perfect sense. This is also the case with the confusion reported via self-report questions, which represented typical mistakes, such as exchanging letters within words and exchanging words within sentences – mistakes that might result from impaired cognitive control of attention. We think that the possibility exists that reduced GM volume in the caudate nucleus in relation to rhyme/confusion might represent a more fundamental dysfunc-tion of dyslexic people, which might encompass various difficulties of confusion, such as exchanging letters or words. A recent study by Hachmann et al. (2014) seems to support this. These researchers found that impaired performances on tasks of short-term memory are explained by impaired serial order processing rather than impaired storage.

Confusion, spelling and white matter density

Dyslexics as a group had reduced WM volume compared to non-dyslexics when age was partialled out. In contrast, within the group of dyslexics, rhyme/confusion correlated negatively with total WM volume, meaning that the dyslexics who are more severely impaired regarding rhyme/confusion have larger total WM volume. Apparently, different behavioural constructs have different effects on WM volume. It is generally assumed that WM volume represents connectivity in the brain. Regarding dyslexia, it has been hypothesised that dyslexics suffer from impaired connectivity (e.g. Steinbrink et al., 2008). Based on the results of the present study, we alternatively hypothesise that confusion may result from too much connectivity in some areas. Connection efficiency has also been investigated using diffusion tensor imaging (DTI), which quantifies the relative diffusivity of water in a voxel into directional components. However, while the relation between WM volume and the so called fractional anisotropy (FA) remains somewhat unclear, a meta-analysis of DTI studies (Vandermosten et al., 2012) only resulted in reduced FA values (mainly in a left temporoparietal region which hosts two WM tracts: the left arcuate fasciculus and the left corona radiate). But, higher FA values were reported in the splenium, the posterior end of the corpus callosum which connects the left and right cerebral hemispheres (Frye et al., 2008; Odegard et al., 2009). This might be viewed as support for the idea that confusion correlates with too much connectivity. But actually, the main thing which is supported by all these results regarding WM volume alterations is the complex nature of dyslexia. This is emphasised even more, for instance, by theories of increased WM gyral depth in the brains of dyslexics (Casanova et al., 2010). The idea is that reduced WM volume is the result of broader gyri or any other change in the thickness of the cortex, involution of sulci and/or complexity of cortical folding.

Spelling and the cerebellum

Better performances on spelling tasks correlated with reduced GM volume in the left posterior cerebellum (and a small part of the left occipital fusiform gyrus) in the whole group of students. In the meta-analysis by Richlan, Kronbichler & Wimmer (2013), cerebellar abnormalities did not survive significant thresholds. In the meta-analysis by Linkersdörfer et al. (2012), reduced GM volumes were found bilaterally in the cerebellum, although located more anterior than the area of increased GM volume in the cerebellum in this study. Despite somewhat different coordinates, this seems to be in contrast with each other. However, in a study by Pernet et al. (2009b), using a classification approach, the right cerebellar declive was one of the two best predictors of dyslexia, with dyslexics falling either above or below the control group's 95% confidence interval boundaries. Remarkably, our cluster of increased GM volume in the left cerebellum was found more or less on the opposite site of the cluster found by Pernet et al. In the study by Jednoróg et al. (2014), increased GM volume for one subtype of dyslexics was reported in the left cerebellum/lingual gyrus, while in the same area, reduced GM volume was reported for another subtype of dyslexics. It becomes even more puzzling when we compare these findings with findings of increased symmetry in dyslexics as opposed to non-dyslexics showing more right GM than left GM (Rae et al., 2002) or with findings of differences in asymmetry between dyslexics with and without a phonological deficit (Leonard et al., 2001). One alternative explanation for inconsistent findings in the cerebellum might be that the cerebellum can be difficult to segment because of volume averaging in the folia in comparison to cerebral cortex.

Another explanation can be derived from the following. Generally, spelling is one of the most commonly reported symptoms of dyslexia. However, in schools, poor-performing children also receive extra training when they are not dyslexic. This might explain why spelling correlates with the cerebellum across groups. The cerebellum is associated with skill acquisition and automatisation and specifically with aspects of language processing (Hodge et al., 2010; Murdoch, 2010). In dyslexia, impaired functioning of the cerebellum is associated with impaired reading fluency and motor deficits (Nicholson & Fawcett, 2007). These findings seem to support increased GM volume in the cerebellum from training in spelling abilities, instead of reductions in GM volume. A strong argument in favour of these learning effects related to dyslexia is that cerebellar findings seem to depend on the age of the subjects. For example, a VBM study of pre-reading dyslexic children did not report alterations in cerebellar areas (Raschle, Chang & Gaab, 2011), while a VBM study of 11 dyslexic school children reported increased GM volume in the right anterior cerebellum after an 8-week training focused on mental imagery; articulation; and tracing of letters, groups of letters and words (Krafnick et al., 2011).

Frontal and temporoparietal areas

We observed five areas of GM alterations in temporo-parietal areas and three in frontal areas. Generally, dyslexia (especially in relation to phonological impairments) has been linked with atypical activation of the left perisylvian fronto-temporo-parietal network (e.g. Richlan, Kronbichler & Wimmer, 2011). However, in the meta-analysis by Richlan, Kronbichler & Wimmer (2012), reduced GM volumes were observed in both hemispheres: one in the left superior temporal sulcus and unexpectedly one in the right superior temporal gyrus. In the present study, all temporo-parietal and frontal GM abnormalities failed to survive corrections for multiple comparisons. Our areas in the left inferior parietal lobe extending to the supramarginal gyrus (increased GM volume for dyslexics) and in the right angular gyrus (reduced GM volume for dyslexics) are close to areas of reduced GM volume reported in the meta-analysis by Linkersdörfer et al. (2012). Six other areas were observed in parietal, temporal and frontal areas, regions close to or overlapping with areas which were reported before, either in anatomical or in functional studies.

A possibility is that unbalanced inclusion of different subtypes of dyslexia might have enhanced the finding of significant and inconsistent results in these areas in individual studies. In other words, when dyslexics exhibit different cognitive impairments, it can be expected that highly educated students apply different alternative compensation techniques leading to various clusters of augmentations or reductions. Thus, some dyslexics might try to improve their phonological abilities and others their reading abilities. This view was confirmed in a study by Peyrin et al. (2012) who observed various functional differences in both hemispheres between a young dyslexic adult with only phonological impairments as opposed to a young dyslexic adult with only an impairment of visual attention span. Another explanation for inconsistent findings in the perisylvian fronto-temporo-parietal network may be gender effects as reported by Evans et al. (2014). They observed typical left and right hemispheric alterations in men, but in women mainly right hemispheric alterations, which seems to be consistent with our observations. We only found three small areas in the left hemisphere, but three small and two large areas in the left hemisphere. We should include these areas in discussions as being relevant tendencies which require further exploration.

Limitations of this study

This study confirmed that the complex nature of dyslexia cannot easily be clarified by anatomical brain correlates. Although findings of this study contribute to the accumulating knowledge about brain correlates of dyslexia, we should also emphasise some limitations.

Although we found significant correlations, we found no significant group differences after corrections for multiple comparisons. Instead, we reported large tendencies and looked whether these tendencies correlated with behavioural measures. These tendencies were defined by clusters of 200 connected voxels with a *p*-value lower than 0.05 in the VBM analysis, which is, of course, an arbitrary decision. We referred to another study which used the same threshold (Rouw & Scholte, 2010). This is a relative large threshold. A disadvantage is that small and relevant clusters may be overlooked. However, we wanted to study large tendencies without running the risk of analyzing small clusters that result from noise.

Another limitation of this study is related to the sample, which consisted of students. However, we found that using a student sample might also be an advantage. For instance, students received extensive language training at school (students with as well as students without dyslexia). This probably was related to the significant correlation between spelling abilities and reduced GM volume in the cerebellum. We argued that also other findings of the present study might be related to different compensation strategies which can assumed to be characteristic for highly intelligent students. However, as a result of this, this study could not separate brain correlates of dyslexia that result from training from brain correlates that may be present at birth.

Conclusion

We found no significant group differences in local GM volumes between dyslexics and non-dyslexics although we used a large sample that accounted for different cognitive profiles of dyslexics. Instead, we found four significant correlations between five behavioural measures of dyslexia and local GM and total GM and WM volumes. These measures specify various specific relations with local GM volume alterations. Specifically, we found that the caudate nucleus is involved in abilities related to confusion, that the cerebellum is involved in abilities related to spelling and that both spelling and confusion are related to total WM volume. These results reveal that understanding of anatomical alterations in dyslexia is best identified when various cognitive aspects of dyslexia are acknowledged. Other findings of this study were more difficult to interpret, such as the involvement of temporo-parietal areas. Effects of sample differences cannot be ruled out, such as gender differences, age differences, differences in selection methods, differences in education and differences in experience and compensation strategies. Nevertheless, also insignificant findings might contribute across studies to accumulate evidence of brain alterations in dyslexia.

7

Machine Learning and Dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia

Meta-analytic studies suggest that dyslexia is characterized by subtle and spatially distributed variations in brain anatomy, although many variations failed to be significant after corrections of multiple comparisons. To circumvent issues of significance which are characteristic for conventional analysis techniques, we applied a machine learning technique – support vector machine – to differentiate between subjects with and without dyslexia.

In a sample of 22 students with dyslexia (20 women) and 27 students without dyslexia (25 women) (18– 21 years), a classification performance of 80% ($p < 0.001$; d-prime = 1.67) was achieved on the basis of differences in grey matter (sensitivity 82%, specificity 78%). The voxels that were most reliable for classification were found in the left occipital fusiform gyrus (LOFG), in the right occipital fusiform gyrus (ROFG), and in the left inferior parietal lobule (LIPL). Additionally, we found that classification certainty (e.g. the percentage of times a subject was correctly classified) correlated with severity of dyslexia (r = 0.47). Furthermore, various significant correlations were found between the three anatomical regions and behavioural measures of spelling, phonology and whole-word-reading. No correlations were found with behavioural measures of short-term memory and visual/attentional confusion. These data indicate that the LOFG, ROFG and the LIPL are neuro-endophenotype and potentially biomarkers for types of dyslexia related to reading, spelling and phonology.

In a second and independent sample of 876 young adults of a general population, the trained classifier of the first sample was tested, resulting in a classification performance of 59% (p = 0.07; d-prime = 0.65). This decline in classification performance resulted from a large percentage of false alarms.

Keywords: Dyslexia, MRI, SVM classification, gray matter, VWFA

Based on: Tamboer, P., Vorst, H. C. M., Ghebreab, S., & Scholte, H. S. (2016). Machine learning and dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia. *NeuroImage: Clinical*, *11*, 508-514.

Introduction

Dyslexia is usually defined as a specific reading disorder characterized by a specific and significant impairment in the development of reading skills that are unrelated to problems with visual acuity, schooling or overall mental development (World Health Organisation, 2010). For (sub-) groups of dyslexics, reading difficulties have been related to various symptoms of which the most frequently reported are related to phonological difficulties although, in recent years, visual/attentional deficits are reported frequently as well (e.g. Ramus and Ahissar, 2012). Generally, it is assumed that early learning delays cannot be overcome completely despite remedial teaching programs, and that these learning delays interfere with academic achievement into adulthood for most of the dyslexics which are estimated to represent 5% to 15% of the population.

Reliable diagnoses can currently only be determined behaviourally and after some years of education, when the discrepancy between nor-mal cognitive and reading abilities becomes visible. Alternatively, researchers have been searching for biomarkers of dyslexia using MRI or fMRI. Meta-analyses showed that these differences do exist (Richlan, Kronbichler & Wimmer, 2011, 2013; Vandermosten et al., 2012), although many findings failed to be significant after corrections of multiple comparisons.

A potentially more powerful technique than the univariate voxel-wise evaluation and correction of multiple comparisons are multivariate classification techniques from machine learning. This technique has recently successfully been applied in several clinical neuroimaging studies. For instance, a high accuracy rate of 90% has been reported for discriminating major depressive disorder and controls (Mwangi et al., 2012). An accuracy rate of 81% has been found for autism (Ecker et al., 2010).

Also with regard to dyslexia, this classification approach has been applied. In a study of Hoeft et al. (2011), a multivariate pattern analysis of brain activation during a reading task over the whole brain using linear support vector machine and cross-validation, showed that reading gains over a 2.5 year period in children with dyslexia can be predicted with N90% classification accuracy. A study of Tanaka et al. (2011) showed that in two samples of typical and poor reading children 79% and 80% were classified correctly using leave-one-out linear SVM analyses of brain activation during phonological processing. In a study of Pernet et al. (2009b), classification of dyslexic readers brains resulted in dyslexics falling outside the 95%

confidence boundaries of the controls in two areas (the right cerebellar declive and the right lentiform nucleus).

The aim of this study is to investigate whether young adults with and without dyslexia can reliably be classified based on anatomical differences. We examined neuro-anatomical networks involved in dyslexia using a whole-brain classification employing SVM and cross-validation. We used the T_1-weighted magnetic resonance images of GM structure of a sample of 22 students with dyslexia and 27 students without dyslexia for acquiring a trained classifier. Next, we determined which voxels were involved with the correct classification. Furthermore, we explored to what degree these results can be used to investigate the relation between different cognitive aspects of dyslexia and neural substrates. We also tested the reliability of the trained classifier in an independent sample of 876 young adults.

Method

Subjects & procedure

The first sample – used to find a trained classifier – consisted of 22 students with dyslexia (20 women; 4 left-handed; mean age 20.7 years, SD 1.8 years) and 27 students without dyslexia (25 women; 4 left-handed; mean age 20.3 years, SD 0.9 years). All participating subjects were first-year psychology students, native Dutch speakers, had at least twelve years of school education, were free from medical or psychiatric diseases and had no history of sensory deficits or head trauma. None of the participants had a diagnosis of ADHD. Handedness was assessed with a short self-report questionnaire, which included questions about writing hand, general hand preference, as well as 20 specific questions. There were no students with inconsistent reports which could indicate being ambidextrous.

The 49 students of the first sample were invited to participate in the present study by mail and telephone. The students gave informed written consent and were debriefed afterwards. All participants had the option to choose between acquiring participation points required for the first year of study, or a financial reward. This study was approved by the ethics committee at the University of Amsterdam.

The second sample – used to test the trained classifier – consisted of young adults of a general population. Brain data of this sample were available for various studies at the University of Amsterdam. We excluded participants with a serious medical condition, with a diagnosis of autism spectrum disorder and participants using psychiatric drugs or psychiatric medicine. The remaining sample consisted of 876 subjects who were native Dutch speakers and who had at least twelve years of school education. Of this sample, 60 (7%) subjects (27 women; mean age 22.5 years, SD 1.6 years) were diagnosed with dyslexia whilst at-tending school, and 816 subjects (433 women; mean age 22.9 years, SD 1.7 years) had no reported history of dyslexia.

Neuropsychological assessment

The first sample was acquired from a sample of 480 students who participated in a previous study (Tamboer, Vorst & Oort, 2014). In that study, dyslexia and non-dyslexia was assessed using three sources of information: (1) a history of language difficulties, (2) a self-report of language difficulties, and (3) a test-battery measuring numerous abilities such as spelling, reading, pseudoword reading, phonology, attention, and short-term memory. Severity of dyslexia was determined with a regression formula which consisted of 13 test items and 10 self-report questions, and which classified all subjects with and without dyslexia correctly. In a follow-up study (Tamboer, Vorst & Oort, 2016), five behavioural factors accompanying dyslexia were determined using exploratory and confirmatory factor analyses. On the basis of these analyses we acquired five Z-transformed sum scores: spelling, phonology, short-term memory, visual/attentional confusion, and whole-word reading.

We assumed that intelligence of all participants was within the nor-mal range because all had finished the highest level of secondary school education in the Netherlands. Group differences of intelligence were analysed as follows. In the original sample of 480 students, we performed factor analyses over six subtests of a cognitive battery that was based on the Structure of Intellect Model of Guilford and *Raven Progressive Matrices* for a better interpretation of various aspects of intelligence. Three factors (non-verbal intelligence, speed of numeric processing, vocabulary) were extracted and factor scores were acquired with a mean of zero and standard deviation of one. The smaller sample of the present study shows small deviations from the mean and SD of 1 because this was a selection of the original sample. In the present sample, the groups did not differ on the three aspects of general intelligence. Furthermore, no differences were found on school grades of English language, mathematics, and other courses. However, the dyslexic group had compared to the non-dyslexic group lower final school grades of Dutch language and other languages such as

French or German. We conclude that the groups did not differ in terms of general intelligence. Specific details can be found in the Appendix.

The data of the second sample were collected to be used in various studies regarding brain correlates accompanying various developmental disorders. The subjects of this sample were not tested for dyslexia, because the present study was performed after the collection of data. Available was a large self-report questionnaire which included two questions about dyslexia. One question was whether the subjects had an official certificate of dyslexia and a second question was whether a subject was tested for dyslexia whilst attending school.

Image acquisition and pre-processing

For both samples, we used the standard population acquisition protocol of the Spinoza Centre for NeuroImaging in Amsterdam. We acquired three 3DT1 whole-brain scans for each subject (3D T1, Turbo Field Echo sequences, voxel size =1 mm^3, FOV = 256^2mm, 160 slices, FA = 8°, TE = 3.81ms, TR = 8.24ms), using a 3T Philips Achieva scanner with a 32 channel headcoil. Each sequence lasted approximately 6 min to acquire. The three T1 scans were aligned to the 2nd recorded T1 scan and subsequently averaged. Each averaged brain was manually inspected and subsequently placed in a common space using VBM (Good et al., 2001) as implemented in FSL (Smith et al., 2004).

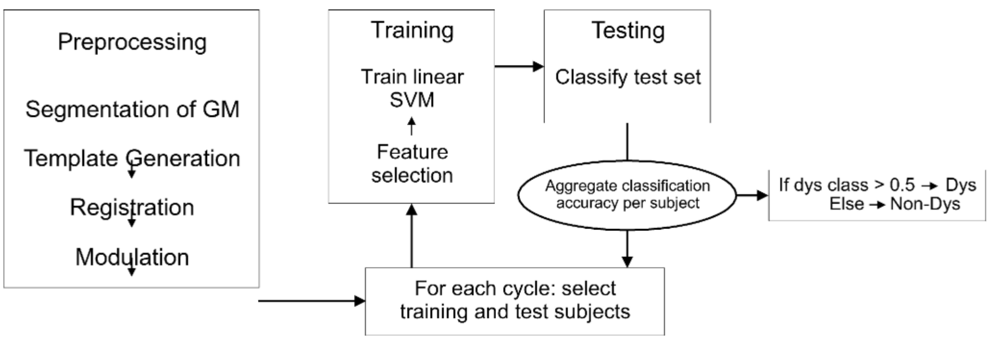
First, structural images were brain-extracted. Next, tissue-type seg-mentation was carried out using FAST4 (Zhang, Brady & Smith, 2001). The resulting GM partial volume images were then aligned to MNI152 standard space using the affine registration. The resulting images were aver-aged to create a study-specific template, to which the original GM images were then non-linearly re-registered with a method that uses a B-spline representation of the registration warp field (Rueckert et al., 1999). The registered partial volume images were then modulated (to correct for local expansion or contraction) by dividing by the Jacobian of the warp field. The modulated segmented images were then smoothed with an isotropic Gaussian kernel with a kernel of 4mm.

Pattern classification

We used SVM to train a classifier to distinguish between subjects with and without dyslexia of the first sample (http://www.csie.ntu.edu.tw/~cjlin/libsvm/). The SVM classifier was trained on using 21 randomly selected subjects with dyslexia (of 22) and 21 randomly selected subjects without dyslexia (of 27). The voxels used during the training stage were determined by subtracting the average VBM transformed brain of the selected controls from the average VBM transformed brain of the dyslexia group and Z-transforming the resulting difference image. Next, we estimated the linear hyperplane that maximally separates the subjects with and without dyslexia using those voxels that surpassed a Z-threshold of 3, 3.5, 4 and 4.5. We used the default values for training a linear classifier, one-class classifier. This procedure yielded four classifiers which performance was evaluated using the one remaining dyslexic subject and one of the (randomly selected) six controls and we repeated this procedure 10,000 times. Thus, we applied a cross-validation procedure in which each subject was classified many times on the basis of a hyperplane acquired with 21 subjects with and 21 subjects without dyslexia. The Z-threshold of 4 and 4.5 yielded the highest classification performance and we subsequently also tested a threshold value of 4.25. This Z-value controls how many voxels are used for classification and resulted in these conditions into a value of 759 voxels. The procedure using the Z-threshold of 4.25 yielded the highest classification performance and only these results were used for subsequent analysis. This result yielded many classifications of each subject as either dyslexic or control that could either be correct or incorrect. See Figure 1 for the classification scheme.

Figure 1

Classification Scheme



Analyses

For each subject, group membership was determined by the majority of hits or misses. Thus, when the proportion of hits for a particular subject was ≥0.5, the subject was considered correctly classified. This also yielded, per subject, a continuous score representing the classification accuracy for each subject. The prediction accuracy of the trained classifier was estimated with the proportion of correctly classified subjects. We quantified the estimated sensitivity with the proportion of correctly classified subjects with dyslexia, and the estimated specificity with the proportion correctly classified subjects without dyslexia. We calculated d-prime with Z(proportion hits) – Z(proportion false alarms).

In a second analysis, we repeated the classification procedure described in the previous paragraph but using 20% of the selected voxels per iteration. For each voxel, we subsequently scored the percentage of times a classification was successful for that drawing of voxels. This yielded after 100,000 iterations, for each voxel, a count for the number of times that a voxel was selected and the number of times that this selection resulted in a correct classification. In this way we expressed the importance of a voxel for classifying dyslexia.

Next, we examined the relation between brain regions of the voxels that were most reliable for classification and the behavioural measures (spelling, phonology, short-term memory, whole-word reading and visual/attentional confusion) with bivariate Pearson correlations. We also calculated the Pearson correlation between the classification accuracy for each subject and severity of dyslexia.

In the second sample, group membership of each subject was deter-mined by the trained classifier which was acquired with the first sample. The accuracy of the trained classifier in this sample was estimated with the proportion of correctly classified subjects. We quantified the estimated sensitivity with the proportion of correctly classified subjects with dyslexia, and the estimated specificity with the proportion correctly classified subjects without dyslexia. We calculated d-prime with Z(proportion hits) – Z(proportion false alarms).

Results

Classification of subjects with and without dyslexia (first sample)

The SVM technique resulted in a total prediction accuracy of 39 / 49 = 80%. Permutation testing, in which we repeated this entire procedure but now with permuted labels, revealed that in 1000 simulations this level of accuracy was never reached yielding a significance of $p < 0.001$. Furthermore, we found a sensitivity (proportion correctly classified subjects with dyslexia) of 18 / 22 = 82%, and a specificity (proportion correctly classified subjects without dyslexia) of 21 / 27 = 78%. We found that d-prime = 1.67. See Table 1. Positive predictive value (proportion of all subjects classified as dyslexic who have in fact dyslexia) was 75% and negative predictive value (proportion of all subjects classified as not dyslexic who have in fact no dyslexia) was 84%.

Table 1

Classification of subjects with and without dyslexia

Sample 1 (N = 49)	Predicted as	
	Dyslexic	Not dyslexic
Subjects with dyslexia	18	4
Subjects without dyslexia	6	21
Prediction accuracy	80% (p < 0.001)	
d-prime	1.67	
Sensitivity	82%	
Specificity	78%	
Positive predictive value	75%	
Negative predictive value	84%	

Sample 2 (N = 876)	Predicted as	
	Dyslexic (43%)	Not dyslexic (57%)
Subjects with dyslexia (7%)	40	20
Subjects without dyslexia (93%)	338	478
Prediction accuracy	59% (p = 0.07)	
d-prime	0.65	
Sensitivity	67%	
Specificity	59%	
Positive predictive value	11%	
Negative predictive value	96%	

Mean classification accuracy (first sample)

For each subject separately, the accuracy of the prediction was calculated with the proportion hits or – for subjects incorrectly classified – misses. This resulted in a continuous score representing the classification accuracy for each subject, ranging from 0.51 to 1.00. In the whole sample the mean of this classification accuracy was 0.89. We also calculated the mean classification accuracy for four subgroups. The correctly classified dyslexics had a mean classification accuracy of 0.87 (SD 0.13). The false negatives had a mean classification accuracy of 0.87 (SD 0.15). The correctly classified non-dyslexics had a mean classification accuracy of 0.90 (SD 0.16). The false positives had a mean classification accuracy of 0.94 (SD 0.12). One conclusion was that the correctly predicted subjects were correctly classified in a large majority of trials. A second conclusion was that the false negatives and the false positives were correctly classified as being false negatives and false positives in a large majority of trials as well. In other words, the subjects who were incorrectly classified were consistently incorrectly classified in a large majority of trials, while more inconsistency over trials could have been expected. Simple explanations can be ruled out because we found that left-handedness, gender, and age had no influence on prediction accuracy. We also found no differences on factors of intelligence, school grades, factors of dyslexia, and severity of dyslexia between the groups of correctly and incorrectly classified subjects with dyslexia and between the groups of correctly and incorrectly classified subjects without dyslexia.

Anatomical classifier (first sample)

Figure 2 shows three brain regions of GM (averaged between trials), which discriminated between subjects with and without dyslexia. One cluster of reduced GM volume for subjects with dyslexia was found in the LIPL (65 voxels; −53, −28, 24). Two clusters of augmented GM volume for subjects with dyslexia were found bilateral in the LOFG (150 voxels; −35, −72, −21), and in the ROFG (187 voxels; 35, 67, −19). Table 2 presents the coordinates of the clusters and the direction of the differences between the groups of subjects with and without dyslexia on GM volume of the three clusters. These differences are statistically not relevant, but evaluate the relative contribution of the separate regions to the overall classification of subjects with and without dyslexia.

Figure 2

Regions with voxels involved in discriminating between subjects with and without dyslexia. Post hoc analysis revealed that the region in blue (LIPL) is smaller in subjects with dyslexia and that the regions in red (LOFG and ROFG) are larger in subjects with dyslexia.

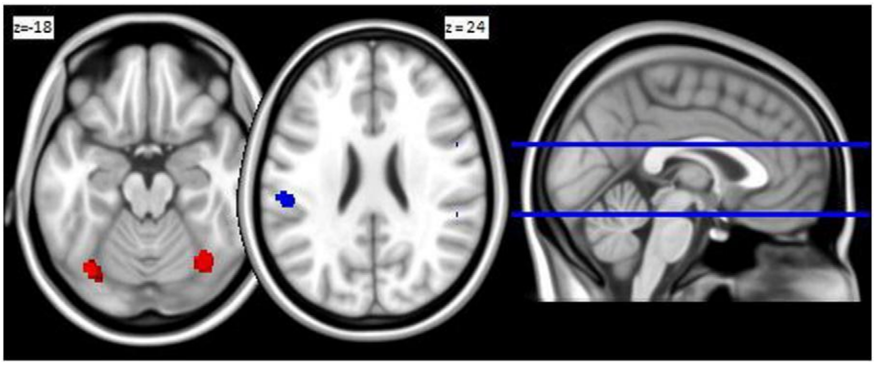


Table 2

GM regions discriminating between subjects with and without dyslexia

Anatomical Region	MNI- coordinates (x, y, z)			Cluster Size (mm^3)	N voxels	Direction of effect
Left occipital fusiform gyrus (visual word form area)	-35	-72	-21	1200	150	Dyslexic > Control (p = 0.02)
Right occipital fusiform gyrus	35	-67	-19	1496	187	Dyslexic > Control (p = 0.01)
Left inferior parietal lobule	-53	-28	24	520	65	Control > Dyslexic (p = 0.21)

Correlation MCA – severity of dyslexia (first sample)

We linearly transformed the mean classification accuracy of each subject to a continuous score that represented severity of dyslexia according to the classifier, ranging from 0 (no dyslexia) to 1 (severe dyslexia). This score was correlated with a behavioural representation of severity of dyslexia (e.g. a regression score). We found a correlation of r = 0.47 (p = 0.0007).

Correlations between GM indices and symptoms of dyslexia (first sample)

Table 3 summarizes correlations between GM volumes of the three clusters with measures and severity of dyslexia. There are four main findings. First, severity of dyslexia correlates significantly with GM volume in the ROFG and LOFG, meaning that dyslexics have higher GM volume in those areas. Second, spelling (good performance) correlates significantly negative with GM volume in the LOFG and significantly positive with GM volume in the LIPL. Third, whole-word-reading (good performance) correlates significantly negative with GM volume in the LOFG. Fourth, phonology (good performance) correlates significantly positive with GM volume in the LIPL. No significant correlations were found for short-term memory and visual/attentional confusion. Scatter plots of the significant correlations are presented in Supplementary Information.

Table 3

Sample 1: *Pearson correlations (whole group: N = 49)*
of brain indices with measures and severity of dyslexia
(Severity of dyslexia ~ low score)

	Right Occipital Fusiform Gyrus	Left Occipital Fusiform Gyrus	Left Inferior Parietal Lobule
Spelling	-0.267	**-0.311 (*p* = 0.029)**	**0.322 (*p* = 0.024)**
Phonology	-0.242	-0.134	**0.301 (*p* = 0.036)**
Short-term memory	-0.177	-0.215	0.222
Visual/attentional Confusion	-0.152	-0.137	0.128
Whole-word reading	-0.239	**-0.318 (*p* = 0.026)**	0.208
Severity of Dyslexia	**-0.377 (*p* = 0.008)**	**-0.337 (*p* = 0.018)**	0.224

Classification of subjects of the second sample

The anatomical classifier of the first sample resulted in a total prediction accuracy of 59% in the second sample. Permutation testing, in which we shuffled the labels from Sample 1, revealed that the classification is lower than the observed 59% in 93% of the cases yielding a significance of $p = 0.07$. Furthermore, we found sensitivity (proportion correctly classified subjects with dyslexia) of 67%, and specificity (proportion correctly classified subjects without dyslexia) of 59%. We found that d-prime = 0.65. See Table 1. In this sample, dyslexia was overpredicted with a total number of predicted subjects with dyslexia of 43%. Positive predictive value (proportion of all subjects classified as dyslexic who have in fact dyslexia) was 11% and negative predictive value (proportion of all subjects classified as not dyslexic who have in fact no dyslexia) was 96%. We also found that total prediction accuracy did not improve when selecting the same age range as in the first sample. Neither did the prediction accuracy improve after selecting only males or females. In both cases, prediction accuracy was 60%.

Discussion

Overview of main results

With SVM, a trained anatomical classifier correctly classified 80% of students with and without dyslexia (82% of students with dyslexia; 78% of students without dyslexia; d-prime = 1.67). Regions that were important in discriminating between these groups were the LOFG and the ROFG and the LIPL. Severity of dyslexia was defined with mean classification accuracy and correlated positively with severity of dyslexia ac-cording to behavioural measures ($r = 0.47$). We found six significant correlations between the three regions and behavioural measures of dyslexia. In an independent sample of a general population, the anatomical trainer of the first sample correctly classified 59% of young adults with and without dyslexia (67% of students with dyslexia; 59% of students without dyslexia; d-prime = 0.65).

Evaluation of prediction accuracy

In a well-balanced sample of students with and without dyslexia, a majority was correctly classified using SVM. In a general population sample, the trained classifier of the first sample resulted in a much lower classification performance, but still above chance. The advantage of using this classification approach over traditional analyses of group differences is that discussions about statistical corrections for multiple comparisons are not relevant. The statistical significance of prediction accuracy of the anatomical classifier was supported by a cross-validation approach and by a low *p*-value (< 0.001). The reliability of the classifier was supported by the finding that on average, subjects were classified with high consistency between trials, which was expressed by a mean classification accuracy of 0.89. The validity of the classifier was supported by the finding that this mean classification ac-curacy correlated positively with a measure of behavioural severity of dyslexia. The reliability of the anatomical classifier was further confirmed in a second sample, although prediction accuracy and calculated d-prime were much lower in the second sample than in the first sample. This decline resulted mainly from many false alarms.

From a diagnostic point of view, predictive values are useful measures. These measures are, however, sample specific. In the first sample, positive and negative predictive value are high but meaningless because the equal groups of the first sample do not represent reality with a prevalence of dyslexia of 5–15%. However, we still can draw meaningful conclusions. We can imagine a sample with a more representative fraction of students with dyslexia, for instance, with the number of students without dyslexia being ten times higher than the 27 students in the first sample. Classification performance would then approach specificity, which is 78%, only 2% below the overall classification performance of 80%. However, positive and negative predictive value would change. Negative predictive value would then be 98%, meaning that a prediction of no dyslexia by the classifier is in most cases correct. In contrast, positive predictive value would be 23%, meaning that in about one of four cases a prediction of dyslexia is correct, but incorrect in three out of four cases. How to explain this large number of false alarms? Analyses of mean classification accuracy revealed that not only the correctly classified students but also false positives and false negatives were consistently classified by the trainer in most of the cases (90%). Apparently, the anatomical features represented by the classifier may represent something else than dyslexia in some of the cases.

In the second sample, which represented a general population, we found a much lower classification performance of 59%. Although negative predictive value was high (96%), meaning that most people who are classified as not dyslexic are indeed not dyslexic, positive predictive value was very low (11%), also lower than in the first sample. This resulted from a high percentage false alarms of 43%, much higher than estimations of prevalence of 5–15%. Before drawing conclusions about the generalizability of the trained classifier, we should discuss three issues.

First, the first sample was small and consisted of equal groups with subjects only being selected when having dyslexia or no dyslexia beyond any reasonable doubt, while the second sample was large and represented a general population. Generally, it is widely accepted that subtypes of dyslexia can be distinguished (e.g. Ramus and Ahissar, 2012). Although we found that the students with dyslexia in the first sample can be characterized by five different impairments, we cannot exclude the possibility that in such a small sample of a specific subpopulation (first-year college students) one or more cognitive aspects of dyslexia are over- or under-represented as compared to a general population. If either of these possibilities would have been the case, the trained classifier was trained on a subgroup of people with dyslexia. This might have compromised classifications in both samples. For instance, specific subtypes of dyslexia may be characterized by specific compensation strategies with specific anatomical consequences. It cannot be ruled out that also students without dyslexia are characterized by the same anatomical consequences because their training histories during school days resemble those of students with dyslexia.

A second issue is the criterion of dyslexia in the second sample. This criterion was established based on diagnoses during school days by specialists. But these records were not specified and were uncontrollable. Based on this criterion, we found a prevalence of dyslexia of seven percent, which may be too low. It can be assumed that some students without records of dyslexia still have dyslexia, while some students with an official certificate of dyslexia have no dyslexia. Although it cannot explain the large number of false alarms, overall classification performance could have been better with better criterion groups.

A third issue is that the two samples were different regarding intelligence, socio-economic status, or other characteristics. One clear difference between students of a university and other young people is that students have received more training than other young people in all kinds of language-related and other cognitive abilities. And in the Netherlands, school children with dyslexia usually receive additional remedial teaching. When we hypothesize that the anatomical differences of the anatomical trainer in this study partly result from training effects, it might be explained why the number of false alarms in the second sample was larger than in the first sample. Subjects without dyslexia, but with low socio-economic status or low intelligence, may have received additional training as well, resulting in a diagnosis of dyslexia in this study. The hypothesis that the anatomical trainer in this study results from training differences is supported by various studies showing effects of training on anatomical alterations in dyslexia (e.g. Hoeft et al., 2011; Krafnick et al., 2011). Two studies report that many GM volume differences between dyslexics and controls in general result from differences in reading experience (Clark et al., 2014; Krafnick et al., 2014).

In short, the classifier found in the first sample performed above chance in the second sample. This underlines its reliability and justifies its further theoretical examination of the areas that contributed to this classifier. However, we conclude that the trained classifier based on anatomical scans of students of this study cannot be used for clinical purposes. Although negative predictive values were high in both samples, positive predictive values were low in both samples. This means that many people without dyslexia would be labelled with dyslexia.

Anatomical classifier

While the usefulness of a trained classifier based on anatomical scans for clinical purposes requires further examination in future studies, the nature of the classifier in this study provides useful information as compared to previous results of brain imaging studies. In the present study, brain regions that contributed to the classification were found in the LIPL, the LOFG and ROFG. These results are in line with converging evidence of involvement of these areas in dyslexia. For instance, in the classification study of Tanaka et al. (2011), poor reading children exhibited significantly reduced activations in the LIPL and the LOFG during phonological processing. We will discuss the areas of the present study one by one.

In the present study, GM volume in the LIPL correlated positively with performances of spelling and phonology, which is consistent with various previous findings. Reduced GM volume in the LIPL has been reported in pre-reading children with a family-history of developmental dyslexia (Raschle, Chang & Gaab, 2011). These researchers suggested that some structural alterations in developmental dyslexia may be present at birth or develop in early childhood prior to reading onset. They also found a significant positive correlation between this area and a rapid automatized naming test, which is assumed to be related to phonological skills (Vaessen, Gerretsen & Blomert, 2009; Vaessen and Blomert, 2010) and which is reported to be one of the main precursors of later reading ability in children (e.g. de Jong and van der Leij, 1999). These results suggest that some anatomical differences related to phonology in the LIPL may be present already at birth. Furthermore, the LIPL has been reported in functional brain imaging studies that showed that multiple specializations along the visual word-form system were found to be impaired in dyslexics (van der Mark et al., 2011), which is consistent with the reduced activations found in the study of Tanaka et al. (2011).

While in the present study dyslexics exhibited less GM volume in the LIPL, they exhibited more GM volume in the LOFG and ROFG. And although GM volume in both areas correlated (negatively) with severity of dyslexia, only the LOFG correlated with behavioural measures: negatively with whole-word reading and negatively with spelling (in contrast to the positive correlation found between the area in the LIPL and spelling). These negative correlations are remarkable: better performances on whole-word reading and spelling are accompanied by reduced GM volume. Maybe this should be interpreted as the result of training effects with poor performances leading to more training and thus to augmented GM volume. Interesting here is also the finding that poor reading children exhibited reduced activations in the LOFG during phonological processing (Tanaka et al., 2011), while no significant correlation was found between this area and phonology in the present study.

Clearly, the relation between GM volume and functionality in the LOFG is hard to understand. However, it is also clear that the LOFG is an important area in dyslexia. In previous studies, support was found for the involvement of the LOFG in dyslexia, but mainly in the VWFA. In the present study, the area in the LOFG is located close to where the VWFA is usually reported, although we were not able to establish whether this area is actually the VWFA. Nevertheless, the correlations between the LOFG and spelling and whole-word reading are consistent with previous findings related to the VWFA. For instance, various studies reveal that the VWFA plays an important role in early stages of whole-word recognition and serial sub-lexical coding of letter strings (Dehaene and Cohen, 2011; Glezer, Jiang & Riesenhuber, 2010; Schurz et al., 2010). Furthermore, lesions in the VWFA cause pure alexia, a selective deficit in word recognition characterized by a disproportionate prolongation of reading time as a function of word length (Pflugshaupt et al., 2009). Possible training effects are supported by showing that the category-selective nature of the VWFA for visually presented words is dependent of experience with specific orthographies (Baker et al., 2007).

Anatomy and functionality

Although the brain areas found in this study can be related to previous findings, interpretations are hard to make. In general, dyslexics are found to have less GM volume in various areas, but some studies report more GM volume in some areas (Silani et al., 2005; Vinckenbosch, Robichon & Eliez, 2005). Likewise, reduced as well as augmented activations have been reported in the literature (e.g. Richlan, Kronbichler & Wimmer, 2011). Some cognitive aspects of dyslexia that are typically impaired in people with dyslexia (e.g. phonological awareness, visual/attentional processing) correlate and others do not correlate with brain volume or activation.

In this study, we found a relationship between three behavioural measures of dyslexia (spelling, phonology and whole-word reading) and brain anatomy but no correlations for short-term memory and visual/attentional confusion, confirming results of previous studies showing that different symptoms of dyslexia exist at different levels of brain organisation. A complicating factor is that anatomical and functional differences may change in the course of a lifetime as the result of differences in training. Another complicating factor is that subgroups of dyslexia exist that suffer from different types of symptoms (e.g. Bosse, Tainturier & Valdois, 2007), which may be related to differences between languages, differences in socioeconomic status, differences in intelligence or differences in schooling.

For example, previous studies have shown that some aspects of dyslexia appear to be genetically induced (e.g. Carrion-Castillo, Franke & Fisher, 2013), while other aspects are related to the development of cognitive abilities and training effects in general (e.g. Clark et al., 2014; Hoeft et al., 2011; Krafnick et al., 2011; Krafnick et al., 2014). This might result in puzzling findings such as those in the classification study of Pernet et al. (2009b). It was found that voxels in the right cerebellar declive and in the right lentiform nucleus classified subjects with and without dyslexia correctly. Remarkably, regarding the cerebellar declive two subtypes of dyslexics could be distinguished: one subtype having more GM volume than controls, and one subtype having less GM volume than controls. After behavioural analyses, the researchers found that these brain phenotypes relate to different deficits of automatization of language-based processes. Thus, it even may be the case that different subgroups of people with dyslexia are characterized by different training histories, either induced by additional training programs or by compensation strategies.

Even more complicating for the interpretation of the relation between brain measures and behavioural measures of dyslexia is that we found both positive and negative correlations between behavioural measures and the brain. For instance, good spelling performances correlated positively with GM volume in the LIPL, but negatively with GM volume in the LOFG. In general, it is hard to interpret the difference between positive and negative correlations because the relation between brain anatomy and functionality remains unclear. The negative correlations in our study may point to effects of more training of people with dyslexia in comparison with people without dyslexia. Especially in our student sample, it might be expected that the students with dyslexia were encouraged to participate in remedial teaching programs during childhood, because dyslexia tends to be more disturbing when the discrepancy with intelligence is high. Support for this view can be found in a recent study in which fMRI was used to investigate the extent of anatomical overlap between three neural systems which are associated with dyslexia in the literature: the auditory phonological, the visual magnocellular and the motor/cerebellar systems (Danelli et al., 2013). Various areas of conjunction were found in the occipito-temporal cortex at more or less the same locations as our two areas in the LOFG and ROFG.

In short, we found that both augmented and reduced GM volumes contribute to an anatomical classifier. Possibly, reduced GM volume in the LIPL may be caused by genetic influences while augmented GM volume is related to training effects which differ between students with and without dyslexia, but also between various subtypes of students with dyslexia and between students without dyslexia who are characterized by different training histories. We also found both positive and negative correlations between these areas and behavioural measures. The sample that was used for creating the classifier consisted of well-educated students which might have influenced the classifier with regard to training effects. It is unknown to what extent all subgroups of dyslexia were represented by the classifier in a balanced way. Nevertheless, the areas that contributed to the classifier are consistent with various findings from previous studies. Observations made in this and other studies not only show that different aspects of dyslexia exist at different levels of neural organisation but also that dyslexia is not a unified phenomenon. Dyslexia results from an interplay between anatomy and functionality which result from both genetics as training effects, while it also should be accounted for that different subtypes of dyslexics exhibit different combinations of anatomy and functionality.

Conclusion

In summary, we report prediction accuracy of dyslexia using machine learning of anatomical scans in two samples. Various predictive values showed that the anatomical classifier of this study is still far away from use in clinical settings. However, the areas that contributed to the classification of students with and without dyslexia contribute to the understanding of brain anatomy in dyslexia. We concluded that relations between brain anatomy and functionality are hard to interpret, especially when considering effects of training and the existence of subtypes of dyslexia. We furthermore concluded that not all symptoms of dyslexia exist at the same cortical level of organisation, suggesting not only that dyslexia is not a uniform syndrome, but also illustrating that these multivariate techniques can be used to arrange and evaluate the symptoms that are believed to belong to a syndrome. Findings in this and previous studies may give directions to further research for suitable biomarkers in the future that can be used in a clinical setting.

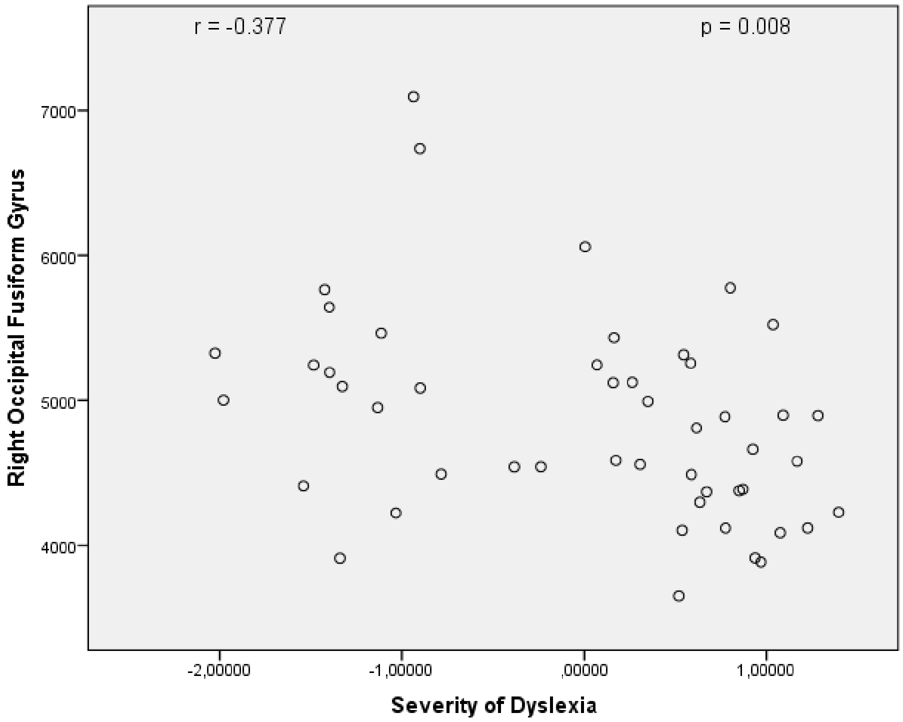
Appendix (Sample 1)

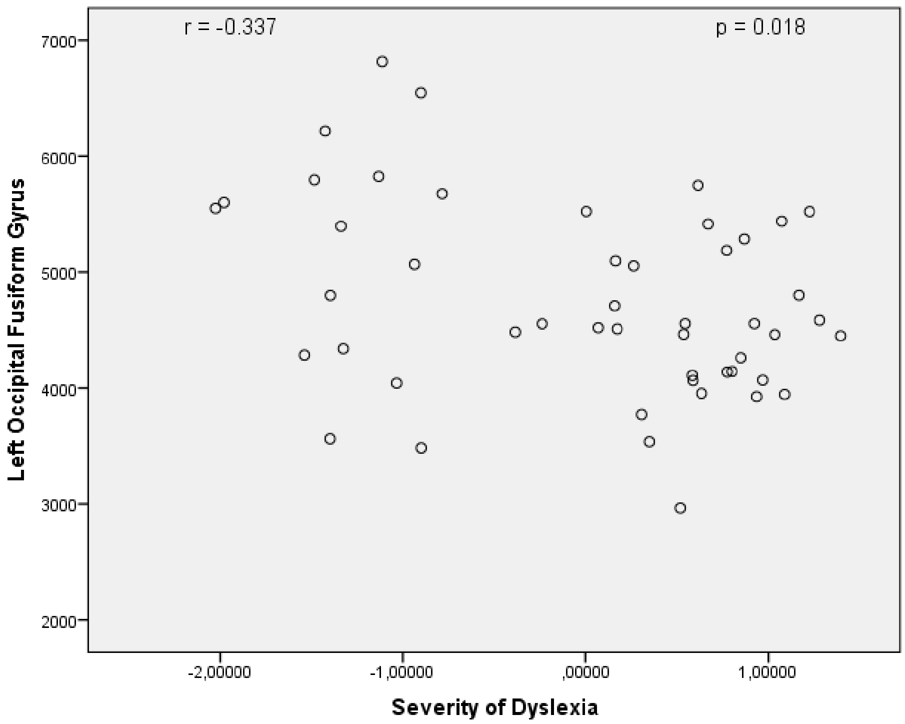
Group differences on school grades, intelligence and five elements of dyslexia

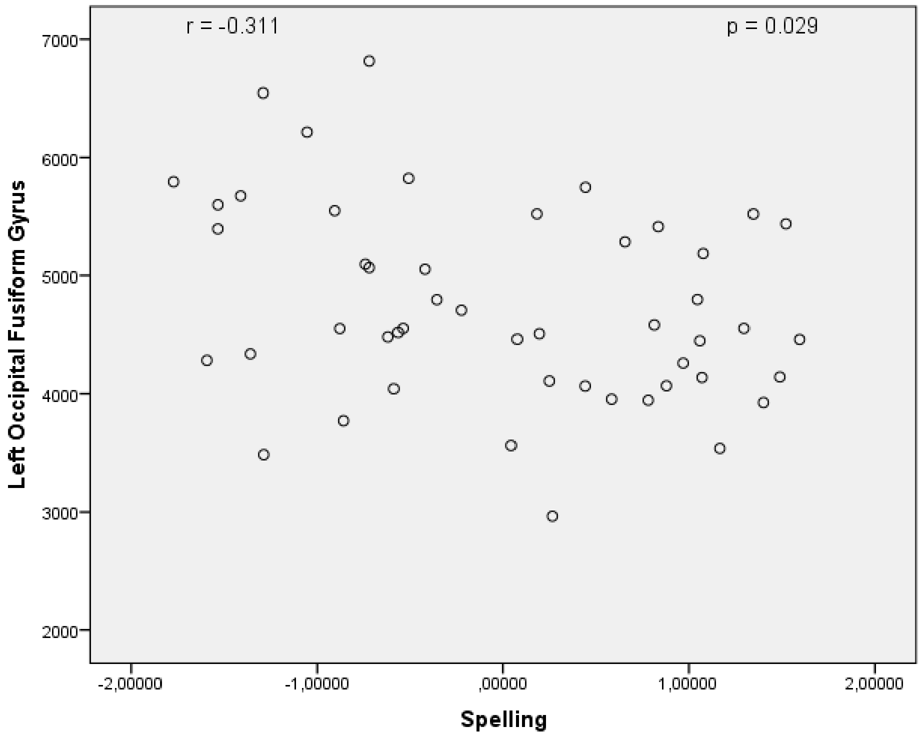
	Dyslexia (N=22)		No dyslexia (N=27)			
	Mean	SD	Mean	SD	T	*p*
Final School Grades						
Dutch	6.50	0.89	7.16	0.75	2.71	0.010
English	6.71	1.01	6.92	1.32	0.59	0.562
Other Languages	6.61	0.70	7.15	0.58	2.77	0.008
Mathematics	6.19	1.66	6.41	0.73	0.55	0.587
Other Courses	6.95	0.59	7.00	0.55	0.27	0.788
Intelligence (factor scores, normalized to a student population)						
Non-verbal intelligence	-0.28	0.86	-0.06	0.98	0.74	0.466
Speed of numeric processing	-0.65	0.62	-0.30	0.82	1.52	0.136
Vocabulary	-0.54	1.02	-0.03	1.00	1.62	0.113
Elements of Dyslexia (Z-scores)						
Spelling	-1.60	0.65	0.54	0.79	10.23	< 0.001
Phonology	-1.23	1.02	0.33	0.86	5.77	< 0.001
Short-term memory	-1.32	0.78	0.42	0.99	6.71	< 0.001
Visual/attentional confusion	-0.62	0.66	0.49	0.82	5.12	< 0.001
Whole-word reading	-1.23	0.71	0.43	0.78	7.70	< 0.001

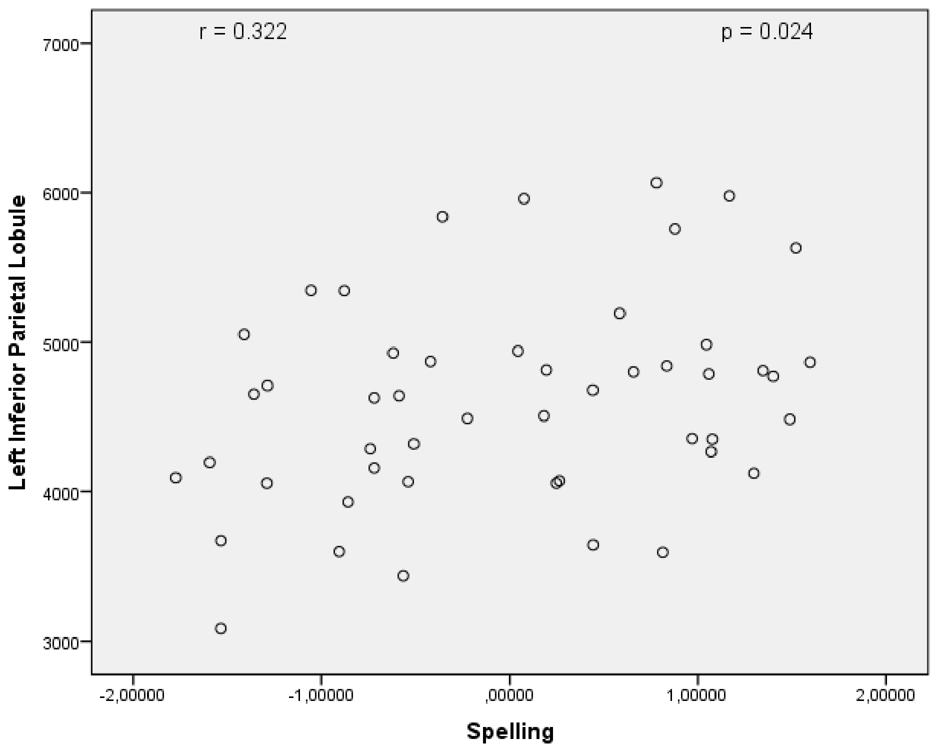
Figure 3

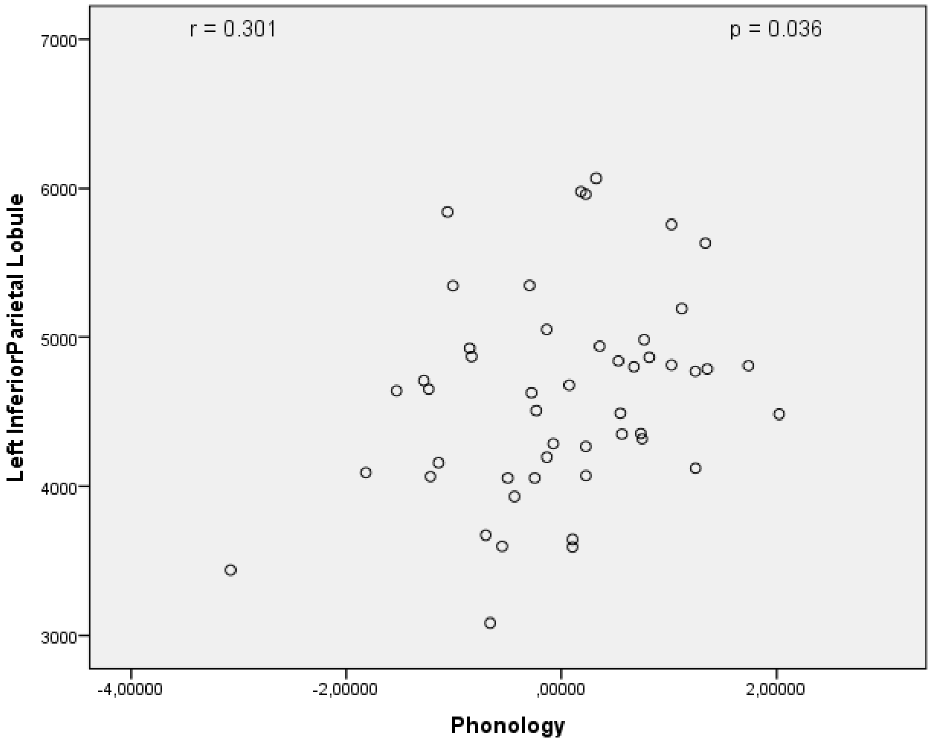
Scatter plots of six significant correlations between behavioural measures and GM volume

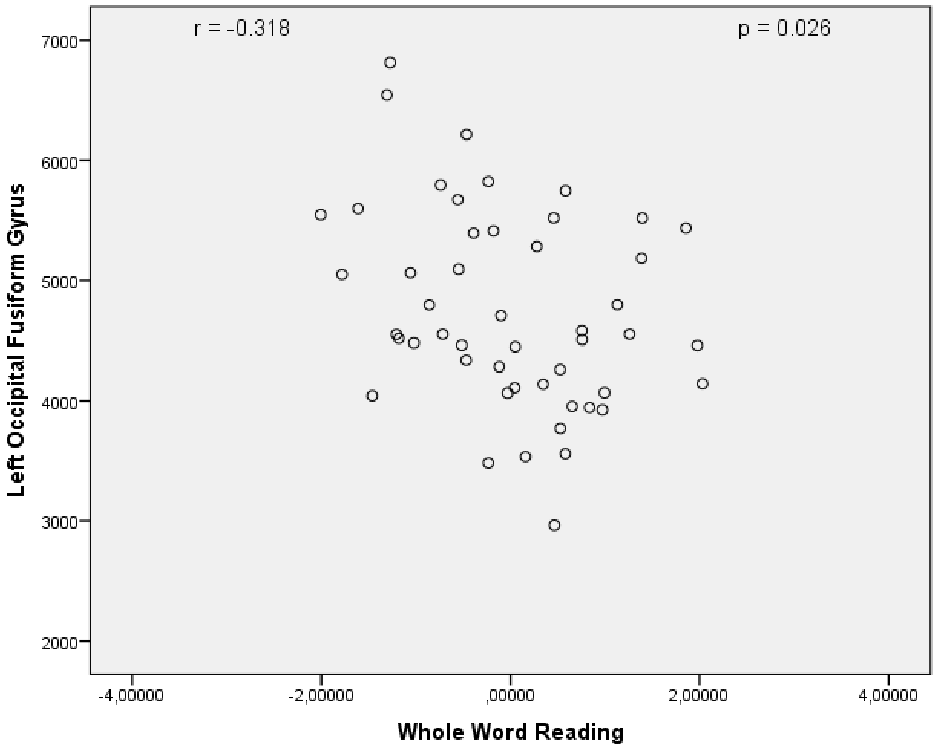





r = -0.311
p = 0.029
7000
6000
5000
4000
3000
2000
Left Occipital Fusiform Gyrus
-2,00000
-1,00000
,00000
1,00000
2,00000
Spelling


r = 0.322
p = 0.024
7000
6000
5000
4000
3000
Left Inferior Parietal Lobule
-2,00000
-1,00000
,00000
1,00000
2,00000
Spelling


r = 0.301
p = 0.036
Left InferiorParietal Lobule
7000
6000
5000
4000
3000
-4,00000
-2,00000
,00000
2,00000
Phonology


r = -0.318
p = 0.026
Left Occipital Fusiform Gyrus
7000
6000
5000
4000
3000
2000
-4,00000
-2,00000
,00000
2,00000
4,00000
Whole Word Reading

8

General Discussion

The purpose of this thesis was to investigate whether diagnoses of dyslexia can be improved with better a methodology or with self-report questionnaires and brain-imaging techniques. Additionally, we hypothesised that improved diagnostics can contribute to a better understanding of theories of dyslexia, especially regarding the cause or causes of dyslexia.

All conclusions will be discussed thoroughly and not briefly, as is often the case in conclusions of theses, because the methodology of dyslexia is extremely complex. Each study in this thesis only deals with a limited number of issues of this complexity. In each of these studies, we introduced some new points of view regarding diagnoses and theories of dyslexia. Therefore, an extensive overview of the results is warranted for a good understanding of new hypotheses. Moreover, some of my conclusions are based on the overview of results itself.

References to the six separate studies are provided by numbers in the order of appearance in this thesis. Study 1 stands for the first publication (i.e. Chapter 2 in this book), and so on.

1. Methods for diagnosing dyslexia

In Study 1 (Chapter 2), Study 2 (Chapter 3), Study 3 (Chapter 4), and Study 4 (Chapter 5), we found that the methodology of diagnosing dyslexia can be improved in various ways. In Study 1 (Chapter 2), we investigated how dyslexia can best be identified in a large sample of students. We used a new digital battery of tests and a large self-report questionnaire. In Study 2 (Chapter 3), we investigated how many underlying and independent cognitive variables could be distinguished from the whole set of tests and questions, and how these variables are related to dyslexia. In Study 3 (Chapter 4), we investigated predictive and construct validity of the self-report questionnaire in a new sample. In Study 4 (Chapter 5), we compared predictive and construct validity of a short version of the digital test battery and questionnaire with an existing test battery. In the four studies together, we found four improvements for diagnosing dyslexia.

The best statistical analyses

We investigated classification accuracy with regression analyses that determine group membership (i.e. discriminant analysis and logistic regression analysis). These analyses determine the optimal combination of tests and assign a weight to each test that was selected. In Study 1 (Chapter 2), Study 3 (Chapter 4), and Study 4 (Chapter 5), we found high classification accuracy using these analyses, higher than it could have been accomplished using arbitrarily determined cut-off scores.

Theoretically, using regression scores for determining dyslexia and non-dyslexia could show where the best benefits of different samples lie. In one sample, a spelling test may be the better predictor, while in another sample a phonological or visual test could be the better predictor. Unfortunately, the use of regression analyses is not common practice. It may be even so, that not the quality of the different test batteries determines predictive validity but the method used for analysing the tests.

Pohar, Blas and Turk (2004) evaluated differences between logistic regression analysis and discriminant analysis in various situations. According to these researchers, discriminant analysis should be preferred over logistic regression analysis in the case of small sample sizes, when violations of normality of predictor variables are not too bad and in the case of categorical explanatory variables with four or more answer categories. We analysed this and found confirmation for these findings in Study 1 (Chapter 2) and decided to administer discriminant analysis in Study 3 (Chapter 4) and Study 4 (Chapter 5). In general, each researcher should carefully decide which analysis is most appropriate for the test results at hand.

It is also important to realise that, with discriminant analysis, various choices can be made. For an objective determination of which tests to use as predictors of dyslexia, the *stepwise method* is preferable over the *enter method*, because this results in a reduced set of predictors without assigning some predictors higher priority than others beforehand. Setting prior probabilities to *all groups equal* is also most objective, because prevalence of dyslexia is unknown in a general population, but it is especially unknown in specific samples. A cross-validation procedure by selecting the option *leave-one-out classification* is highly recommended to prevent that a single case would be classified based partly on itself, and to minimise the effects of outliers.

Finally, we recommend to always carefully interpret classification accuracy. Any classification accuracy is almost meaningless when a sample is not randomly drawn from a certain population or when participants are excluded (e.g. participants with low intelligence, other disorders, etc.). However, even when the sample is random, it is still recommended to present sensitivity and specificity separately. When one group (i.e. the group with no dyslexia) exceeds the other group (i.e. the group with dyslexia), it can happen that overall classification accuracy is high while sensitivity is low. Also, for diagnostic purposes, positive and negative predictive values should be presented when a sample is randomly drawn.

Self-report statements

We found high construct validity and high predictive validity for self-report questions, with predictive validity even higher than predictive validity of tests. In Study 2 (Chapter 3) and Study 4 (Chapter 5), exploratory and confirmatory factor analyses showed that underlying variables of the questionnaire highly correlate with underlying variables of test batteries that measure all well-known symptoms of dyslexia. What explains this success of self-report questions in comparison with test scores?

One reason is that the questions in our studies are specific statements with seven answer categories of a Likert scale, which requires respondents to specify their level of agreement or disagreement on a symmetric agree-disagree scale. The creation of many statements was based on recommendations of Belson (1986). Most statements are long and specific and not short, which is common in many existing questionnaires. Thus, we did not use statements such as ‘I experience difficulties with spelling’, but ‘When I write a word that sounds like another word, I get confused with spelling of the correct letters’. To make statements as specific as possible, we frequently had them start with a condition (‘When I …’) and end with a difficulty. The specific nature of the statements, in combination with seven answer categories, offers the opportunity to respondents to report various specific difficulties as well as strengths with high precision. This is not possible with test items because these usually only have two possible outcomes (i.e. correct or incorrect), which makes them useless for measuring levels of ability. Specific measurement is also difficult to accomplish with sum scores of tests because many test scores represent performances on global abilities, such as ‘spelling’ or ‘phonological awareness’ and not on specific abilities.

A second reason is that some statements enhance recognition. The fact that people with dyslexia usually experience difficulties learning a language is well-known to both researchers and people with dyslexia themselves. It remains unknown, however, what specific differences there are between dyslexia-related difficulties and language difficulties that are experienced by many other people as well. Some self-report statements were found to have high predictive value in all samples investigated in the studies of this thesis. These statements represent combinations of specific difficulties in a context that exactly represents the essence of dyslexia. For instance, one such ‘super statement’ is: ‘When doing a dictation, I almost automatically write down the words without mistakes.’ This statement correlated with seven different factors that were found in Study 4 (Chapter 5) (Table 12). Apparently, this statement represented the exact nature of difficulties that are specific for dyslexia as compared to other, more general difficulties. Respondents may have recognised themselves especially in this type of statements that combine various specific difficulties.

A third reason is that self-report statements can do what no test can do: measure *relative* performances. The one thing that all symptoms of dyslexia have in common is that they relate to cognitive abilities. Thus, every test of dyslexia – whether it is a phonological, visual, attentional, or memory test – measures some kind of cognitive performance. Thus, as I argued before, it is almost impossible to make a test that only measures one single cognitive ability. There is always variance explained by general intelligence, schooling background, or motivation to perform well on a test. Smart people, on average, perform better than low intelligent people on every test. The advantage of a questionnaire is that people tend to compare themselves with people from their own social environment. Therefore, that is exactly what we need for reliable predictions. We do not want to know whether a respondent has good spelling abilities; instead, we want to determine whether they have good or poor spelling abilities as compared to people with a comparable background. For instance, classrooms in a primary school, on average, represent children from the same social environment. Classrooms in a secondary school or in higher education represent children or students with comparable intelligence. A child assesses his or her specific abilities – as indicated by specific statements – with children who mostly resemble them. For instance, ‘My friend is just like me, but I make more spelling errors when writing difficult words’. This *relative* performance reflects the difference between ‘normal errors’ and ‘dyslexia-specific’ errors. Some colleagues of mine put forward the notion that questionnaires cannot be objective. However, what does that even mean? People are generally in the perfect position to evaluate their own abilities. Additionally, in comparison with tests, someone who fills in a questionnaire does not have to be well-rested or perfectly concentrated. There is no need to be nervous, and when a participant is nervous, it will not affect the answers. There is also no reason to lie because, in general, people wish that both their strengths and their weaknesses are acknowledged. Children (but also adults) very much like to answer questions about themselves, mostly because this rarely happens. Teachers and parents tend to give instructions, advice, and recommendations, but they rarely ask children what they think about their own capacities. Of course, all the above must be investigated more thoroughly but it is consistent with very high predictive validity of self-report statements in all samples of this thesis.

Considering the arguments above, self-report statements may provide high predictive validity in all sorts of samples of various ages and various languages. For instance, let us consider the symptom of poor short-term memory. This symptom may express itself in various languages and ages in very different ways. A child with dyslexia at primary school may, for instance, experience difficulties in learning the capitals of countries for an exam the next day, while an adult with dyslexia can experience difficulties in memorising the tasks that have to be performed at work. An English child can experience difficulties in remembering the spelling of difficult words, while an Italian child can experience difficulties in remembering the rules of grammar. A questionnaire consisting of many statements makes it possible to represent the same symptom in different ways, so that all people with dyslexia are confronted with the specific difficulty they experience. It suffices to adjust the regression formulas in such a way that different statements or the same statements with different weights are selected.

High predictive validity of self-report questions means that questionnaires can be considered as an alternative for test batteries in various situations. Test batteries are used for clinical diagnoses, preliminary screening on schools or institutions, and for classifying participants of samples for studies. In all situations, questionnaires can have an additional value. For screening purposes and for studies, questionnaires might be considered as a full alternative for test batteries. Test batteries are time consuming for a participant, as well as for a psychologist or scientific researcher. In various situations, high costs are required. Questionnaires can be digitalised and subsequently assessed within thirty minutes, they require few costs and hardly any time investment from the psychologist or researcher.

Measuring as many abilities as possible

In Study 2 (Chapter 3) and Study 4 (Chapter 5), we found that at least five but probably more cognitive abilities are related to dyslexia. These findings are highly consistent with the findings in a comparable study of Callens et al. (2014). For a theoretical interpretation of these abilities, see Chapter 3. The relevant conclusion here pertains to the methodology: highest predictive validity is achieved when many of these abilities are represented in a predictive regression equation.

The explanation that came forward from this thesis is that the number of variables that is needed in a regression equation for high classification accuracy depends on the homogeneity or heterogeneity of the given sample. Some samples are relatively homogeneous after excluding various participants. This means that the remaining participants within both groups resemble each other. It then makes perfect sense that only a few disabilities characterise the participants with dyslexia. For instance, in a study of Tops et al. (2012), only three tests sufficed for high classification accuracy of students who were recruited from a specialised dyslexia centre. In our studies, no participants were excluded, apart from fraudulent behaviour on tests. Thus, compared to the study of Tops et al., our sample not only included students with a prior diagnosis of dyslexia and students without any typical difficulties but also students with some typical difficulties without a prior diagnosis of dyslexia. This resulted in a relatively heterogenous sample.-What we consistently found is that more variables were needed for high predictive validity when a criterion of dyslexia is more flexible and more heterogenous.

The main conclusion is that, apparently, people with dyslexia are very different from each other, just like people without dyslexia. People differ not only in regard to schooling, additional training, and general intelligence but also in more specific forms of intelligence, such as fluid and crystallised intelligence as well as verbal, numeric, and spatial intelligence, and so on. For people with dyslexia, these individual differences result in various ways of dealing with difficulties resulting from dyslexia. Some people with dyslexia may compensate for some difficulties in different ways, depending on the individual strengths they have. As a result, reliable diagnoses of large and heterogenous samples require measurements of many difficulties. In other words, fishing is easier in a homogeneous pond.

Flexible criteria and multiple predictions

We investigated the so-called criterion of dyslexia and non-dyslexia. A criterion means that it is determined beforehand who has dyslexia and who has not. This is, as argued, an arbitrary decision. Usually in diagnostic studies, a strict criterion is chosen: participants of whom we can be sure whether they have dyslexia or not. This is mostly based on former test results, for instance, at an official dyslexia centre. In our studies, we started in the same way. We collected all kinds of biographical information (e.g. former test results, official documentation of dyslexia, history of dyslexic family members, and self-assessment of dyslexia) and categorised all participants in three groups: (1) dyslexia with high certainty, (2) no dyslexia with high certainty, and (3) the remaining. Prediction analyses were performed based on the first two groups, but all participants were categorised.

An important improvement in this standard procedure is to re-evaluate the criterion groups. What we did, was to categorise some of the remaining participants into the groups dyslexia and non-dyslexia, based on the predictions of the analysis. This resulted in more flexible criterion groups. Then, we performed the whole procedure again. The amazing result was higher predictive validity, with more variables in the regression equation; however, this made sense. Strict criterion groups consist of people with severe difficulties, resulting from dyslexia and super-performing controls. Adding more 'moderate' people to the criterion groups requires measuring more variables, but it also allows more people with dyslexia to get 'recognised' by the analysis. An objection against this procedure is that it is statistically incorrect because predictions are partly based on previous predictions. In Study 1 (Chapter 2), however, we found that even after repeating this procedure three times, the final classification was highly consistent with the original biographical criterion groups. Thus, no participants from the strict criterion groups were re-assigned to another group. The main change in classification was that participants of the remaining group were added to the group of dyslexia or to the group of non-dyslexia with increased certainty.

Although most participants could reliably be classified in all studies of this thesis, there was always a small remaining group of participants who could not reliably be classified (about 10%). Perhaps some of these participants were highly intelligent and able to compensate for most of the symptoms, while some others were low intelligent people with no dyslexia who were a bit sloppy during the assessment of tests. It is important to realise that in a general population, there will always be people who are difficult to diagnose, no matter what diagnostic procedure is used. Thus, this should further be investigated.

Conclusion

The methodology for diagnosing dyslexia can be improved in four ways. Highest predictive validity is achieved when using self-report statements in combination with test results as possible predictors in regression equations, representing many symptoms of dyslexia, and when predictions are repeated with various criterion groups ranging from very strict to more flexible. Currently, we refer to the improved procedure as the 'Repeated Criterion Classification Procedure' (RCCP).

2. Investigating dyslexia with MRI

In Study 5 (Chapter 6) and Study 6 (Chapter 7), we investigated how dyslexia is represented in the anatomy of the brain with Magnetic Resonance Imaging (MRI). Although the methodological interpretation of brain-imaging results is far more complex than that of results obtained from tests or questionnaires, such complexity added many new insights to the diagnostic process of dyslexia and provided theoretical insights in general. The main conclusion is that every interpretation of structural findings regarding dyslexia depends on the interpretation of what should be considered as statistically significant. Regarding this, there are several important issues.

The first issue is about the methods used for analysing anatomical data. First, anatomical data are usually processed with the automatised technique, Voxel Based Morphometry (VBM), but the choices made by researchers regarding the pre-processing of structural images – such as modulation, adjustment of head size, or the choice of smoothing kernel size – vary between studies and may alter the significance of the results (Henley et al., 2010). Second, the significance of the results depends on choices about which technique is used for controlling for confounding variables, such as age, gender, and brain size (Snoek, Miletić, & Scholte, 2018). Third, the significance of the results depends on choices about which technique is used for corrections for multiple comparisons (i.e. Bonferroni, False Discovery Rate, and random fields thresholding), and these vary between the anatomical dyslexia studies.

The second issue is about what it means when a group difference is not significant. Tests of significance with, for instance, a required *p*-value of less than 0.05 were once introduced to prevent researchers from interpreting group differences that could easily be attributed to normal probability distributions. When a *p*-value is below 0.05, this means that there is strong support for an effect. When a *p*-value is above 0.05, there is no strong support for an effect. However, a *p*-value of 0.05 is not a border between ‘true’ and ‘not true’. A *p*-value below 0.05 can still be the result of chance, and a *p*-value of just above 0.05 does not tell us anything about whether a group difference exists or not. For instance, when a *p*-value is 0.052, a group difference more likely than not exists. In ‘normal’ research, this is not relevant because it is relatively easy to replicate findings using large samples. With the limited data of small and different samples available of MRI, however, this issue is very relevant. Thus, the conclusion that group differences do not exist when the *p*-value is just above 0.05 is not justified. Nevertheless, this conclusion may be a leading factor in a journal’s decision to accept or reject an MRI study. Study 5 (Chapter 6) was accepted only when we added significant correlations between various factor scores of symptoms of dyslexia with brain anatomy, which is something we originally planned as a separate study. This raises the question of how many relevant studies were never published. Thus, publication bias should be accounted for when interpreting general findings from MRI studies. For instance, Ramus et al. (2017) suggest that many of the reported differences in small-scale studies may be false-positive results. However, if non-published insignificant results were just a little bit below any cut-off for significance, these results may, after combining with published results, very well be significant after meta-analysis. Moreover, the fact that various significant correlations were found with brain areas that represented non-significant group differences (in our study and in a study of Pernet et al., 2009a), supports the possibility that these group differences may exist. Only considering a few meta-analytical results as being relevant is a false interpretation of the limited MRI data, which may lead to false conclusions regarding any theory of dyslexia.

The third issue is about samples. Most anatomical studies used small samples of less than 20 participants per group (either dyslexic or non-dyslexic). Most of these studies also reported several areas of significant group differences. In our study with a larger sample, no significant group differences were found, consistent with one earlier study that used a relatively large sample (Pernet et al., 2009a). Two meta-analyses of nine MRI studies showed that hardly any significant group differences exist (Linkersdörfer et al., 2012; Richlan, Kronbichler & Wimmer, 2012). Thus, how should we interpret all of this? Most striking is that low power of small samples results in more significant group differences than high power with large samples or with meta-analyses. Ramus et al. (2017) even found that across 20 published whole-brain MRI studies there is a clear trend in the opposite direction of what would be expected, with small-scale studies contributing disproportionately to the differences reported in the literature. A statistical explanation may be derived from differences between the samples. Dyslexia is characterised by various difficulties. Small samples can be expected to be relatively homogeneous regarding biographical variables (i.e. age, gender, intelligence, level of education, social background, and native language) that affect the severity of various symptoms of dyslexia. Thus, in small samples, some difficulties may be more prominent than others. On the other hand, large samples may be more heterogeneous regarding biographical variables, which might result in more heterogeneity regarding various typical difficulties. Therefore, large samples may also include more people without one specific difficulty, and thus decreasing power for that difficulty. Only small differences regarding heterogeneity between small and large samples may result in decreasing power, especially considering that even the large samples just exceed the minimum of about 30 participants per group which is required for group comparisons to be reliable. Combining small samples to one large sample does not help. The small samples are completely different regarding age, gender, intelligence, level of education, social background, and native language. Moreover, some of the typical difficulties accompanying dyslexia vary to a large extent depending on these biographical differences. Summation of these samples may decrease the power of various results that are only characteristic for specific small samples. This is consistent with conclusions from the previous paragraph in which I argued that homogenous samples are characterised by only a few specific symptoms of dyslexia, while heterogenous samples are characterised by a large diversity of symptoms across individual participants. Larger diversity of symptoms results in less statistical power for each separate symptom, resulting in no significant group differences.

The fourth issue is about using a classification approach as an alternative for traditional *t*-testing. In Study 6 (Chapter 7), we found a significant classification accuracy of 80%. In a large independent sample, we found a much lower but still significant classification accuracy of 60%. For clinical purposes, these percentages are of course too low. However, another important result of this study is that differences between groups were found using the same sample as in Study 5 (Chapter 6), whereas traditional analyses of group differences found nothing. Three areas of the brain were involved in this classification, and the most stunning result is that these areas were completely different areas than the ones found in Study 5 (Chapter 6). The best explanation is that this represents a more general methodological issue: the difference between measurement and prediction. The combination of small (non-significant) differences may distinguish people with and without dyslexia, while these separate differences may even be too small to detect with analyses of group differences. This confirms three things. First, diagnoses of dyslexia using anatomical MRI cannot be done based on a single area, just as diagnoses with tests or questions require more than one score in a regression equation. Second, group differences cannot be used as a diagnosis. This confirms that regression equations are better predictors of dyslexia than arbitrarily determined cut-off scores. Third, in a relatively large sample, dyslexia is represented by various symptoms. Determining which of these symptoms is the best predictor will depend on the characteristics of different samples. Here, we used a student sample. The second sample was a population sample, probably with some other characteristics when compared with the student sample.

Summarising, we found support for the hypothesis that dyslexia is represented in the brain by various anatomical varieties, using correlational analyses and classification analyses. The interpretation that non-significant group differences mean that there is nothing going on anatomically, is not correct. These findings support that the methodology of dyslexia research is extremely complex. Furthermore, these results support that dyslexia is characterised by many symptoms that vary between small and homogenous samples to a large extent.

3. Dyslexia and theoretical conclusions

From results to conclusions

The exploratory study of diagnostic methods in this thesis resulted in the conclusion that dyslexia is characterised by many different symptoms that vary between individuals to a large extent. So, what does this mean for theories of dyslexia? What else did we learn from the results presented in this thesis? Can we draw more conclusions? Before discussing this, I believe it is warranted to emphasise various aspects of drawing theoretical conclusions about dyslexia in general.

When something is fully unknown, such as the cause or causes of dyslexia, this always results in many theories. Thus, it is quite possible that no area of research of the human mind yielded so many different theories as dyslexia, with maybe intelligence as the only exception. On first sight, this seems very positive. There are, however, adverse effects. The existence of many theories makes it almost impossible for any individual researcher to keep an overview of all new theories and hypotheses and, at the same time, mirror his or her new research data to all of them. This results in islands of research groups, usually nested in different countries, with each group focussing on only a subset of theories. In England, for instance, dyslexia research focused mainly on phonological aspects of dyslexia, while dyslexia research in Italy also focused on visual and attentional difficulties. In many scientific papers about dyslexia that were published in leading journals, competing theories were not mentioned at all. How then can these different insights be combined into one universal model of dyslexia?

At present, the situation is completely unclear. New theories of dyslexia are proposed every year. Many completely different definitions of dyslexia circulate across publications. There are even researchers who do not mention the word ‘dyslexia’ anymore but replaced it with ‘reading disorder’ (e.g. Black, Xia, & Hoeft, 2017; Xia, Hancock, & Hoeft, 2017). The chaos is now complete. In this situation, some theories are declared sacred, despite contradicting evidence. This has been the case with the theory of impaired phonological awareness, which is still generally considered as the key theory of dyslexia, even though some serious objections have been proposed. After a while, it has become impossible for researchers not to look through coloured glasses anymore.

Admissions of ignorance and temporary mystification

As stated in the introduction, we wanted to avoid an initial discussion about the different theories because of the danger that this could distract us from clear methodological reasoning. In the separate chapters of this thesis, however, we evaluated many theories of dyslexia based on the methodological results. In this conclusion, I want to present a clear overview of these evaluations, and even introduce a new theoretical perspective.

Generally, there are two ways to develop a new theory. One way of developing a new theory is by using the inductive or bottom-up approach. This means that a theory is built upon existing knowledge. A common procedure is: (1) data collection, (2) formulating a hypothesis, (3) testing this hypothesis in a new sample, (4) rejecting, accepting or adjusting the hypothesis, and so on. Regarding dyslexia, the problem with this method, so far, is that the accumulating amount of data has resulted in many competing hypotheses. There is still no single hypothesis or theory that can explain all the data. Each hypothesis or theory can be rejected based on existing data, although it will always remain a topic of debate. Moreover, as explained before, no single theory can predict who has dyslexia and who does not. Of course, we can always continue this way. However, we do not know how long the road to salvation is, nor do we know whether salvation will come in the end. Maybe the origins of dyslexia are too subtle to detect with the current methods. Then again, this strategy of developing theories might just as well end in a complexity that is beyond our comprehension.

A second way of developing a new theory is by using the deductive or top-down approach. This means that a theory is built based on an intuitive idea that encompasses the core problems of a topic. In the case of dyslexia, this should be an idea that explains that dyslexia is characterised by a variety of symptoms that differ on an individual level, without one symptom being symptomatic for all people with dyslexia. The main disadvantage of this strategy is that the proposed theory could be complete nonsense, even when it explains many of the symptoms. The advantage is that it searches for an explanation directly on a deeper level.

In this conclusion, I will use a combination of both strategies to develop a new theory. First, I will present five general theoretical propositions that are crucial for understanding further steps in theory development. All propositions are based on the studies of this thesis, in combination with existing knowledge from literature, and on the fact that, despite this knowledge, we still have no clue about what causes dyslexia. Admitting that our knowledge is very limited may in fact be a crucial step for embracing completely new insights (admissions of ignorance). These propositions may lead to a theory that, based on the data that are currently available, or will be in the near future, and can neither be proven nor falsified (temporary mystification) but may inspire us to further insights and research.

Theoretical proposition 1:

Dyslexia has an unknown cause or causes, but many symptoms

All researchers agree that dyslexia is characterised by more than one symptom. Moreover, most researchers agree about the existence of impaired phonological awareness, impaired spelling abilities, reduced short-term memory, poor rapid naming abilities, and that various aspects of visual attention may be impaired. Disagreement exists, however, about the exact number of the symptoms, the severity of the symptoms, the causal nature of the symptoms, the relations between the symptoms (subtypes of dyslexia), and individual variation regarding the various symptoms. In many studies, relations were investigated between various abilities related to reading and underlying symptoms such as impaired phonological awareness and rapid naming (e.g. Hulme et al., 2015; de Jong & van der Leij, 1999; Lervåg, & Hulme, 2009; van Viersen et al., 2018). However, only a handful of studies contained analyses of relations between underlying symptoms, such as phonological, visual and attentional symptoms (e.g. Bosse, Tainturier & Valdois, 2007).

In Study 2 (Chapter 3) and Study 4 (Chapter 5), we conducted various explorative and confirmative factor analyses on both test results and the questionnaire, with Study 4 (Chapter 5) being an extended replication of Study 2 (Chapter 3). An important outcome in Study 2 (Chapter 3) was that latent variables of the tests highly correlated with latent variables of the questionnaire with the same interpretation. Furthermore, the outcomes of the analyses of the two studies showed both similarities and differences. In short, we found that the number of 'independent' variables that characterise dyslexia is at least five, but probably more, maybe even ten or so. Some variables came up in all analyses: spelling, phonology, and short-term memory. The main difference between the numerous analyses was that some variables were taken together in one analysis but separated as independent factors in other analyses, such as spelling and rapid automatised naming. We found support for eight latent variables: spelling, reading, phonology, rapid automatised naming, short-term memory, attention, confusion (rhyme), and complexity (whole-word-processing). Callens et al. (2014) found similar results: seven latent variables with high-effect sizes were distinguished from 53 tests, using exploratory factor analysis.

The best conclusion is that variables underlying dyslexia are not independent from each other. Some may seem independent in the analyses, but the analyses only describe phenotypical outcomes. The main difference between people with dyslexia and those without is that those with dyslexia suffer from relatively more impairments than those without dyslexia. We also found that none of these latent variables by itself could distinguish people with dyslexia from those without, where the combination of variables resulted in very high classification accuracy. People with dyslexia suffer from many symptoms, but not necessarily all. Attempts to distinguish subtypes of dyslexia failed in this thesis. No subtypes such as phonological versus visual were found, but the samples may have been too small for the detection of subtypes.

The first theoretical proposition is that dyslexia is caused by something unknown, but with many consequences. These consequences vary on an individual level depending on age, gender, intelligence, schooling, additional training, native language, social background, motivation, and so on. Relations between different symptoms are complex and influenced by different additional variables.

Theoretical proposition 2:
Dyslexia has probably a single cause

The proposition from the previous paragraph is that neither of the latent variables that were found in this thesis can be considered to represent a cause of dyslexia on its own. The next question to address is why most people with dyslexia suffer from many, but not all symptoms of dyslexia. A logical explanation is that symptoms of dyslexia result from an underlying cause in different ways among individuals, or that some individuals can compensate for some of the symptoms because of additional training, good schooling, or high intelligence. Another view is the multiple deficit theory (MDT) of dyslexia, as proposed by Pennington (2006), which assumes that there is no single deficit causing dyslexia but declares that dyslexia is the outcome of multiple causes.

In recent years, the MDT has become increasingly popular (e.g. van Bergen, van der Leij & de Jong, 2014; de Jong & van Bergen, 2017; Peterson & Pennington, 2015). The MDT defines dyslexia as poor reading that results from the interplay between risk factors and protective factors. For instance, impaired phonological awareness is, according to the MDT, not only a symptom of dyslexia but also one of the possible causes or risk factors that can lead to dyslexia. According to this view, for instance, poor reading (dyslexia) can be the result of poor phonological abilities while rapid naming is unimpaired, or poor reading (dyslexia) can be the result of poor rapid naming abilities while phonological abilities are unimpaired. The basic question here is as follows: Is there one single cause of dyslexia, or should dyslexia be defined as being the outcome of the interplay between various factors? Although none of these views can be proven right or wrong with the data currently available, I will argue based on the present thesis that the view of a single cause prevails over the view of multiple causes. There are five issues regarding the two competing views, which are evaluated below.

The first issue is that a single cause of dyslexia has not yet been identified. Some have argued that this is one of various general shortcomings of any single cognitive deficit model of dyslexia (e.g. van Bergen, van der Leij & de Jong, 2014), or even one of the key reasons for rejecting these models (Pennington, 2006). However, that something has not yet been found does not mean that it does not exist. Much is unknown about the human brain and we just started to investigate it. That much is unknown was confirmed in a study of Van Bergen et al. (2015) which investigated parental reading influence on children's reading. It was found that about half of the genetic risk of dyslexia was mediated by known cognitive causes. Thus, the other half of the genetic risk should be mediated by cognitive correlates that are to date unknown (de Jong & van Bergen, 2017). This is consistent with an earlier conclusion regarding developmental disorders in general: that the multiple deficit model is universally applicable to developmental disorders but, therefore, remains abstract, and that it is not specified which etiological factors, neural systems, and cognitive processes interact to produce a given disorder (van Bergen, van der Leij, de Jong, 2014). What does this mean? Should we keep looking for other possible causes of dyslexia, but forget about a single cause? I believe that it is too early to decide this. There are so many unknown symptoms that can accompany dyslexia; therefore, it would be premature to state that neither of those may represent a key feature of dyslexia, at least not before this has been thoroughly investigated, no matter how difficult this is and how long this search will last. As I will argue further on, there are reasons to believe that a single cause of dyslexia exists.

The second issue involves the number of underlying cognitive variables that has been proposed so far. In this thesis, it was shown that at least eight cognitive variables are involved in dyslexia: spelling, reading, phonology, rapid automatised naming, short-term memory, attention, confusion (rhyme), and complexity (whole-word-processing). To my knowledge, confusion and complexity have not been reported before, whereas the other six cognitive variables are well-known. In the literature of dyslexia, many theories have been proposed and, as described in the introduction of this thesis, at least 28 symptoms were described, including various visual and attentional deficits or, for instance, impaired reading of music notes (Hébert et al., 2007; Miles, Westcombe & Ditchfield, 2008; Tunmer & Greaney, 2010). However, in studies that propose the MDT as an alternative for single deficit explanations, visual and attentional theories of dyslexia were hardly considered relevant (e.g. Peterson and Pennington, 2015). This is inconsistent with the large number of theories and symptoms of dyslexia that were described in the literature. For instance, it is found that impaired phonological awareness cannot be the central mechanism that explains all characteristics of dyslexia; yet, that does not mean other possible central mechanisms should be discarded. In the studies of the present thesis it became clear that some underlying variables of dyslexia cannot be entirely independent from one another. Various exploratory factor analyses showed that certain abilities loaded on one latent variable in one analysis but on two latent variables in another analysis. The existence of interrelations between many cognitive variables justifies the assumption that there must be something more elementary that connects these variables, an elementary single cause maybe, as the *g*-factor in intelligence research. The question is then: Does this elementary factor exist, or does it merely represent the statistical outcome of interrelations between various variables related to dyslexia?

The third issue is about defining dyslexia. According to the International Dyslexia Association (IDA) (2002), dyslexia is defined as follows: ‘Dyslexia is a specific learning disability that is neurobiological in origin. It is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language that is often unexpected in relation to other cognitive abilities and the provision of effective classroom instruction. Secondary consequences may include problems in reading comprehension and reduced reading experience that can impede growth of vocabulary and background knowledge.’ Based on this definition, many researchers consider dyslexia as a (word) reading disorder, which resulted in using ‘dyslexia’, ‘reading disorder’, and ‘poor reading’ interchangeably across studies. Where the definition of the IDA only mentions a phonological deficit as underlying cause, the MDT proposes that many underlying causes are possible. However, despite the number of causes, the definition of the IDA and the use of terms like ‘poor reading’ imply that dyslexia, in fact, represents the low end of a continuous and normal distribution of a continuous disorder. Moreover, in order to diagnose dyslexia, a somewhat arbitrary cut-off must be set on a continuous variable that makes a threshold between affected and unaffected rather arbitrary (van Bergen, van der Leij & de Jong, 2014; Peterson & Pennington, 2015). In some studies, for instance, categorisation as either dyslexic or non-dyslexic is based on performances below or above the 10th percentile cut-off on task of word reading fluency (e.g. van der Leij et al., 2013). This distinction between people with dyslexia and those without has some strange consequences. Assume that a child has been diagnosed with dyslexia at the age of ten years because his or her reading performance was just below the threshold for dyslexia. Based on this diagnosis, this child starts to practice extensively under the supervision of teachers, remedial teachers, and parents. After a year, his or her reading performance has improved to a level just above the threshold for dyslexia. Does this child have dyslexia or not? This means that, for some people, a diagnosis of either dyslexia or no dyslexia will never be possible because their performances may vary just around the threshold throughout their lives. Another consequence is that people without reading difficulties have no dyslexia per definition. In fact, this is the key difference between the single-cause and the multiple-cause views of dyslexia. When multiple causes are involved, a condition such as dyslexia is by definition the low end of a continuous distribution. A single cause of a condition should be defined by the cause and not by one of many possible results of that cause, even when this cause is still unknown. For example, what happens when someone has dyslexia but does not suffer from reading difficulties, but that person does suffer from other difficulties? Some children in the Netherlands with early reading difficulties managed to overcome most of these difficulties in an early stage of the primary school; however, they still experienced major difficulties all over again in secondary school when they started to learn English, Greek, mathematics, or reading music notes. These children will never understand why they experience such difficulties. In fact, children with dyslexia who tried their very best learning to read and write on primary school were punished in comparison with children who did not practice a little extra. Only the lazy children will then be rewarded with extra time during exams in secondary school or at a university. That some people with dyslexia do not suffer too much from poor reading does not mean that their dyslexia disappeared, but only that other symptoms may be more prominent. In short, people are not very well supported by defining a disorder based on an arbitrarily chosen threshold on only one task (reading), just because underlying causes are unknown.

In this thesis, something was found that even contradicts the continuous character of dyslexia. In Study 1 (Chapter 2), the severity of dyslexia was expressed by a score, which was the outcome of the regression formula that had highest classification accuracy. The frequency distribution is presented in Study 2 (see Chapter 3, Figure 1). This score represents many symptoms of dyslexia with different weights for each regressor (with each regressor representing one or a combination of symptoms). According to the definition of the IDA and according to the MDT, a normal distribution should be expected. However, the distribution of this score shows two separate normal and continuous distributions, which do not overlap; one representing people with dyslexia and the other representing people without dyslexia. This strongly supports that dyslexia is only continuous within the separate groups of people with dyslexia and those without.

The fourth issue is about the genetic difference between dyslexia and impaired ‘normal reading skills’. Both are familial and moderately heritable; more specifically, however, the neuropsychological deficits associated with dyslexia are considered to be heritable (Christopher et al., 2013; Peterson & Pennington, 2015). Not more than ten candidate genes have been identified in relation to dyslexia (Carrion-Castillo, Franke & Fisher, 2013; Darki et al., 2012; Gialluisi et al., 2014; Kere, 2011; Scerri & Schulte-Körne, 2010; Scerri et al., 2011). Some of these genes were found to affect neuronal migration during prenatal processes in rodents (Kere, 2011). However, none of the studies that related genes to dyslexia have convincingly been replicated. Thus, although some of these genes may be involved in the development of dyslexia, it is premature to draw conclusions regarding the cause or causes of dyslexia. If we assume that dyslexia is in fact poor reading, and some genes exist that are related to reading, those genes must code cognitive functions that were present long before we started to read on a large scale. A study of Van Bergen et al. (2014) showed that, in children with dyslexia, verbal intelligence and nonverbal (arithmetic) intelligence are indeed affected as well. An explanation is that the set of genes involved in dyslexia has pleiotropic effects. These genes are involved in neuronal migration during early brain development and, as a result, also affect the development of other abilities (Plomin & Kovas, 2005; Kovas et al., 2007). However, are we now still talking about dyslexic genes? What about the consensus that dyslexia is independent from general intelligence?

In summary of this issue, there are numerous genetically determined cognitive functions underlying reading, dyslexia, and (verbal) intelligence. However, most of these functions are unknown. For example, rapid naming cannot be such a function, because this task involves reading, which we started to do only recently on an evolutionary scale. The same argument can be used for phonological awareness. Therefore, candidate genes must code for more elementary functions. Reading involves various cognitive abilities, some of which were present long before people started to read. These functions may very well be genetically determined. However, if dyslexia is defined as an unknown underlying single deficit instead of a reading disorder, findings about genes that are involved in (poor) reading cannot be used as an argument for the probability of multiple causes of dyslexia.

The fifth issue is about the comorbidity between dyslexia and for instance Attention Deficit Hyperactivity Disorder (ADHD). The MDT claims that this comorbidity is easier to explain with a multiple deficit model than with a single deficit model because this comorbidity is mediated by shared etiologic and neurocognitive risk factors (Boada, Willcutt, & Pennington, 2012; Peterson & Pennington, 2015; van Bergen, van der Leij, & de Jong, 2014; Willcutt et al. 2010). It is estimated that 15–40% of children with dyslexia are also diagnosed with ADHD, and 25–40% of children with ADHD also have dyslexia (Eden & Vaidya, 2008; Purvis & Tannock, 1997; Willcutt, Pennington, & DeFries, 2000a). According to the *DSM-V* (APA, 2013), ADHD is characterised by a persistent pattern of 'inattention and/or hyperactivity-impulsivity that interferes with functioning or development'. ADHD is also familial and significantly heritable. Furthermore, a distinction is made between three subtypes: the predominantly inattentive type, the predominantly hyperactive or impulsive type, and the combined type. For a diagnosis of each type, children must meet criteria for at least six of the nine symptoms. An alternative explanation for the comorbidity between ADHD and dyslexia may be that people with the inattentive subtype of ADHD are people with dyslexia who suffer from visual or attentional impairments. Some findings seem to be consistent with this. For instance, in various twin studies, it was found that most of the phenotypic covariance between reading difficulties and inattention was explained by common genetic influences, whereas phenotypic covariance between reading and hyperactivity-impulsivity was less explained by common genetic influences (Boada, Willcutt, & Pennington, 2012; Willcutt, Pennington, & DeFries, 2000b; Willcutt et al., 2007). This raises the question whether ADHD represents one disorder, or two disorders with one of them being dyslexia. Another explanation for the comorbidity between ADHD and dyslexia may be that ADHD might be over-diagnosed whereas dyslexia might be be under-diagnosed. We must keep in mind that diagnoses of ADHD are frequently based on symptom ratings obtained from parents or teachers, who might exaggerate the severity of some symptoms. More importantly, however, some symptoms of ADHD may just be the result of dyslexia. Children with dyslexia who are not able to read, or spell, may become frustrated, especially when teachers do not acknowledge that these children might be dyslexic, but instead tell them that they are stupid, something which is still common in many schools. This can lead to a lack of motivation and impaired concentration at school. These children may start to show either internalised or externalised behaviour (Terras, Thompson, & Minnis, 2009). The internalised behaviour can lead to a diagnosis of the inattentive type of ADHD, and the externalised behaviour can lead to a diagnosis of the hyperactive or impulsive type of ADHD. Other comorbidities can easily be explained as well. For example, approximately 30% of children with early language or speech problems go on to develop dyslexia (van Bergen, van der Leij, & de Jong, 2014). However, when we assume that dyslexia is a single heritable trait, whether or not it leads to poor reading, these early language or speech problems can also be caused by dyslexia. In the remaining 70% of the children, these problems may result from something else, such as hearing problems or low intelligence. However, it is possible that these children have dyslexia as well, although the severity of their reading problems is not sufficient for a diagnosis of dyslexia based on the MDT. Summarising, comorbidity between dyslexia and ADHD can be explained by the MDT. However, a single deficit model of dyslexia can explain this comorbidity as well. For all models, though, it should be kept in mind that the prevalence of this comorbidity depends on unknown over- or underdiagnoses of both dyslexia and ADHD, as well as on the various definitions and subtypes of these conditions.

In summation of this theoretical proposition, there are no data or convincing arguments that make a multiple deficit explanation of dyslexia more likely than one that assumes a single underlying deficit, even knowing that a single deficit still must be identified. In a study by Pennington et al. (2012), a multiple deficit model and various single deficit models were tested on diagnostic power, but none of the models could reliably predict dyslexia. Defining dyslexia as poor reading or reading disability on a continuous scale is arbitrary and only based on the assumption that dyslexia has multiple causes, and its clinical implications are not helpful for some of the children who experience serious difficulties. Evidence from genetic studies is weak because most findings could not be replicated. Furthermore, on a behavioural level, a single deficit model of dyslexia can explain comorbidity just as well as the MDT. Most importantly, in this thesis, it was found that a regression score representing severity of dyslexia was characterised by two separate normal distributions. Actually, the MDT represents the traditional bottom-up approach of developing a new theory. A single deficit model can be developed using a top-down approach, as will be explained in the paragraphs to follow.

The second theoretical proposition is that dyslexia probably has a single underlying but unknown cause. Symptoms of dyslexia result from this unknown cause in different ways among individuals, depending on the ability to compensate for some of the symptoms as a result of additional training, good schooling, motivation, or high intelligence. This results in a severity of dyslexia being normally distributed but separated from a normal distribution of people without dyslexia.

Theoretical proposition 3:

Anatomical variety results from training differences

The proposition that dyslexia is characterised by many symptoms is consistent with our findings regarding the diagnosis of dyslexia using MRI (chapter 2 of this conclusion). In Study 5 (Chapter 6) and Study 6 (Chapter 7), we concluded that dyslexia is represented in the brain by various anatomical varieties that vary between different small and homogenous samples. Thus, there may be something that causes dyslexia, which has many consequences, both on an anatomical level and on a phenotypical level.

In chapter 2, we discussed the results of MRI studies regarding sample size and significance. Yet, we should further question what it means when people with dyslexia have a different volume of grey matter in a specific area of the brain than people without dyslexia. In the introduction, we already introduced this issue and suggested that anatomical variations regarding dyslexia may be induced by genetics but may also result from training differences. Seeing that dyslexia is genetically determined, it already plays a factor in the earliest stages of cognitive development. Depending on the variables mentioned in the previous paragraph (i.e. age, gender, intelligence, schooling, additional training, native language, social background, and motivation), children may intuitively use their strengths to compensate for their weaknesses and, thus, develop different strategies when dealing with dyslexia. This may result in many differences in the severity of symptoms and in many different anatomical variations.

Maybe what we are seeing here are not genetically induced variations at all. Maybe we are only seeing the effects of training. Many areas are hardly used because they are involved in some weaknesses, and other areas are used intensively for compensation. The area that is most actively involved in training and automaticity is the cerebellum. Especially this area of the brain shows very different results between studies, with people with dyslexia in one study showing more grey matter volume in one particular area of the brain but, in a different study, showing less grey matter in another area of the brain (for good overviews of results, see Black, Xia & Hoeft, 2017; Ramus et al., 2017; Xia, Hancock & Hoeft, 2017). Most of the results of Study 5 (Chapter 6) and Study 6 (Chapter 7) support this view. For instance, a factor related to performances in spelling correlated negatively with grey matter volume in both the left and right *occipital fusiform gyrus* and the left posterior cerebellum. This may imply that reduced spelling abilities lead to more training and, thus, to augmented grey matter volume in those areas.

The third theoretical proposition is that the brain displays an anatomical variety resulting from training differences. Various symptoms result in additional training of these symptoms or, instead, of abilities that might compensate for such symptoms. On an individual level, this anatomical variety depends on age, gender, intelligence, schooling, additional training, native language, social background, motivation, and so on. Anatomical variety, resulting from genetical influences, cannot be excluded but may very well be too small to detect.

Theoretical proposition 4:

The cause of dyslexia is subcortical

The previous propositions were used to argue that dyslexia probably has a single cause that results in phenotypical and anatomical variety, with training effects as an underestimated factor. However, what could be the cause? Moreover, what could be the origin of inflammation that leads to this widespread fire?

One hypothesis is that there is something wrong with the functional connectivity in specific pathways of the brain, caused by reduced white matter density. Indeed, there are many studies supporting this hypothesis (e.g. Vandermosten et al., 2012). Another hypothesis is that the key variation is a reduced folding of grey matter (e.g. Płoński et al., 2017). However, these hypotheses do not necessarily provide a causal explanation. Reduced connectivity or folding may result from reduced functioning of separate areas in the brain and may just as well be caused by something more fundamental.

Throughout the literature of dyslexia, based on MRI or fMRI, the reported areas are located in almost the entire cortex, thus, in the cerebellum, occipital lobe, temporal lobe, parietal lobe and frontal lobe, and in both hemispheres. Until recently, however, no studies were published about possible relations between dyslexia and subcortical functioning. In Study 5 (Chapter 6), we found that within the group of people with dyslexia, a variable related to performances in Dutch-English rhyme words correlated positively with GM volume in the left and the right caudate nucleus. In another recent study, a reduction of grey matter was reported in the left thalamus (Jednoróg et al., 2015). The thalamus is involved in the distribution of sensory information to the cortex. The caudate nucleus is involved in several executive control functions, and in the processing of information. Thus, both areas are involved in processes that affect various specific functions throughout the brain.

The question now is whether anatomical variations in these subcortical areas result from training differences as well. This could be so, but if something fundamental causes dyslexia, it is more likely to be found subcortically than cortically. Only a subcortical variation can have widespread effects, such as those found in dyslexia. Furthermore, it is possible that subcortical variations result in specific effects, without affecting all functions. General and specific forms of intelligence may be spared. See Chapter 4 for a further discussion of the functionality of the thalamus and caudate in relation to dyslexia.

A subcortical variation may be too small to detect in small samples, but it can be detected in large samples. We argued that small samples lead to more significant results than large samples because small samples tend to be more homogenous. However, if a cause of dyslexia is to be found subcortically, it should be common for all people with dyslexia and power should accumulate with larger samples. This is exactly what was found. Both Study 5 (Chapter 6) and the study of Jednoróg et al. used relatively large and heterogenic samples. The sample of Jednoróg et al. even consisted of various subsamples from different countries with different languages. Very recently, a new study (Jagger-Rickels, Kibby & Constance, 2018) confirmed involvement of the caudate nucleus in dyslexia and ADHD, while thalamus involvement was found for a comorbid dyslexia and ADHD group.

The fourth theoretical proposition is that dyslexia is nested somewhere in the subcortex, but differences between people with and without dyslexia are too small to be detected in small samples. In large samples, anatomical effects in the caudate and thalamus are sometimes visible. The subcortical cause of dyslexia causes widespread anatomical effects that influence specific cognitive functions while general and various specific forms of intelligence remain intact.

Theoretical proposition 5:

Dyslexia exists, is not a disorder, but a dichotomous perceptual variation

In summary of the first four propositions, the cause of dyslexia is unknown and maybe too small to detect, but with widespread effects throughout the brain which results in a large variety of symptoms. About the same could be stated about (poor) intelligence; but then, would that classify dyslexia as a specific form of intelligence? Does dyslexia exist at all? These questions were raised in an impressive overview of studies on dyslexia by Elliot and Grigorenko (2014). What can we conclude from this thesis regarding such questions?

In ‘Theoretical proposition 2’, I discussed the findings of Study 1 (Chapter 2), indicating that a severity score of dyslexia showed two separate normal and continuous distributions. The existence of two separate normal distributions instead of one, supports the notion that dyslexia exists and is a dichotomous condition, and not that dyslexia is the utter left side of a normal and continuous distribution that represents some kind of cognitive ability, such as reading. Variations within both distributions are easily explained by variations regarding age, gender, intelligence, schooling, additional training, native language, social background, motivation, and so on. The distribution of this score shows two normal distributions that do not overlap: one representing people with dyslexia and the other representing people without dyslexia.

The idea that dyslexia is a dichotomous condition is not strange. There are more dichotomous conditions that originate from the brain, such as hand preference and sexual preference, although some people are ambidextrous or bisexual. Thus, these human traits are divided into three groups. Estimations of prevalence of these groups are not very reliable in the literature. However, it is striking that these estimations are about the same for dyslexia. For these three traits, there is a majority group of about 75–90% of all people (i.e. heterosexual, right-handed, and no dyslexia), a small group of about 5–15% of all people (i.e. homosexual, left-handed, and dyslexia), and a very small group of about 3–8% (i.e. bisexual, ambidextrous, dyslexia, or no dyslexia is unclear). Of course, this proves nothing, except that dichotomous traits with one subgroup being much larger than the other, are common in the brain. Despite decades of research, it is still unresolved what causes dichotomous traits, such as hand and sexual preferences, just like with dyslexia. However, let us not forget that brain research is still in its infancy. So far, we were looking at a large forest, hoping to see the trees in the middle of it.

If dyslexia is indeed a dichotomous condition, it also can be ruled out that dyslexia is a specific form of intelligence, such as spatial intelligence or arithmetic intelligence. All these specific forms of intelligence are abilities with normal distributions. Another possibility is that dyslexia is not a cognitive but a perceptual condition. Support for this can be found in Study 2 (Chapter 3) and Study 4 (Chapter 5). In these studies, we found various well-known cognitive variables that are often impaired in people with dyslexia, such as spelling and reading abilities and short-term memory. However, other variables – such as phonological abilities, automatised rapid naming, attention, and confusion – may not represent cognitive but more perceptual processes (see Chapter 4).

So far, I have argued that dyslexia is probably a dichotomous perceptual condition without knowing what causes dyslexia. One thing all researchers agree on is that dyslexia has a genetic origin. The offspring of a parent with dyslexia will, depending on the cut-off, have a chance of about 30–65% to have dyslexia as well (Snowling, & Melby-Lervåg, 2016). If dyslexia is genetically determined, it must already exist for a long time, much longer than when people started to write and read. This has been stated by other researchers as well, however, without drawing any clear conclusions. Taken one step further, it can be argued that dyslexia is not a disorder at all. Dyslexia is a genetic condition with a relatively high prevalence for a disorder. One could argue that anything that results in a disadvantage, whether it is impaired reading or slow naming of numbers, should be considered as a disorder. Regarding homosexuality, however, most scholars would not argue in the same way, although the natural disadvantage of being homosexual is clear (unlikely to produce offspring). As long as we do not know what causes dyslexia, prevalence should go to considering dyslexia as a genetic variation, resulting in a perceptual dichotomy. This variation has disadvantages but may just as well have advantages. The disadvantages arose only when people started to write and read, so evolution had no reasons to let it disappear.

For years, I have been intrigued by a study of Lachmann & Van Leeuwen (2007). They introduced a Functional Coordination Deficit Model, proposing that dyslexia reflects the inability to suppress symmetry of letters. They suggested that only people without dyslexia learned to suppress symmetry. This might be an explanation for confusion between the letters [p] and [b]. People with dyslexia may tend to perceive these letters as the same, which can be beneficial in certain tasks. The results of this study are consistent with dyslexia as a genetic, dichotomous and perceptual condition that has disadvantages and advantages. When I spoke with a colleague about this (Kretzschmar, 2017), he proposed, as a joke, that if this were true, dyslexia may never have been a genetic adaptation, but instead, the adaptation was evolving away from dyslexia. By now, I think this joke might be true. People with dyslexia are the original people and getting rid of dyslexia was a variation with benefits for not getting confused. It would offer a perfect explanation for the existence of the variable confusion as posited in this thesis, which as far as I know is for the first time in the literature.

Perception and cognition evolved thousands of years ago in ways we do not understand, and we can only have an educated guess about it. A possible educated guess is that, alongside with cognitive development of intelligence, a variation evolved that made it possible to perceive small details and symbols without getting confused. That not all people acquired this ability was not a large disadvantage for group survival. Only centuries later, these people have difficulties in the complex and specified abilities that are, for instance, needed to learn to write and read.

The fifth theoretical proposition is that dyslexia is a dichotomous, perceptual, and genetic variation that should not be considered as a disorder. It affects various cognitive processes, such as learning to write and read, but it is not part of cognition itself.

4. Theory of impaired inhibition and excitation

From knowledge to theory

In Chapter 3, I presented five propositions that can be summarised in one sentence: Dyslexia should not be considered as a disorder but as a dichotomous, perceptual, and genetic variation with an unknown cause in the subcortex, resulting in widespread effects on various perceptual and cognitive functions and displaying a high variability between individuals depending on intelligence, schooling, additional training, native language, social background, and motivation. Throughout this conclusion, I argued that all aspects of this hypothesis follow directly from the results of this thesis in combination with existing knowledge.

Although the purpose of this thesis was to study methods for diagnosing dyslexia, I hypothesised that it might also result in conclusions of theoretical importance. The above summary of conclusions is based on this thesis as a whole. In the separate studies though, we discussed various results with existing theories, putting forward various objections against some of them and suggesting various new interpretations. Here, I could give a summary of this. However, when I evaluated these objections and suggestions, I found support for something that is shared by all theories. Therefore, I decided to present it as a new theoretical perspective of findings; a new theory, so to say, that encompasses all theoretical conclusions of this thesis.

Any new theory of dyslexia though, as I argued above, must adhere to some general requirements. In the case of dyslexia, so many theories have been proposed; thus, it seems pointless to propose yet another theory that opposes aspects of other theories. A better solution is to use a top-down strategy by developing a theory that encompasses all other theories. At the same time, it is paramount not to speculate but to respect the limits of our knowledge (admissions of ignorance). We must also accept that a broad perspective may not be put into use easily, seeing that a broad perspective may not be falsifiable directly with the scientific tools available at present (temporary mystification). The only way to test whether any theory of dyslexia is correct is to combine the top-down strategy with the bottom-up strategy by verifying that a new theory is consistent with all available data.

According to my conclusions, a theory of dyslexia must be a perceptual variation that originates in the subcortex, for instance, in the areas that were found to be involved in dyslexia: the thalamus and the caudate nucleus. The thalamus is usually described as a relay station where sensory information is filtered and transferred to the cerebral cortex, received back from the cerebral cortex and sent back again, and so on. As was described in Study 5 (Chapter 6), the caudate nucleus is implicated in several executive control functions and language functions, such as language switching in bilinguals, second language learning, and suppression of irrelevant words. A full quote from Study 5 (Chapter 6) will explain it best: *'In relation to these findings, the relation between the caudate nucleus and the test Dutch-English Rhyme Words, which requires fast switching between languages, makes perfect sense. This is also the case with the confusion reported via self-report questions, which represented typical mistakes, such as exchanging letters within words and exchanging words within sentences – mistakes that might result from impaired cognitive control of attention. We think that the possibility exists that reduced grey matter volume in the caudate nucleus in relation to rhyme/confusion might represent a more fundamental dysfunction of dyslexic people, which might encompass various difficulties of confusion, such as exchanging letters or words.'*

It is not perception itself that is altered in dyslexia, but it is the higher processing of incoming sensory information, whether it is visual or auditory. The involved key functions are as follows: control of attention, parallel processing, switching, suppression, inhibition, and excitation. Suppression of symmetry in the study of Lachman and Van Leeuwen (2007) is fully consistent with these key functions. More in general, the perceptual variation might have something to do with both suppression or inhibition and excitation of all kinds of sensory information. People with this variation in perception, such as people with dyslexia, may have a different way of looking and listening at things. When irrelevant information is not suppressed, these people have more information to process at the same time, thus having a more global way of perceiving things. For example, this quality might be beneficial for architects. However, this global view has disadvantages when cognitive tasks are performed, such as reading, which require fast switching between suppression and excitation and between relevant and irrelevant information. Reading requires the processing of so much sensory information at the same time; therefore, people with dyslexia get confused about what is relevant and what is irrelevant.

The theory I propose is that, originally, people perceived sensory information in a global way. Alongside cognitive development, a genetic variation benefitted a more specific way of processing information. This variation is related to inhibition and excitation of relevant versus irrelevant sensory information. The variation involves a sort of on and off button in the brain. This has widespread effects on various perceptual and cognitive functions. Nowadays, advanced cognitive functions are required for, for instance, reading. Nature was not aware that people were ever going to do something that is so complicated. Thus, unfortunately, in modern times, disadvantages are more prominent than advantages.

This theory might explain all other theories of dyslexia because it focuses on more fundamental traits. It is not necessary at this stage to specify where exactly in the subcortex this perceptual variation is nested. It is not possible to describe all details of this variation in this conclusion. It is a theory that comes forward from the results in this thesis. However, as stated above, it is required to explain existing data from this thesis and all other research available, both theories and symptoms. This will be done extensively in a separate manuscript. Here, we only highlight the most important findings regarding the theory of inhibition and excitation and existing theories of dyslexia.

In one sentence, the new theory can be summarised as follows: Dyslexia is a dichotomous, genetic variation, originating in the subcortex, which involves higher processing of sensory information through an on and off-button (inhibition and excitation) that is crucial for the control of attention, which has widespread effects on various perceptual and cognitive functions, with high variability between individuals depending on intelligence, schooling, additional training, native language, social background, and motivation.

Testing the theory of inhibition and excitation

Dyslexia as a language disorder

According to most definitions, dyslexia is a language disorder, predominantly in the domain of reading. However, the exact nature of the language disorder has never been described. Reading is an extremely complicated process that involves many separate cognitive and perceptual functions. A disorder in only one of them can easily disturb the whole learning process of reading.

At a very early stage of cognitive development (about five or six years), the human brain receives the assignment to learn reading, while during evolution, the human brain never had the opportunity to prepare itself for such a complex task. Mastering this ability requires years of practice, while general cognition has hardly been developed. During this period the brain develops a reading network based on many separate functions that were originally meant for entirely different tasks. While spoken language had a chance to develop in the brain on an evolutionary scale, resulting in various specified areas (e.g. Broca’s area and Wernicke’s area), there is no specific reading area in the brain.

It is paramount to realise that the first developmental stages of reading involve visual processes. Large amounts of letters, words, and sentences enter the brain as visual sensory information with high speed. Quick decisions must be made, which requires a perfect control of attention. Which letters, words, and sentences are crucial for understanding a text? When irrelevant letters or words are not inhibited, confusion arises: letters and words get mixed up and letters and words that resemble the ones already processed are added, replaced or forgotten. If there is no working control mechanism, the brain gets overloaded with sensory information. To make things even more confusing, sometimes identical letters represent different sounds in different words, and visually resembling words can have a completely different pronunciation. On the other hand, the same sound can be represented by different letters, and words that sound similar can have completely different writing. This is probably the reason that English is considered to be the most difficult language to learn for people with dyslexia.

A more specific explanation can be derived from studies of noise exclusion and visual crowding. Various deficits of noise exclusion are reported in the literature (Benassi et al., 2010; Sperling et al., 2005; Sperling et al., 2006). This is fully consistent with a lack of control due to too little inhibition and too much excitation. Noise is not inhibited, but it remains active or even becomes more excited, which leads to interference with relevant stimuli. The same mechanism may be happening during visual crowding. Visual crowding is a form of masking, in which single-letter identification is compromised by the presence of additional letters or other simple visual forms in close proximity (e.g. Crutch & Warrington, 2007; Levi, 2008; Pelli & Tillman, 2008). In one study, 60% of dyslexics' slowness in word analysis was explained by crowding effects (Martelli et al., 2009). Visual crowding is a key process of reading. During reading, foci of attention quickly shift from point to point with regular intervals, and a so-called 'crowding window' moves alongside these foci. This crowding window is an area around the focus of attention, in which everything can be identified. A disturbance of normal shifting between inhibition and excitation may have various effects on the crowding window. It might become larger, resulting in overlap between succeeding windows. Also, the distance between succeeding foci of attention may be smaller or larger, resulting in overlap or skipping letters or words. Any disturbance of these processes may result in impaired control of attention and, thus, in slow and inaccurate reading.

In summary, it is important to realise that when people with dyslexia are learning to read, this not only involves the development of phonological abilities but also visual abilities. If certain perceptual processes are different in people with dyslexia, these processes may affect both visual and phonological abilities required for reading.

Phonological deficit theory

The phonological deficit theory is by far the most frequently proposed to be a causal theory. This theory posits that dyslexia is caused by impairments in phonological information processing, probably caused by problems in the access to or fuzziness of phonological representations of words in the phonological lexicon (e.g. Vellutio et al., 2004; Shaywitz & Shaywitz, 2005; Snowling & Hulme, 2005). An important finding is that, in an fMRI study (Boets et al.2013), no differences between people with and without dyslexia were found regarding phonetic representations, but only deficient *access* to phonological representations. This clearly supports the idea that there is something going on during the processing of phonological information, which is an idea also suggested by Blomert, Mitterer and Paffen (2004).

Explaining the phonological deficit theory with an on and off button requires acknowledgement of the many forms of resemblance that exist in languages. Many letters and words resemble other letters and words, either visually or auditorily, depending on variety between languages. So, what is happening? When you read a word, a message is sent to the phonological lexicon. In that lexicon, various sounds are activated that may represent the correct pronunciation of the word and are sent back to, for instance, the working memory. Then, the incorrect pronunciations must quickly be inhibited, and the correct one activated. When this inhibition is too slow or even absent, confusion arises about the correct pronunciation. With letters, an additional problem can arise. If the symmetry of letters is not correctly inhibited, then reading the letter [p], for instance, may activate all the sounds of [p], [b], [d], and [q] simultaneously. Deciding which one is the correct one then becomes exponentially more complex. Sometimes an identical letter can have different pronunciations, depending on the word; however, different letters can also share the same pronunciation. These processes are probably far more complex than we can imagine, but clearly, any lack of control of inhibition and activation will affect the entire process.

According to the allophonic speech perception theory (Serniclaes et al., 2001; Serniclaes et al., 2004; Bogliotti et al., 2008; Dufor et al., 2009), people with dyslexia are characterised by poor discrimination between phoneme categories. According to this theory, people with dyslexia maintain a higher sensitivity to acoustic differences *within* one phoneme category. For example, the letter [p] is perceived by dyslexics as a wide range of sounds that overlaps the range of sounds of the letter [b]. Normally, young children lose this sensitivity, which leads to categorical perception of phonemes. However, this may not happen to people with dyslexia; thus, as a result, they remain more sensitive to differences within one phoneme category. They perceive allophones, which are contextual variants of phonemes. Somebody with dyslexia, who perceives an allophone, will have problems with attributing the same written symbol to sounds of different categories. These findings are easily explained by poor inhibition. As a child, people with dyslexia did not learn to inhibit the overlapping sounds between letters. This may have contributed to difficulties in the pronunciation of the correct letters in complex words, or to activation of the wrong letters when reading a letter with a resembling sound, as is the case with the [p] and the [b].

In short, poor inhibition affects all reading and speaking processes that involve both visual and auditory resemblance. This view explains more symptoms of dyslexia than the phonological deficit theory can. For instance, confusion between the letters [p] and [b] can be attributed to a phonological deficit, but visual confusion between letters cannot be explained by this theory. As Greek school teachers told me, in the Greek language a typical confusion of people with dyslexia is between the number ‘3’ (pronounced as /tria/) and the letter epsilon ‘ε’ (roughly pronounced as the ‘e’ in the English word *bet*). In Greek, there is no auditory resemblance at all. This can be explained by poor inhibition of symmetry, but not by the phonological deficit theory.

Impaired automatized rapid naming

Impaired automatised rapid naming is consistently reported as one of the main symptoms of dyslexia (e.g. Jones, Branigan, & Kelly, 2009), especially regarding the naming of letters and numbers but also of pictures and colours. How to explain this with the on and off-button? What happens during rapid naming? Five different letters (or numbers, pictures, or colours) are presented on a paper in columns in a random order and must be read out aloud as quickly as possible without making mistakes. It has been suggested that impaired inhibition of preceding items causes a delay in serial rapid naming (Jones, Snowling, & Moll, 2016). This is consistent with the idea of the on and off-button.

What happens is the following: you see the first letter, and while you are pronouncing this, you look at the next letter. Still pronouncing the first letter, you must remember the second letter because your eye is already going to the third letter. Maybe, your eye is already going to the fourth letter because you can easily remember the second and third letter. However, you must pronounce the second letter before the third letter. Thus, you must inhibit the third letter temporarily and activate the second letter, while reading the fourth letter. After that, the second letter must be forgotten and the third letter, which is still held in memory, must be activated. Meanwhile, your eye has already progressed to the fifth and sixth letter, which must be temporarily inhibited again. This clearly requires a perfect control of attention to the right letter for pronunciation. It is a process of fast switching between inhibition and excitation. One small mistake of inhibition or excitation leads to an error or total confusion.

Impaired short-term memory

Another well-known symptom of dyslexia is impaired short-term memory, usually measured with a test of digit span. Digit span tests mostly consist of numbers with an increasing number of digits. All must be memorised and pronounced in the correct order. To my knowledge, there are no reports of poor long-term memory or poor short-term memory for episodic events. Therefore, the question arises whether there is something wrong with the memory or with something else. Numbers may trigger all kinds of associations with, for instance, birth dates, age of a sister, and so on. All of this must be inhibited. During the repeating of the digits, there are also processes of inhibition and excitation, which are required to keep track of the correct order of the digits. In fact, a study of Hachmann et al. (2014) showed that people with dyslexia tend to remember digits in the incorrect order; however, they do not perform poorer than other people in tests that require remembering digits without the need of a specific order.

Explanation of other symptoms

Evaluating all symptoms of dyslexia with the on and off-button would require a separate thesis. There are many symptoms that must be explained by this button. For instance, poor concentration may be explained by poor inhibition of surrounding noise. Pianists with dyslexia experience difficulties in reading music notes on two staffs at the same time. This might be explained by poor inhibition and excitation during quick shifting between the staffs because notes in the same position on separate staffs denote a different meaning. There is lot more to say about this theory and the symptoms that were evaluated shortly in the above paragraphs.

5. Conclusion

The main conclusion of this thesis is that the study of the methodology of diagnosing dyslexia provided both suggestions for improvements as well as several new theoretical perspectives. Beforehand, the assumption was that diagnoses can only be improved without any preference for a specific theory of dyslexia. This was supported by the results of this thesis. Based on these results, we developed a new theoretical idea. The main reason for this was to show that a theoretical idea that does not exclude other theories will improve the explanations of every complex issue that accompanies the diagnoses of dyslexia.

The final definition of dyslexia, based on this thesis, is as follows. Dyslexia is a dichotomous, genetic variation, originating in the subcortex, involving higher processing of sensory information through an on and off-button (i.e. inhibition and excitation) that is crucial for the control of attention, causing widespread effects on various perceptual and cognitive functions and displaying high variability between individuals, depending on their intelligence, schooling, additional training, native language, social background, and motivation.

This definition explains why self-report statements are more successful in diagnosing dyslexia than tests. Self-report statements address the relative abilities of a large variety of cognitive demands, considering that people with dyslexia vary on various other characteristics, such as intelligence. This definition also explains the relation between the findings of tests and self-report statements with anatomical variations in the brain.

The main characteristic of this definition is that it can be a basis for more improvements in diagnosing dyslexia. The relevance of this new theory of dyslexia in the current thesis is to show that it is thinkable that there is a common cause that encompasses all existing theories, because it is the only way to understand the results of this thesis. Dyslexia was once a perceptual variation that, in the modern human being, initially without intention, causes various difficulties during cognitive learning that vary between individuals to a large extent. If diagnoses of dyslexia are only based on a few specific impairments, they will always lead to misdiagnoses. Only if diagnoses of dyslexia are based on the new broad view that was presented in this thesis, will they be independent from any differences in intelligence, schooling, or between languages.

9

Nederlandse samenvatting

Het doel van dit proefschrift was om te bepalen wat de beste methode is om dyslexie te diagnosticeren en hoe bestaande methodes verbeterd kunnen worden. Het stellen van een diagnose dyslexie wordt bemoeilijkt door de vraag wat dyslexie precies is en waardoor het wordt veroorzaakt. Daarover bestaat weinig consensus. De enige consensus is dat dyslexie erfelijk is en dat het allerlei effecten heeft op het aanleren van en gebruiken van geschreven taal. De afgelopen decennia zijn er veel symptomen beschreven en veel theorieën voorgesteld, maar geen enkele theorie heeft tot dusver alle symptomen kunnen verklaren (Zie hoofdstuk 1). In grote lijnen zijn er op dit moment twee belangrijke stromingen. De eerste stroming behelst het verklaren van dyslexie vanuit een enkelvoudige oorzaak. De tweede stroming behelst het idee dat dyslexie wordt veroorzaakt door de optelsom van een aantal kleine zwakheden die tezamen tot taalproblemen leiden. Omdat er zoveel verschillende visies over dyslexie zijn, is bij de onderzoeken in dit proefschrift geen standpunt ingenomen over de aard en de oorzaak van dyslexie. Een belangrijk uitgangspunt is wel geweest dat de studie van diagnostiek mogelijk kan bijdragen aan theoretische inzichten, en daarmee weer aan het verbeteren van de diagnostiek.

Dyslexie kan op drie verschillende manieren worden vastgesteld: met tests, met vragenlijsten en door te kijken naar hersenstructuren. Bestaande diagnostische procedures zijn voor het grootste deel gebaseerd op testresultaten, en vooral testresultaten van tests die diverse taalvaardigheden meten. Maar daar kleven wel wat nadelen aan. Ten eerste is het nog geen uitgemaakte zaak dat dyslexie uitsluitend een taalstoornis is. Ten tweede kunnen de resultaten van taaltests moeilijk vergeleken worden tussen diverse talen, omdat typische dyslectische moeilijkheden per taal zeer kunnen verschillen. Ten derde is het heel moeilijk om effecten door algemene intelligentie te scheiden van effecten door dyslexie. Ten vierde is het heel moeilijk om een zogeheten criterium van dyslexie vast te stellen. Om te bepalen of iemand dyslexie heeft, moet vastgesteld worden dat zo iemand bepaalde lage prestaties heeft die in overeenstemming zijn met de lage prestaties van andere mensen met dyslexie. Maar het valt niet met absolute zekerheid vast te stellen wie dyslexie heeft zolang dyslexie niet eenduidig is gedefinieerd. Er bestaan dus geen absolute internationale normen voor prestaties op dyslexietests.

In de eerste studie (hoofdstuk 2) zijn twee statistische technieken (discriminantanalyse en logistische regressieanalyse) gebruikt om in een groep van 495 studenten vast te stellen wie dyslectisch is en wie niet. Bij deze technieken worden altijd twee criteriumgroepen gebruikt: een groep personen van wie zeker is of ze dyslectisch zijn en een groep personen van wie zeker is dat ze niet dyslectisch zijn. Voor het criterium dyslectisch wordt doorgaans gebruik gemaakt van de zogeheten dyslexieverklaring, een officieel document dat kinderen kunnen krijgen na uitvoerig onderzoek. Voor het criterium niet-dyslectisch wordt doorgaans een groep studenten geselecteerd die geen enkel probleem hebben met taal. De analyses vergelijken dan de testprestaties van de overige studenten met die van de twee criteriumgroepen en delen vervolgens deze studenten in als dyslectisch of niet-dyslectisch. Een nadeel is dat het criterium dyslectisch mogelijk alleen de zware gevallen betreft en het criterium niet-dyslectisch alleen zeer goed presterende studenten. Wij hebben in deze studie drie nieuwe dingen uitgeprobeerd. Ten eerste hebben we nieuwe vragenlijsten en tests gemaakt waaronder ook vragen en tests die niet alleen afhangen van talige prestaties. Ten tweede hebben we gebruik gemaakt van een biografisch en een flexibel criterium. Allereerst hebben we het hebben van een dyslexieverklaring vergeleken met andere biografische gegevens zoals een zelfoordeel over dyslexie, een zelfoordeel over spellingsvaardigheden, het al dan niet voorkomen van dyslexie in de familie, enzovoorts. Vervolgens hebben we na een eerste ronde van analyses het criterium aangepast en met de nieuwe criteria de analyses opnieuw uitgevoerd. Deze procedure hebben we meerdere keren herhaald. Ten derde hebben we in deze studie is niet alleen prestaties op gehele tests en vragenlijsten geanalyseerd (somscores), maar ook prestaties op items van tests en vragenlijsten. De gedachte hierachter was dat somscores van tests in hoge mate worden beïnvloed door bijvoorbeeld intelligentie, terwijl dyslexie zich kenmerkt door specifieke fouten die beter aan het licht komen in afzonderlijke test- of vragenlijstitems. De eerste conclusie van deze studie was dat vragen of testitems inderdaad tot betere diagnoses leiden dan somscores. De tweede conclusie was dat diagnoses betrouwbaarder zijn als in een criterium niet alleen de zware gevallen van dyslexie en super-presterende studenten zijn opgenomen, maar ook de minder zware gevallen en meer gemiddeld presterende studenten zonder dyslexie. De conclusies uit deze studie zijn nader onderzocht in de drie hierop volgende studies.

In de tweede studie (hoofdstuk 3) hebben we dezelfde dataset gebruikt als in de eerste studie. De vraag was welke cognitieve variabelen te onderscheiden zijn in de gehele set van tests en vragenlijsten en in hoeverre deze variabelen een rol spelen bij dyslexie. Met behulp van exploratieve en confirmatieve factoranalyses werd gevonden dat in zowel de tests als in de vragenlijsten vijf variabelen te onderscheiden zijn. De betrouwbaarheid van de analyses werd bevestigd doordat deze variabelen paarsgewijs zeer hoog met elkaar correleerden tussen de tests en de vragenlijsten. Zo konden we vijf algemene variabelen benoemen: (1) spelling, (2) fonologie, (3) korte termijn-geheugen, (4) verwarring / rijmvaardigheid en (5) complexiteit / woordverwerking. Dat al deze variabelen een rol spelen bij dyslexie werd bevestigd door de zeer significant lagere scores op deze variabelen van dyslectische studenten. Een tweede belangrijke bevinding in deze studie kwam voort uit een algemene dyslexiescore die gebaseerd was op een regressievergelijking met diverse items van tests en vragenlijsten die betrokken zijn bij diagnoses van dyslexie. Deze score bleek niet normaal verdeeld te zijn. Er konden twee verdelingen worden onderscheiden waarbij de verdeling van de studenten met dyslexie lager ligt (lagere scores) dan die van studenten zonder dyslexie (zie Hoodstuk 3, Figure 1). Dit duidt erop dat dyslexie niet zomaar beschouwd kan worden als het onderste deel van een normale verdeling onder een zekere grens, maar als iets wat echt bestaat als een bijzondere eigenschap. Verrassende bevindingen waren ook dat ruim 20 procent van de dyslectische studenten geen ernstige taalproblemen heeft, en dat geen enkele dyslectische student zwak is op alle vijf variabelen, wat erop duidt dat studenten met dyslexie mogelijk voor bepaalde zwakheden kunnen compenseren doordat zij bijvoorbeeld intelligent en hooggeschoold zijn.

In de derde studie (hoofdstuk 4) hebben we een nieuwe groep studenten onderzocht met een verkorte vragenlijst. Uitkomsten uit de eerste twee studies werden bevestigd. Ten eerste kan dyslexie heel betrouwbaar worden vastgesteld zonder gebruikmaking van tests en met uitsluitend biografische informatie als criterium. Niet meer dan 20 specifieke vragen zijn nodig om tot een diagnose te komen die minstens 90 procent betrouwbaar is. En opnieuw werden er vijf variabelen gevonden die ten grondslag liggen aan de vragen, die bovendien grote gelijkenis vertonen met die uit de tweede studie: spelling / lezen, fonologie, korte termijn-geheugen / aandacht, verwarring en complexe woorden. Een specifiek resultaat was dat vragen meer voorspellende waarde hebben naarmate ze gaan over heel specifieke eigenschappen. Dus niet: “Ik ben slecht in spelling”, maar bijvoorbeeld: “Ik haal vaal de ‘ij’ en ‘ei’ door elkaar”.

In de vierde studie (hoofdstuk 5) hebben we opnieuw een nieuwe groep studenten onderzocht. Dit keer hebben we niet alleen onze tests en vragenlijst afgenomen, maar ook een in Nederland algemeen gangbare testbatterij. Er waren drie belangrijke uitkomsten. Ten eerste werd bevestigd dat meerdere cognitieve variabelen een rol spelen bij dyslexie. In deze studie werd steun gevonden voor acht variabelen: spelling, lezen, fonologie, korte termijn-geheugen, verwarring, complexiteit, aandacht en snel benoemen. Ten tweede bleek dat beide testbatterijen (de gangbare en die van onszelf) dezelfde diagnostische waarde hadden (90%), maar de vragenlijst een veel betere (97%). Ten derde werd gevonden dat de voorspellende waarde toeneemt naarmate er meer cognitieve variabelen meegewogen worden. Dit bevestigt dat dyslectische studenten onderling heel verschillend kunnen zijn en dat er dus meerdere variabelen nodig zijn om dyslexie vast te stellen.

In de vijfde studie (hoofdstuk 6) is onderzocht in hoeverre dyslexie zichtbaar is in de anatomie van het brein. Hiervoor werd gebruik gemaakt van Magnetic Resonance Imaging (MRI). Een belangrijk aspect van deze studie was dat het gebruikte sample veel groter was dan gebruikelijk in MRI-studies (37 dyslectisch, 57 niet-dyslectisch). Toch werden er geen significante groepsverschillen gevonden in het brein. Toch hebben wij de gebieden met de grootste verschillen gerapporteerd, omdat er bij kleine samples veel onduidelijk is over de betekenis van niet-significant. Wel werden er diverse significante correlaties gevonden tussen enkele cognitieve variabelen (uit de tweede studie) en diverse gebieden in het brein. In het oog sprong hierbij een correlatie tussen verwarring en de caudate nucleus, iets wat niet eerder was gerapporteerd.

In de zesde studie (hoofdstuk 7) is met een selectie van het sample uit de vijfde studie (22 dyslectisch, 27 niet-dyslectisch) een relatief nieuwe techniek toegepast: classificatie door middel van Machine Learning. Dit was nooit eerder gedaan met betrekking tot dyslexie. Het resultaat was dat 80 procent van de studenten correct kon worden geclassificeerd met behulp van de anatomie van het brein. En wederom werden er diverse correlaties gevonden tussen de belangrijkste gebieden die betrokken waren bij de classificatie en diverse cognitieve variabelen. Kruisvalidatie naar een groot tweede sample was minder succesvol, wat vooral te wijten was aan het gebrek aan biografische informatie over deze groep.

De belangrijkste bevindingen van de zes studies tezamen kunnen als volgt worden samengevat. Ten eerste kunnen bestaande diagnostische methodes verbeterd worden door het gebruik van veel verschillende soorten tests aangevuld met een uitgebreide vragenlijst, door een beter gebruik van een criterium, door het gebruik van items of vragen in plaats van alleen somscores van tests, en door het herhalen van voorspellingen met verschillende criteria. Dat een zelf-rapportage van specifieke moeilijkheden zo goed dyslexie kan voorspellen is misschien onverwacht, maar niet vreemd. De prestaties op tests hangen niet alleen af van dyslexie, maar ook van intelligentie, scholing, motivatie, enzovoorts. Daardoor is het moeilijk om absolute normen vast te stellen die geldig zijn voor mensen van alle cognitieve niveaus. Bij vragenlijsten speelt dit allemaal niet zo’n grote rol. Als iemand zichzelf moet beoordelen, gebruikt hij of zij vooral zijn of haar eigen sociale omgeving als referentie. Daardoor zijn de antwoorden op vragen grotendeels onafhankelijk van mate van intelligentie of scholing.

Een tweede bevinding is dat dyslexie in verband brengen met de anatomie van het brein leidt tot resultaten waarvan de statistische significantie moeilijk is te beoordelen. MRI-onderzoek naar dyslexie is nog niet veel gedaan. We moeten het doen met minder dan twintig studies waarvan de meerderheid kleine samples heeft gebruikt. Dat een bevinding in het brein niet significant is, wil niet zeggen dat dit verschil niet bestaat. Dus is het voor de hand liggend om deze resultaten niet meteen te negeren. Wat we wel kunnen doen is de studies met kleine samples vergelijken met studies met grote samples. Maar het vreemde is dat juist de studies met grotere samples geen significante groepsverschillen rapporteren, terwijl de meeste significante groepsverschillen gevonden worden in de studies met de kleinste samples. Dit is tegen de verwachting in, maar in overeenstemming met diverse meta-analyses waaruit naar voren komt dat vrijwel geen significante groepsverschillen bestaan. Een mogelijke verklaring is dat de kleine samples relatief homogeen zijn omdat ze bijvoorbeeld verkregen zijn in een specifiek deel van de populatie, bijvoorbeeld op één school. Mogelijk kenmerken deze homogene samples zich door slechts enkele cognitieve zwaktes, terwijl voor dezelfde zwaktes gecompenseerd wordt. Met andere woorden: als een sample van dyslectische mensen groter is, bestaat deze uit meer verschillende soorten mensen met onderling verschillende kenmerken van dyslexie. Dat verklaart ook waarom er in onze studie geen groepsverschillen gevonden werden, maar wel correlaties tussen specifieke gebieden in het brein en specifieke cognitieve zwaktes.

Op grond van de bevindingen in dit proefschrift kunnen conclusies getrokken worden die van belang zijn voor de theorieën over dyslexie. Dit is ook hier van belang, omdat elke theoretische conclusie ook van belang is voor onderzoek naar diagnoses van dyslexie. Ik bespreek de implicaties van de bevindingen in dit proefschrift hieronder aan de hand van vijf theoretische stellingen.

Stelling 1 is dat onbekend is wat dyslexie veroorzaakt, maar dat duidelijk is dat dyslexie gekenmerkt wordt door vele symptomen. In de factoranalyses in de vierde studie bleek dat diverse cognitieve variabelen zich soms gedroegen als onafhankelijk van elkaar en soms afhankelijk. De meest strikte conclusie is dat de vele symptomen van dyslexie zich in elk geval openbaren als minimaal vijf cognitieve variabelen die op individueel niveau zeer verschillend kunnen zijn. Dit hangt mogelijk mede af van verschillen in leeftijd, intelligentie, scholing, extra training, moedertaal, sociale achtergrond, motivatie, enzovoorts. Door deze verschillen is het moeilijk om diagnoses te stellen en om algemene uitspraken te doen over de relatie met de anatomie van het brein.

Stelling 2 is dat dyslexie vermoedelijk een enkelvoudige oorzaak heeft. Dit staat haaks op de steeds meer steun krijgende hypothese dat dyslexie slechts de gedragsmatige uitkomst is (slecht lezen) van de optelsom van diverse onafhankelijke cognitieve zwaktes ofwel risicofactoren, de zogeheten Multiple Deficit Theory (MDT), voorgesteld door Pennington in 2006. Een argument dat soms gebruikt wordt tegen een enkelvoudige verklaring van dyslexie, is dat zo’n verklaring nog steeds niet gevonden is. Maar dat lijkt mij nogal voorbarig aangezien onze kennis van de werking van het cognitieve brein uiterst minimaal is. Niet een argument tegen de MDT, maar wel een merkwaardige bijkomstigheid van de MDT is dat iemand niet voor het leven gedefinieerd kan worden als zijnde dyslectisch of niet-dyslectisch. Volgens de MDT dient er een (arbitraire) grens getrokken te worden op een normale curve voor leesvaardigheid. Bijvoorbeeld een kind van negen jaar kan net onder die grens terecht komen en een dyslexieverklaring krijgen. Op grond daarvan gaat zo’n kind extra hard oefenen op taalvaardigheden. Een jaar later presteert zo’n kind dan opeens boven de grens en is daarmee ineens niet meer dyslectisch. Men kan zich afvragen of de MDT mogelijk betekent dat dyslexie niet bestaat, maar alleen zwakke leesprestaties. Maar hoe zit het dan met iemand die bijvoorbeeld op de lagere school erin is geslaagd goed te leren lezen door een hoge intelligentie, motivatie en extra hulp, maar opnieuw de mist in gaat tijdens lessen Engels en Grieks op de middelbare school? Een sterk argument voor het bestaan van een enkelvoudige oorzaak is de bevinding in dit proefschrift van een gesplitste normaalverdeling voor een dyslexiescore. Het bestaan hiervan zou in elk geval kunnen betekenen dat mensen voor altijd eenduidig gedefinieerd worden als dyslectisch of niet-dyslectisch. Extra training leidt dan slechts tot een opschuiven naar de rechterkant van de verdeling voor mensen met dyslexie, maar nog altijd zonder in de buurt te komen van de verdeling voor mensen zonder dyslexie. Een belangrijk argument voor de MDT is dat het eenvoudig is om de comorbiditeit tussen ADHD en dyslexie te verklaren. Maar een enkelvoudige oorzaak kan dat ook. Zwakke cognitieve aandacht is een kenmerk van zowel dyslexie als ADHD. In Nederland wordt zelfs de diagnose ADD gesteld: een aandachtstoornis zonder de bekende hyperactiviteit. Maar mogelijk is er bij deze mensen geen sprake van ADD, maar van dyslexie. Mogelijk hebben dyslexie en ADHD zelfs dezelfde enkelvoudige oorzaak, en is ADHD niet meer dan dyslexie in combinatie met een slechte concentratie en verhoogde hersenactiviteit.

Stelling 3 is dat de anatomische verschillen in het brein tussen verschillende mensen met dyslexie, zoals die gevonden zijn in de studies van dit proefschrift en andere studies, het resultaat zijn van verschillen in training. Dyslexie is in sterke mate genetisch bepaald. Maar daardoor is dyslexie al vanaf de geboorte aanwezig in het brein. Vanaf het allereerste moment dat kinderen met de nadelen hiervan geconfronteerd worden, zullen zij (onbewust) deze zwaktes proberen te compenseren. Afhankelijk van andere eigenschappen zoals intelligentie, zal deze compensatie op individueel niveau heel verschillend verlopen. In sommige MRI-studies is gevonden dat dyslectische mensen soms ook meer grijze stof hebben in een bepaald gebied dan mensen zonder dyslexie. Dit kan duiden op compensatie. Mogelijk zijn alle gevonden verschillen in alle studies slechts het resultaat van sterke verschillen in training. Dit wordt bevestigd door de bevindingen in het cerebellum, een gebied dat betrokken is bij het automatiseren van cognitieve vaardigheden. In relatief veel studies worden verschillen gevonden in het cerebellum, maar steeds weer op een andere plek, en soms met minder grijze stof voor mensen met dyslexie, soms juist met meer grijze stof.

Stelling 4 is dat de oorsprong van dyslexie gezocht moet worden in de subcortex. De bevinding van de caudate nucleus in de vijfde studie ondersteunt dit. Anatomische verschillen zijn in allerlei gebieden van de cortex gevonden, maar niet in de subcortex. Pas na het verschijnen van de studies van dit proefschrift hebben ook andere studies dyslexie in verband gebracht met gebieden in de subcortex: behalve de caudate, ook de aangelegen thalamus. Deze gebieden zijn o.a. betrokken bij diverse hogere functies die verband houden met informatieverwerking. Omdat de subcortex allerlei verbindingen heeft met de omliggende cortex, is het een plausibele gedachte dat slechts een klein, bijna onmeetbaar verschil in de subcortex grote gevolgen kan hebben voor het functioneren van diverse gebieden in de cortex.

Stelling 5 is dat dyslexie bestaat, maar niet als een stoornis, maar als een dichotome perceptuele variatie. Het idee dat dyslexie een stoornis is, is eigenlijk vreemd. De prevalentieschattingen van dyslexie zijn de laatste jaren explosief gestegen van drie procent naar in sommige studies zelfs zo’n 18 procent. Die recente schattingen zijn eigenlijk te hoog om van een stoornis te spreken. Homoseksualiteit en linkshandigheid zijn ook geen stoornis, ook al kan beargumenteerd worden dat er allerlei nadelen kleven aan linkshandigheid en homoseksualiteit (beperkt vermogen om kinderen te krijgen). Ook in de studies van dit proefschrift schommelde de prevalentie van dyslexie rond de 15 procent. Dat dyslexie een perceptuele variatie zou kunnen zijn wordt ondersteund door de bevindingen in de subcortex en door het feit dat dyslexie zich nog steeds onafhankelijk van algemene intelligentie openbaart. In de eerste studie werd bijvoorbeeld gevonden dat studenten met en zonder dyslexie niet verschillen op schoolcijfers wiskunde, maar wel op schoolcijfers voor talige vakken, met name Engels. Dat dyslexie mogelijk te maken heeft met perceptie volgt uit een studie van Lachmann & Van Leeuwen (2007). Hierin werd voorgesteld dat mensen met dyslexie bijvoorbeeld in de war raken met gelijkende letters doordat zij er niet in slagen de symmetrie die bestaat tussen bepaalde letters te onderdrukken, bijvoorbeeld de [b] en de [p]. Deze onderzoekers stelden dat mensen met dyslexie dit nooit hadden geleerd tijdens de vroege ontwikkeling. Als je het zo bekijkt zou het allemaal zelfs andersom kunnen zijn: niet dyslexie is een variatie, maar het hebben van geen dyslexie (Alexander Kretzschmar). Oorspronkelijk was symmetrieonderdrukking helemaal niet nodig, maar in de loop van de cognitieve ontwikkeling van de mens is deze vaardigheid steeds belangrijker geworden. Zo ontstond de mens zonder dyslexie. Hoe het ook zij, perceptie en cognitie is geëvolueerd in de loop van duizenden jaren. Wij weten er nog vrijwel niks van af. Er is dus alle reden om alle mogelijke opties open te houden, van enkelvoudige tot meervoudige verklaringen.

Op grond van bovenstaande stellingen kan nu zelfs een nieuwe theorie geformuleerd worden. Die houdt in dat mensen met dyslexie prikkels van buitenaf ervaren op een globale manier. Om specifieke details waar te nemen is het van belang dat bepaalde gelijkende, maar irrelevante details worden onderdrukt. Mensen met dyslexie kunnen dit niet goed. Daardoor raken zij in de war. Processen zoals inhibitie en excitatie worden geregeld in gebieden zoals de thalamus en caudate nucleus. Het voordeel van deze theorie is dat een enkelvoudige oorzaak toch de grote verscheidenheid aan symptomen kan verklaren. Immers, inhibitie en excitatie spelen een rol juist in symptomen die aan dyslexie gerelateerd zijn. Bijvoorbeeld fonologische verwarring ontstaat niet in het fonologisch lexicon, zoals vroeger werd gedacht, maar juist in de banen van en naar het lexicon toe, waardoor bij een enkele fonologische prikkel mogelijk meerdere schrifttekens worden opgehaald waarvan de verkeerde schrifttekens slecht onderdrukt worden. Visuele verwarring wordt veroorzaakt door gelijkenissen tussen letters of tussen woorden. Slechte onderdrukking van irrelevante letters of woorden leidt tot het lezen van een ander woord. Juist ook bij vaardigheden waarbij snelheid een rol speelt neemt de verwarring toe door een slechte afwisseling van excitatie en inhibitie. Bijvoorbeeld snel benoemen van letters, of het onthouden van cijfers in de juiste volgorde is moeilijk als er voortdurend verwarring optreedt. Het lezen van muzieknoten is zo goed als onmogelijk: alle noten lijken op elkaar, en bij de piano zijn er zelfs twee balken.

Op grond van bovenstaande wordt deze conclusie beëindigd met een voorstel voor een nieuwe definitie van dyslexie. “Dyslexie is geen stoornis, maar een dichotome, genetische variatie met een onbekende oorzaak in de subcortex die verband houdt met hogere orde verwerking van sensorische informatie door een soort aan-uit-knop (inhibitie en excitatie) die cruciaal is voor de controle van aandacht, resulterend in wijdverspreide effecten op diverse perceptuele en cognitieve functies met een grote variëteit tussen individuen afhankelijk van intelligentie, scholing, extra training, moedertaal, sociale achtergrond, concentratie en motivatie.” Dat dit een basis moge zijn voor nader onderzoek en voor het mogelijk verbeteren van diagnostische procedures.

10
References

Ahissar, M. (2007). Dyslexia and the anchoring-deficit hypothesis. *Trends in Cognitive Sciences*, *11*(11), 458-465.

Ali, N., Green, D. W., Kherif, F., Devlin, J. T., & Price, C. J. (2010). The role of the left head of caudate in suppressing irrelevant words. *Journal of Cognitive Neuroscience*, *22*(10), 2369-2386.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Andersson, J. L., Jenkinson, M., & Smith, S. (2007a). Non-linear optimisation. FMRIB technical report TR07JA1. *Practice. 2007a Jun*.

Andersson, J. L., Jenkinson, M., & Smith, S. (2007b). Non-linear registration, aka Spatial normalisation FMRIB technical report TR07JA2. *FMRIB Analysis Group of the University of Oxford*, *2*, 1-21.

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, *105*(4), 678-723.

Ashburner, J., & Friston, K. J. (2000). Voxel-Based Morphometry—The Methods. *NeuroImage, 11*(6), 805-821.

Baddeley, A. (1986). Oxford psychology series, No. 11. Working memory. New York, NY, US.

Baddeley, A. D. (1990). *Human memory*. Boston, MA: Allyn & Bacon.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417-423.

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist*, *7*(2), 85-97.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189-208.

Baker, C. I., Liu, J., Wald, L. L., Kwong, K. K., Benner, T., & Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proceedings of the National Academy of Sciences*, *104*(21), 9087-9092.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*(6), 561-571.

Belson, W. A. (1986). Validity in survey research: with special reference to the techniques of intensive interviewing and progressive modification for testing and constructing difficult or sensitive measures for use in survey research: a report (Vol. 201). Gower Publishing Company, Limited.

Benassi, M., Simonelli, L., Giovagnoli, S., & Bolzani, R. (2010). Coherence motion perception in developmental dyslexia: A meta-analysis of behavioral studies. *Dyslexia*, *16*(4), 341-357.

Bjornsdottir, G., Halldorsson, J. G., Steinberg, S., Hansdottir, I., Kristjansson, K., Stefansson, H., & Stefansson, K. (2014). The adult reading history questionnaire (ARHQ) in Icelandic: Psychometric properties and factor structure. *Journal of Learning Disabilities*, *47*(6), 532-542.

Black, J. M., Xia, Z., & Hoeft, F. (2017). Neurobiological bases of reading disorder part II: The importance of developmental considerations in typical and atypical reading. *Language and Linguistics Compass*, *11*(10), e12252.

Blomert, L. (2005). Dyslexie in Nederland Theorie, Praktijk en Beleid. *Amsterdam: Uitgeverij Nieuwezijds*.

Blomert, L., Mitterer, H., & Paffen, C. (2004). In search of the auditory, phonetic, and/or phonological problems in dyslexia: Context effects in speech perception. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1030-1047.

Blomert, L., & Willems, G. (2010). Is there a causal link from a phonological awareness deficit to reading failure in children at familial risk for dyslexia?. *Dyslexia*, *16*(4), 300-317.

Boada, R., Willcutt, E. G., & Pennington, B. F. (2012). Understanding the comorbidity between dyslexia and attention-deficit/hyperactivity disorder. *Topics in Language Disorders*, *32*(3), 264-284.

Boets, B., de Beeck, H. P. O., Vandermosten, M., Scott, S. K., Gillebert, C. R., Mantini, D., ... & Ghesquière, P. (2013). Intact but less accessible phonetic representations in adults with dyslexia. *Science*, *342*(6163), 1251-1254.

Bogliotti, C., Serniclaes, W., Messaoud-Galusi, S., & Sprenger-Charolles, L. (2008). Discrimination of speech sounds by children with dyslexia: Comparisons with chronological age and reading level controls. *Journal of Experimental Child Psychology*, *101*(2), 137-155.

Bosse, M.-L., Tainturier, M.-J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition*, *104*(2), 198-230.

Bosse, M.-L., & Valdois, S. (2009). Influence of the visual attention span on child reading performance: a cross-sectional study. *Journal of Research in Reading*, *32*(2), 230-253.

Brambati, S. M., Termine, C., Ruffino, M., Stella, G., Fazio, F., Cappa, S. F., & Perani, D. (2004). Regional reductions of gray matter volume in familial dyslexia. *Neurology*, *63*(4), 742-745.

Brovelli, A., Nazarian, B., Meunier, M., & Boussaoud, D. (2011). Differential roles of caudate nucleus and putamen during instrumental learning. *NeuroImage*, *57*(4), 1580-1590.

Brown, W. E., Eliez, S., Menon, V., Rumsey, J. M., White, C. D., & Reiss, A. L. (2001). Preliminary evidence of widespread morphological variations of the brain in dyslexia. *Neurology, 56*(6), 781-783.

Brunswick, N. (2010). Unimpaired reading development and dyslexia across different languages. In *Reading and dyslexia in different orthographies* (pp. 149-172). Psychology Press.

Brus, B. T., & Voeten, M. J. M. (1979). Een-Minuut-Test [One Minute Test], Vorm A en B: Verantwoording en handleiding—Schoolvorderingstest voor de technische leesvaardigheid, bestemd voor het tweede tot en met het zesde leerjaar van het basisonderwijs. *Nijmegen: Berkhout*.

Buchholz, J., & Davies, A. A. (2005). Adults with dyslexia demonstrate space-based and object-based covert attention deficits: Shifting attention to the periphery and shifting attention between objects in the left visual field. *Brain and Cognition*, *57*(1), 30-34.

Callens, M., Tops, W., Stevens, M., & Brysbaert, M. (2014). An exploratory factor analysis of the cognitive functioning of first-year bachelor students with dyslexia. *Annals of Dyslexia, 64*(1), 91-119.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81-105.

Carrion-Castillo, A., Franke, B., & Fisher, S. E. (2013). Molecular genetics of dyslexia: an overview. *Dyslexia*, *19*(4), 214-240.

Casanova, M. F., El-Baz, A. S., Giedd, J., Rumsey, J. M., & Switala, A. E. (2010). Increased white matter gyral depth in dyslexia: Implications for corticocortical connectivity. *Journal of Autism and Developmental Disorders*, *40*(1), 21-29.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, *47*(2), 149-180.

Cavalli, E., Casalis, S., El Ahmadi, A., Zira, M., Poracchia-George, F., & Colé, P. (2016). Vocabulary skills are well developed in university students with dyslexia: Evidence from multiple case studies. *Research in Developmental Disabilities*, *51*, 89-102.

Chase, C., Dougherty, R. F., Ray, N., Fowler, S., & Stein, J. (2007). L/M speed-matching ratio predicts reading in children. *Optometry and Vision Science*, *84*(3), 229-236.

Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B., ... & Olson, R. K. (2013). Modeling the etiology of individual differences in early reading development: Evidence for strong genetic influences. *Scientific Studies of Reading*, *17*(5), 350-368.

Clark, K. A., Helland, T., Specht, K., Narr, K. L., Manis, F. R., Toga, A. W., & Hugdahl, K. (2014). Neuroanatomical precursors of dyslexia identified from pre-reading through to age 11. *Brain*, *137*(12), 3136-3141.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204-256.

Crutch, S. J., & Warrington, E. K. (2007). Foveal crowding in posterior cortical atrophy: a specific early-visual-processing deficit affecting word reading. *Cognitive Neuropsychology*, *24*(8), 843-866.

Danelli, L., Berlingeri, M., Bottini, G., Ferri, F., Vacchi, L., Sberna, M., & Paulesu, E. (2013). Neural intersections of the phonological, visual magnocellular and motor/cerebellar systems in normal readers: Implications for imaging studies on dyslexia. *Human Brain Mapping*, *34*(10), 2669-2687.

Darki, F., Peyrard-Janvid, M., Matsson, H., Kere, J., & Klingberg, T. (2012). Three dyslexia susceptibility genes, DYX1C1, DCDC2, and KIAA0319, affect temporo-parietal white matter structure. *Biological Psychiatry*, *72*(8), 671-676.

Dawkins, R. (2016). *The god delusion*. Random House.

Davis, R. D., & Braun, E. M. (2010). The Gift of Dyslexia, Revised and Expanded: Why Some of the Smartest People Can't Read... and How They Can Learn. Penguin.

de Groot, A. M. B., Dannenburg, L., & van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, *33*, 600-629.

de Jong, P. F. (1998). Working memory deficits of reading disabled children. *Journal of Experimental Child Psychology*, *70*(2), 75-96.

de Jong, P. F., & van Bergen, E. (2017). Issues in diagnosing dyslexia. Developmental Perspectives in Written Language and Literacy; Segers, E., van den Broek, P., Eds, 349-361.

de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology*, *91*(3), 450-476.

De Pessemier, P., & Andries, C. (2009). *GL&SCHR. Test voor Gevorderd Lezen & Schrijven.* Antwerpen, België: Garant.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15*(6), 254-262.

Di Filippo, G., & Zoccolotti, P. (2012). Separating global and specific factors in developmental dyslexia. *Child Neuropsychology*, *18*(4), 356-391.

Dufor, O., Serniclaes, W., Sprenger-Charolles, L., & Démonet, J. F. (2009). Left premotor cortex and allophonic speech perception in dyslexia: A PET study. *NeuroImage*, *46*(1), 241-248.

Ecker, C., Rocha-Rego, V., Johnston, P., Mourao-Miranda, J., Marquand, A., Daly, E. M., ... & MRC AIMS Consortium. (2010). Investigating the predictive value of whole-brain structural MR scans in autism: A pattern classification approach. *NeuroImage*, *49*(1), 44-56.

Eckert, M. A., Leonard, C. M., Wilke, M., Eckert, M., Richards, T., Richards, A., & Berninger, V. (2005). Anatomical signatures of dyslexia in children: Unique information from manual and voxel based morphometry brain measures. *Cortex*, *41*(3), 304-315.

Eden, G. F., & Vaidya, C. J. (2008). ADHD and developmental dyslexia. *Annals of the New York Academy of Sciences*, *1145*(1), 316-327.

Elliott, J. G., & Grigorenko, E. L. (2014). *The dyslexia debate* (No. 14). Cambridge University Press.

Evans, T. M., Flowers, D. L., Napoliello, E. M., & Eden, G. F. (2014). Sex-specific gray matter volume differences in females with developmental dyslexia. *Brain Structure and Function*, *219*(3), 1041-1054.

Facoetti, A., Trussardi, A. N., Ruffino, M., Lorusso, M. L., Cattaneo, C., Galli, R., ... & Zorzi, M. (2010). Multisensory spatial attention deficits are predictive of phonological decoding skills in developmental dyslexia. *Journal of Cognitive Neuroscience*, *22*(5), 1011-1025.

Friedmann, N., Dotan, D., & Rahamim, E. (2010). Is the visual analyzer orthographic-specific? Reading words and numbers in letter position dyslexia. *Cortex*, *46*(8), 982-1004.

Frye, R. E., Hasan, K., Xue, L., Strickland, D., Malmberg, B., Liederman, J., & Papanicolaou, A. (2008). Splenium microstructure is related to two dimensions of reading skill. *Neuroreport, 19*(16), 1627-1631.

Furlanetto, L. M., Mendlowicz, M. V., & Bueno, J. R. (2005). The validity of the Beck Depression Inventory-Short Form as a screening and diagnostic instrument for moderate and severe depression in medical inpatients. *Journal of Affective Disorders*, *86*(1), 87-91.

Gathercole, S. E., & Baddeley, A. D. (2014). *Working memory and language*. Psychology Press.

Gialluisi, A., Newbury, D. F., Wilcutt, E. G., Olson, R. K., DeFries, J. C., Brandler, W. M., ... & SLI Consortium. (2014). Genome-wide screening for DNA variants associated with reading and language traits. *Genes, Brain and Behavior, 13*(7), 686-701.

Glezer, L. S., Jiang, X., & Riesenhuber, M. (2009). Evidence for highly selective neuronal tuning to whole words in the “visual word form area”. *Neuron*, *62*(2), 199-204.

Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N., Fristen, K. J., & Frackowiak, R. S. (2002, June). A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage, 14*, 21-36.

Goswami, U. (2015). Visual attention span deficits and assessing causality in developmental dyslexia. *Nature Reviews Neuroscience*, *16*(4), 225-226.

Grahn, J. A., Parkinson, J. A., & Owen, A. M. (2008). The cognitive functions of the caudate nucleus. *Progress in Neurobiology*, *86*(3), 141-155.

Hachmann, W. M., Bogaerts, L., Szmalec, A., Woumans, E., Duyck, W., & Job, R. (2014). Short-term memory for order but not for item information is impaired in developmental dyslexia. *Annals of dyslexia*, *64*(2), 121-136.

Harrison, A. G., & Nichols, E. (2005). A validation of the Dyslexia Adult Screening Test (DAST) in a post-secondary population. *Journal of Research in Reading*, *28*(4), 423-434.

Hatcher, J., Snowling, M. J., & Griffiths, Y. M. (2002). Cognitive assessment of dyslexic students in higher education. *British Journal of Educational Psychology*, *72*(1), 119-133.

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., & Shakespeare, B. (2009). Speech perception abilities of adults with dyslexia: Is there any evidence for a true deficit?. *Journal of Speech, Language, and Hearing Research*, *52*(6), 1510-1529.

Hébert, S., Béland, R., Beckett, C., Cuddy, L. L., Peretz, I., & Wolforth, J. (2008). A case study of music and text dyslexia. *Music Perception: An Interdisciplinary Journal*, *25*(4), 369-381.

Heim, S., Tschierse, J., Amunts, K., Wilms, M., Vossel, S., Willmes, K., ... & Huber, W. (2008). Cognitive subtypes of dyslexia. *Acta Neurobiologiae Experimentalis*, *68*(1), 73-82.

Henderson, L., Barca, L., & Ellis, A. W. (2007). Interhemispheric cooperation and non-cooperation during word recognition: Evidence for callosal transfer dysfunction in dyslexic adults. *Brain and Language*, *103*(3), 276-291.

Henley, S. M. D., Ridgway, G. R., Scahill, R. I., Klöppel, S., Tabrizi, S. J., Fox, N. C., ... & EHDN Imaging Working Group. (2010). Pitfalls in the use of voxel-based morphometry as a biomarker: Examples from Huntington disease. *American Journal of Neuroradiology*, *31*(4), 711-719.

Hensler, B. S., Schatschneider, C., Taylor, J., & Wagner, R. K. (2010). Behavioral genetic approach to the study of dyslexia. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *31*(7), 525-532.

Hodge, S. M., Makris, N., Kennedy, D. N., Caviness, V. S., Howard, J., McGrath, L., ... & Harris, G. J. (2010). Cerebellum, language, and cognition in autism and specific language impairment. *Journal of Autism and Developmental Disorders*, *40*(3), 300-316.

Hoeft, F., McCandliss, B. D., Black, J. M., Gantman, A., Zakerani, N., Hulme, C., ... & Gabrieli, J. D. (2011). Neural systems predicting long-term outcome in dyslexia. *Proceedings of the National Academy of Sciences*, *108*(1), 361-366.

Hoeft, F., Meyler, A., Hernandez, A., Juel, C., Taylor-Hill, H., Martindale, J. L., ... & Deutsch, G. K. (2007). Functional and morphometric brain dissociation between dyslexia and reading ability. *Proceedings of the National Academy of Sciences*, *104*(10), 4234-4239.

Hosoda, C., Tanaka, K., Nariai, T., Honda, M., & Hanakawa, T. (2013). Dynamic neural network reorganization associated with second language vocabulary acquisition: A multimodal imaging study. *Journal of Neuroscience*, *33*(34), 13663-13672.

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The foundations of literacy development in children at familial risk of dyslexia. *Psychological Science*, *26*(12), 1877-1886.

International Dyslexia Association. (2002). The nature of learning disabilities. *IDA Policy*.

Jagger-Rickels, A. C., Kibby, M. Y., & Constance, J. M. (2018). Global gray matter morphometry differences between children with reading disability, ADHD, and comorbid reading disability/ADHD. *Brain and Language*, *185*, 54-66.

Jednoróg, K., Gawron, N., Marchewka, A., Heim, S., & Grabowska, A. (2014). Cognitive subtypes of dyslexia are characterized by distinct patterns of grey matter volume. *Brain Structure and Function*, *219*(5), 1697-1707.

Jednoróg, K., Marchewka, A., Altarelli, I., Monzalvo Lopez, A. K., van Ermingen-Marbach, M., Grande, M., ... & Ramus, F. (2015). How reliable are gray matter disruptions in specific reading disability across multiple countries and languages? Insights from a large-scale voxel-based morphometry study. *Human Brain Mapping*, *36*(5), 1741-1754.

Jones, M. W., Branigan, H. P., & Kelly, M. L. (2009). Dyslexic and nondyslexic reading fluency: Rapid automatized naming and the importance of continuous lists. *Psychonomic Bulletin & Review*, *16*(3), 567-572.

Jones, M. W., Snowling, M. J., & Moll, K. (2016). What automaticity deficit? Activation of lexical information by readers with dyslexia in a rapid automatized naming Stroop-switch task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(3), 465-474.

Kemp, N., Parrila, R. K., & Kirby, J. R. (2009). Phonological and orthographic spelling in high-functioning adult dyslexics. *Dyslexia*, *15*(2), 105-128.

Kere, J. (2011). Molecular genetics and molecular biology of dyslexia. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(4), 441-448.

Kershner, J. R. (2016). Forced-attention dichotic listening with university students with dyslexia: search for a core deficit. *Journal of Learning Disabilities*, *49*(3), 282-292.

Kibby, M. Y., Marks, W., Morgan, S., & Long, C. J. (2004). Specific impairment in developmental reading disabilities: A working memory approach. *Journal of Learning Disabilities*, *37*(4), 349-363.

Kirkby, J. A., Webster, L. A., Blythe, H. I., & Liversedge, S. P. (2008). Binocular coordination during reading and non-reading tasks. *Psychological Bulletin*, *134*(5), 742-763.

Kovas, Y., Haworth, C. M., Dale, P. S., & Plomin, R. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development*, *72*(3), vii-144.

Krafnick, A. J., Flowers, D. L., Napoliello, E. M., & Eden, G. F. (2011). Gray matter volume changes following reading intervention in dyslexic children. *NeuroImage*, *57*(3), 733-741.

Krafnick, A. J., Flowers, D. L., Luetje, M. M., Napoliello, E. M., & Eden, G. F. (2014). An investigation into the origin of anatomical differences in dyslexia. *Journal of Neuroscience, 34*(3), 901-908.

Kramer, D., & Vorst, H. C. M. (2007). *Communicatievragenlijst voor de opsporing van dyslectische leerlingen* [Communication questionnaire for the identification of dyslexic students] (Unpublished thesis). University of Amsterdam, Amsterdam, Netherlands.

Kretzschmar, A.E. (2017). *Dyslexia and the history of reading abilities in the human species.* Kepler, 186f.

Kronbichler, M., Wimmer, H., Staffen, W., Hutzler, F., Mair, A., & Ladurner, G. (2008). Developmental dyslexia: Gray matter abnormalities in the occipitotemporal cortex. *Human Brain Mapping*, *29*(5), 613-625.

Lachmann, T. (2002). Reading disability as a deficit in functional coordination. In *Basic functions of language, reading and reading disability* (pp. 165-198). Springer, Boston, MA.

Lachmann, T., & Geyer, T. (2003). Letter reversals in dyslexia: Is the case really closed? A critical review and conclusions. *Psychology Science*, *45*, 50-72.

Lachmann, T., & Van Leeuwen, C. (2007). Paradoxical enhancement of letter recognition in developmental dyslexia. *Developmental Neuropsychology, 31,* 61-77.

Le Jan, G., Le Bouquin-Jeannès, R., Costet, N., Trolès, N., Scalart, P., Pichancourt, D., Faucon, G., & Gombert, J.-E. (2011). Multivariate predictive model for dyslexia diagnosis. *Annals of Dyslexia*, *61*(1), 1-20.

Lefly, D. L., & Pennington, B. F. (2000). Reliability and validity of the Adult Reading History Questionnaire. *Journal of Learning Disabilities*, *33*(3), 286-296.

Leonard, C. M., Eckert, M. A., Lombardino, L. J., Oakland, T., Kranzler, J., Mohr, C. M., ... & Freeman, A. (2001). Anatomical risk factors for phonological dyslexia. *Cerebral Cortex*, *11*(2), 148-157.

Lervåg, A., & Hulme, C. (2009). Rapid automatized naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science*, *20*(8), 1040-1048.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635-654.

Lewis, S. J., Dove, A., Robbins, T. W., Barker, R. A., & Owen, A. M. (2004). Striatal contributions to working memory: A functional magnetic resonance imaging study in humans. *European Journal of Neuroscience*, *19*(3), 755-760.

Linkersdörfer, J., Lonnemann, J., Lindberg, S., Hasselhorn, M., & Fiebach, C. J. (2012). Grey matter alterations co-localize with functional abnormalities in developmental dyslexia: An ALE meta-analysis. *PloS one*, *7*(8), e43122.

Lobier, M., & Valdois, S. (2015). Visual attention deficits in developmental dyslexia cannot be ascribed solely to poor reading experience. *Nature Reviews Neuroscience*, *16*(4), 225.

Lobier, M., Zoubrinetzky, R., & Valdois, S. (2012). The visual attention span deficit in dyslexia is visual and not verbal. *Cortex*, *48*(6), 768-773.

Lorusso, M. L., Facoetti, A., Pesenti, S., Cattaneo, C., Molteni, M., & Geiger, G. (2004). Wider recognition in peripheral vision common to different subtypes of dyslexia. *Vision Research*, *44*(20), 2413-2424.

Luk, G., Green, D. W., Abutalebi, J., & Grady, C. (2012). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and Cognitive Processes*, *27*(10), 1479-1488.

Marinus, E., & de Jong, P. F. (2010). Variability in the word-reading performance of dyslexic readers: Effects of letter length, phoneme length and digraph presence. *Cortex*, *46*(10), 1259-1271.

Martelli, M., Di Filippo, G., Spinelli, D., & Zoccolotti, P. (2009). Crowding, reading, and developmental dyslexia. *Journal of Vision*, *9*(4):14, 1-18.

McArthur, G. M., & Bishop, D. V. M. (2001). Auditory perceptual processing in people with reading and oral language impairments: Current issues and recommendations. *Dyslexia*, *7*(3), 150-170.

McCarthy, J. H., Hogan, T. P., & Catts, H. W. (2012). Is weak oral language associated with poor spelling in school-age children with specific language impairment, dyslexia or both? *Clinical Linguistics & Phonetics*, *26*(9), 791-805.

Menghini, D., Finzi, A., Benassi, M., Bolzani, R., Facoetti, A., Giovagnoli, S., ... & Vicari, S. (2010). Different underlying neurocognitive deficits in developmental dyslexia: A comparative study. *Neuropsychologia*, *48*(4), 863-872.

Menghini, D., Hagberg, G. E., Petrosini, L., Bozzali, M., Macaluso, E., Caltagirone, C., & Vicari, S. (2008). Structural correlates of implicit learning deficits in subjects with developmental dyslexia. *Annals of the New York Academy of Sciences*, *1145*(1), 212-221.

Miles, T. R., Westcombe, J., & Ditchfield, D. (Eds.). (2008). *Music and dyslexia: A positive approach*. John Wiley & Sons.

Mortimore, T., & Crozier, W. R. (2006). Dyslexia and difficulties with study skills in higher education. *Studies in Higher Education*, *31*(2), 235-251.

Murdoch, B. E. (2010). The cerebellum and language: historical perspective and review. *Cortex*, *46*(7), 858-868.

Mwangi, B., Ebmeier, K. P., Matthews, K., & Steele, J. D. (2012). Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder. *Brain*, *135*(5), 1508-1521.

Nergård-Nilssen, T., & Hulme, C. (2014). Developmental dyslexia in adults: Behavioural manifestations and cognitive correlates. *Dyslexia*, *20*(3), 191-207.

Nicolson, R. I., & Fawcett, A. J. (1990). Automaticity: A new framework for dyslexia research? *Cognition*, *35*(2), 159-182.

Nicolson, R. I., & Fawcett, A. J. (1995). Dyslexia is more than a phonological disability. *DYSLEXIA-CHICHESTER-*, *1*, 19-36.

Nicolson, R. I., & Fawcett, A. J. (1997). Development of objective procedures for screening and assessment of dyslexic students in higher education. *Journal of Research in Reading*, *20*(1), 77-83.

Nicolson, R. I., & Fawcett, A. J. (1998). The dyslexia adult screening test. London: Psychological Corporation.

Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: reuniting the developmental disorders? *Trends in Neurosciences*, *30*(4), 135-141.

Nicolson, R. I., Fawcett, A. J., Brookes, R. L., & Needle, J. (2010). Procedural learning and dyslexia. *Dyslexia*, *16*(3), 194-212.

Odegard, T. N., Farris, E. A., Ring, J., McColl, R., & Black, J. (2009). Brain connectivity in non-reading impaired children and children diagnosed with developmental dyslexia. *Neuropsychologia*, *47*(8-9), 1972-1977.

Orton, S. T. (1925). Word-blindness in school children. *Archives of Neurology & Psychiatry*, *14*(5), 581-615.

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, *11*(10), 1129-1135.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *Cognition*, *101*(2), 385-413.

Pennington, B. F., Santerre–Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., ... & Olson, R. K. (2012). Individual prediction of dyslexia by single versus multiple deficit models. *Journal of Abnormal Psychology*, *121*(1), 212-242.

Pernet, C., Andersson, J., Paulesu, E., & Demonet, J. F. (2009a). When all hypotheses are right: a multifocal account of dyslexia. *Human Brain Mapping*, *30*(7), 2278-2292.

Pernet, C. R., Poline, J. B., Demonet, J. F., & Rousselet, G. A. (2009b). Brain classification reveals the right cerebellum as the best biomarker of dyslexia. *BMC Neuroscience, 10*(1), 67-86.

Peterson, R. L., & Pennington, B. F. (2015). Developmental dyslexia. *Annual Review of Clinical Psychology*, *11*, 283-307.

Peyrin, C., Lallier, M., Demonet, J. F., Pernet, C., Baciu, M., Le Bas, J. F., & Valdois, S. (2012). Neural dissociation of phonological and visual attention span disorders in developmental dyslexia: FMRI evidence from two case reports. *Brain and Language*, *120*(3), 381-394.

Pflugshaupt, T., Gutbrod, K., Wurtz, P., von Wartburg, R., Nyffeler, T., de Haan, B., ... & Mueri, R. M. (2009). About the role of visual field defects in pure alexia. *Brain*, *132*(7), 1907-1917.

Plomin, R., & Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin*, *131*(4), 592-617.

Płoński, P., Gradkowski, W., Altarelli, I., Monzalvo, K., van Ermingen-Marbach, M., Grande, M., ... & Jednoróg, K. (2017). Multi-parameter machine learning approach to the neuroanatomical basis of developmental dyslexia. *Human Brain Mapping*, *38*(2), 900-908.

Pohar, M., Blas, M., & Turk, S. (2004). Comparison of logistic regression and linear discriminant analysis: a simulation study. *Metodoloski Zvezki*, *1*(1), 143-161.

Poljac, E., Simon, S., Ringlever, L., Kalcik, D., Groen, W. B., Buitelaar, J. K., & Bekkering, H. (2010). Impaired task switching performance in children with dyslexia but not in children with autism. *The Quarterly Journal of Experimental Psychology*, *63*(2), 401-416.

Provost, J. S., Petrides, M., & Monchi, O. (2010). Dissociating the role of the caudate nucleus and dorsolateral prefrontal cortex in the monitoring of events within human working memory. *European Journal of Neuroscience*, *32*(5), 873-880.

Purvis, K. L., & Tannock, R. (1997). Language abilities in children with attention deficit hyperactivity disorder, reading disabilities, and normal controls. *Journal of Abnormal Child Psychology*, *25*(2), 133-144.

Rae, C., Harasty, J. A., Dzendrowskyj, T. E., Talcott, J. B., Simpson, J. M., Blamire, A. M., ... & Richardson, A. J. (2002). Cerebellar morphology in developmental dyslexia. *Neuropsychologia*, *40*(8), 1285-1292.

Ramus, F., & Ahissar, M. (2012). Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance. *Cognitive Neuropsychology*, *29*(1–2), 104-122.

Ramus, F., Altarelli, I., Jednoróg, K., Zhao, J., & di Covella, L. S. (2017). Neuroanatomy of developmental dyslexia: Pitfalls and promise. *Neuroscience & Biobehavioral Reviews, 84,* 434-452.

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain, 126*(4), 841-865.

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *The Quarterly Journal of Experimental Psychology*, *61*(1), 129-141.

Raschle, N. M., Chang, M., & Gaab, N. (2011). Structural brain alterations associated with dyslexia predate reading onset. *NeuroImage*, *57*(3), 742-749.

Raven, J. C., Court, J. H., & Raven, J. (1979). *A manual for Raven's Progressive Matrices and Vocabulary Tests*. London, UK: H. K. Lewis.

Reynolds, A. E., & Caravolas, M. (2016). Evaluation of the Bangor Dyslexia Test (BDT) for use with Adults. *Dyslexia*, *22*(1), 27-46.

Richlan, F., Kronbichler, M., & Wimmer, H. (2009). Functional abnormalities in the dyslexic brain: A quantitative meta-analysis of neuroimaging studies. *Human Brain Mapping, 30*(10), 3299-3308.

Richlan, F., Kronbichler, M., & Wimmer, H. (2011). Meta-analyzing brain dysfunctions in dyslexic children and adults. *Neuroimage*, *56*(3), 1735-1742.

Richlan, F., Kronbichler, M., & Wimmer, H. (2013). Structural abnormalities in the dyslexic brain: A meta-analysis of voxel-based morphometry studies. *Human Brain Mapping*, *34*(11), 3055-3065.

Romani, C., Tsouknida, E., di Betta, A. M., & Olson, A. (2011). Reduced attentional capacity, but normal processing speed and shifting of attention in developmental dyslexia: Evidence from a serial task. *Cortex*, *47*(6), 715-733.

Rouw, R., & Scholte, H. S. (2010). Neural basis of individual differences in synesthetic experiences. *Journal of Neuroscience*, *30*(18), 6205-6213.

Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L., Leach, M. O., & Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, *18*(8), 712-721.

Romani, C., Tsouknida, E., di Betta, A. M., & Olson, A. (2011). Reduced attentional capacity, but normal processing speed and shifting of attention in developmental dyslexia: Evidence from a serial task. *Cortex, 47*(6), 715-733.

Savage, R., Lavers, N., & Pillay, V. (2007). Working memory and reading difficulties: What we know and what we don't know about the relationship. *Educational Psychological Review, 19*(2), 185-221.

Saviour, P., Padakannaya, P., Nishanimutt, S., & Ramachandra, N. B. (2009). Familial patterns and biological markers of dyslexia. *International Journal of Human Genetics*, *9*(1), 21-29.

Scerri, T. S., Morris, A. P., Buckingham, L. L., Newbury, D. F., Miller, L. L., Monaco, A. P., ... & Paracchini, S. (2011). DCDC2, KIAA0319 and CMIP are associated with reading-related traits. *Biological Psychiatry*, *70*(3), 237-245.

Scerri, T. S., & Schulte-Körne, G. (2010). Genetics of developmental dyslexia. *European Child & Adolescent Psychiatry*, *19*(3), 179-197.

Schraeyen, K., Janssens, V., Aerts, A., Meersschaert, E., Baeyens, D., & Geudens, A. (2009). Diagnose van dyslexie bij jongvolwassenen: Validiteitsonderzoek van de Interactieve Dyslexietest Amsterdam-Antwerpen (IDAA).

Schulte-Körne, G., & Bruder, J. (2010). Clinical neurophysiology of visual and auditory processing in dyslexia: A review. *Clinical Neurophysiology*, *121*(11), 1794-1809.

Schulte-Körne, G., Deimel, W., & Remschmidt, H. (1997). Can self-report data on deficits in reading and spelling predict spelling disability as defined by psychometric tests? *Reading and Writing*, *9*(1), 55-63.

Schumacher, J., Hoffmann, P., Schmäl, C., Schulte-Körne, G., & Nöthen, M. M. (2007). Genetics of dyslexia: the evolving landscape. *Journal of Medical Genetics*, *44*(5), 289-297.

Schurz, M., Sturm, D., Richlan, F., Kronbichler, M., Ladurner, G., & Wimmer, H. (2010). A dual-route perspective on brain activation in response to visual words: Evidence for a length by lexicality interaction in the visual word form area (VWFA). *NeuroImage*, *49*(3), 2649-2661.

Serniclaes, W., Sprenger-Charolles, L., Carré, R., & Démonet, J. F. (2001). Perceptual discrimination of speech sounds in developmental dyslexia. *Journal of Speech, Language, and Hearing Research*, *44*(2), 384-399.

Serniclaes, W., van Heghe, S., Mousty, P., Carré, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology*, *87*(4), 336-361.

Shaywitz, S. E., & Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biological Psychiatry*, *57*(11), 1301-1309.

Siegel, L. S. (2008). Morphological awareness skills of English language learners and children with dyslexia. *Topics in Language Disorders*, *28*(1), 15-27.

Silani, G., Frith, U., Demonet, J. F., Fazio, F., Perani, D., Price, C., ... & Paulesu, E. (2005). Brain abnormalities underlying altered activation in dyslexia: A voxel based morphometry study. *Brain*, *128*(10), 2453-2461.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., ... & Niazy, R. K. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*(S1), 208-219.

Smith-Spark, J. H., Henry, L. A., Messer, D. J., Edvardsdottir, E., & Zięcik, A. P. (2016). Executive functions in adults with developmental dyslexia. *Research in Developmental Disabilities*, *53*, 323-341.

Snoek, L., Miletic, S., & Scholte, H. S. (2018). How to control for confounds in decoding analyses of neuroimaging data. *bioRxiv*, 290684.

Snowling, M. J., Dawes, P., Nash, H., & Hulme, C. (2012). Validity of a protocol for adult self-report of dyslexia and related difficulties. *Dyslexia, 18*(1), 1-15.

Snowling, M. J., & Hulme, C. (Eds.). (2005). *The science of reading: A handbook*. Oxford, UK: Blackwell.

Snowling, M. J., & Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological Bulletin*, *142*(5), 498-545.

Snowling, M. J., Nation, K., Moxham, P., Gallagher, A., & Frith, U. (1997). Phonological processing skills of dyslexic students in higher education: A preliminary report. *Journal of Research in Reading*, *20*(1), 31-41.

Soto, C. J., & John, O. P. (2009). Ten facet scales for the Big Five Inventory: Convergence with NEO PI-R facets, self-peer agreement, and discriminant validity. *Journal of Research in Personality*, *43*(1), 84-90.

Sperling, A. J., Lu, Z. L., & Manis, F. R. (2004). Slower implicit categorical learning in adult poor readers. *Annals of Dyslexia*, *54*(2), 281-303.

Sperling, A. J., Lu, Z. L., Manis, F. R., & Seidenberg, M. S. (2005). Deficits in perceptual noise exclusion in developmental dyslexia. *Nature Neuroscience*, *8*(7), 862-863.

Sperling, A. J., Lu, Z. L., Manis, F. R., & Seidenberg, M. S. (2006). Motion-perception deficits and reading impairment: It's the noise, not the motion. *Psychological Science*, *17*(12), 1047-1053.

Stein, J., & Walsh, V. (1997). To see but not to read: The magnocellular theory of dyslexia. *Trends in Neuroscience,* 20(4), 147-152.

Steinbrink, C., Vogt, K., Kastrup, A., Müller, H. P., Juengling, F. D., Kassubek, J., & Riecker, A. (2008). The contribution of white and gray matter differences to developmental dyslexia: Insights from DTI and VBM at 3.0 T. *Neuropsychologia*, *46*(13), 3170-3178.

Tabachnik, B. G., & Fidell, L. S. (2007). Using multivariate statistics (5th ed.). Boston: Pearson/Allyn & Bacon.

Tallal. P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language, 9*(2), 182-198.

Tallal, P., Miller, S., & Fitch, R.H. (1993). Neurological basis of speech: A case for the preeminence of temporal processing. *Annals of the New York Academy of Sciences, 682*(1), 27-47.

Tamboer, P., Scholte, H. S., & Vorst, H. C. M. (2015). Dyslexia and voxel-based morphometry: Correlations between five behavioural measures of dyslexia and gray and white matter volumes. *Annals of Dyslexia*, *65*(3), 121-141.

Tamboer, P., & Vorst, H. C. M. (2015). A new self-report inventory of dyslexia for students: Criterion and construct validity. *Dyslexia*, *21*(1), 1-34.

Tamboer, P., Vorst, H. C. M., & de Jong, P. F. (2017). Six factors of adult dyslexia assessed by cognitive tests and self-report questions: Very high predictive validity. *Research in Developmental Disabilities*, *71*, 143-168.

Tamboer, P., Vorst, H. C. M., Ghebreab, S., & Scholte, H. S. (2016). Machine learning and dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia. *NeuroImage: Clinical*, *11*, 508-514.

Tamboer, P., Vorst, H. C. M., & Oort, F. J. (2014). Identifying dyslexia in adults: An iterative method using the predictive value of item scores and self-report questions. *Annals of Dyslexia*, *64*(1), 34-56.

Tamboer, P., Vorst, H. C. M., & Oort, F. J. (2016). Five describing factors of dyslexia. *Journal of Learning Disabilities*, *49*(5), 466-483.

Tan, L. H., Chen, L., Yip, V., Chan, A. H. D., Yang, J., Gao, J. H., & Siok, W. T. (2011). Activity levels in the left hemisphere caudate–fusiform circuit predict how well a second language will be learned. *Proceedings of the National Academy of Sciences*, *108*(6), 2540-2544.

Tanaka, H., Black, J. M., Hulme, C., Stanley, L. M., Kesler, S. R., Whitfield-Gabrieli, S., ... & Hoeft, F. (2011). The brain basis of the phonological deficit in dyslexia is independent of IQ. *Psychological Science*, *22*(11), 1442-1451.

Terras, M. M., Thompson, L. C., & Minnis, H. (2009). Dyslexia and psycho-social functioning: An exploratory study of the role of self-esteem and understanding. *Dyslexia, 15*(4), 304-327.

Tops, W., Callens, M., Lammertyn, J., Van Hees, V., & Brysbaert, M. (2012). Identifying students with dyslexia in higher education. *Annals of Dyslexia*, *62*(3), 186-203.

Tunmer, W., & Greaney, K. (2010). Defining dyslexia. *Journal of Learning Disabilities, 43*(3), 229-243.

Vaessen, A., & Blomert, L. (2010). Long-term cognitive dynamics of fluent reading development. *Journal of Experimental Child Psychology*, *105*(3), 213-231.

Vaessen, A., Gerretsen, P., & Blomert, L. (2009). Naming problems do not reflect a second independent core deficit in dyslexia: Double deficits explored. *Journal of Experimental Child Psychology, 103*(2), 202-221.

van Bergen, E., Bishop, D., van Zuijen, T., & de Jong, P. F. (2015). How does parental reading influence children's reading? A study of cognitive mediation. *Scientific Studies of Reading*, *19*(5), 325-339.

van Bergen, E., de Jong, P. F., Maassen, B., Krikhaar, E., Plakas, A., & van der Leij, A. (2014). IQ of four-year-olds who go on to develop dyslexia. *Journal of Learning Disabilities*, *47*(5), 475-484.

van Bergen, E., van der Leij, A., & de Jong, P. F. (2014). The intergenerational multiple deficit model and the case of dyslexia. *Frontiers in Human Neuroscience*, *8*, 346.

van den Boer, M., van Bergen, E., & de Jong, P. F. (2015). The specific relation of visual attention span with reading and spelling in Dutch. *Learning and Individual Differences*, *39*, 141-149.

van den Bos, K. P., Lutje Spelberg, H. C., Scheepstra, A. J. M., & De Vries, J. R. (1994). De Klepel. Vorm A en B. Een test voor de leesvaardigheid van pseudowoorden. Verantwoording, handleiding, diagnostiek en behandeling. *Berkhout, Nijmegen, The Netherlands*.

van der Leij, A., van Bergen, E., van Zuijen, T., de Jong, P., Maurits, N., & Maassen, B. (2013). Precursors of developmental dyslexia: An overview of the longitudinal Dutch dyslexia programme study. *Dyslexia*, *19*(4), 191-213.

van der Mark, S., Klaver, P., Bucher, K., Maurer, U., Schulz, E., Brem, S., Martin, E., & Brandeis, D. (2011). The left occipitotemporal system in reading: Disruption of focal fMRI connectivity to left inferior frontal and inferior parietal language areas in children with dyslexia. *NeuroImage, 54*(3), 2426-2436.

Vandermosten, M., Boets, B., Wouters, J., & Ghesquière, P. (2012). A qualitative and quantitative review of diffusion tensor imaging studies in reading and dyslexia. *Neuroscience & Biobehavioral Reviews*, *36*(6), 1532-1552.

van Strien, J.W. (1992). Classificatie van links- en rechtshandige proefpersonen. *Nederlands tijdschrift voor de psychologie, 47,* 88-92.

van Viersen, S., de Bree, E. H., Zee, M., Maassen, B., van der Leij, A., & de Jong, P. F. (2018). Pathways into literacy: The role of early oral language abilities and family risk for dyslexia. *Psychological Science*, *29*(3), 418-428.

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*(1), 2-40.

Vidyasagar, T. R., & Pammer, K. (2010). Dyslexia: A deficit in visuo-spatial attention, not in phonological processing. *Trends in Cognitive Sciences*, *14*(2), 57-63.

Vinckenbosch, E., Robichon, F., & Eliez, S. (2005). Gray matter alteration in dyslexia: Converging evidence from volumetric and voxel-by-voxel MRI analyses. *Neuropsychologia, 43*(3), 324-331.

Vorst, H. C. M. (2006). Dyslexie Opsporingstest Niveau groep 7/8, DOT 7/8. Handleiding voor leerkrachten. Intern Rapport. Amsterdam: PsiTestuitgevers.

Vorst, H. C. M., & Bermond, B. (2001). Validity and reliability of the Bermond–Vorst alexithymia questionnaire. *Personality and Individual Differences*, *30*(3), 413-434.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV)*. San Antonio, TX: The Psychological Corporation.

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2018). The procedural learning deficit hypothesis of language learning disorders: We see some problems. *Developmental Science, 21*(2), e12552.

White, N. M. (2009). Some highlights of research on the effects of caudate nucleus lesions over the past 200 years. *Behavioural Brain Research*, *199*(1), 3-23.

White, S., Milne, E., Rosen, S., Hansen, P., Swettenham, J., Frith, U., & Ramus, F. (2006). The role of sensorimotor impairments in dyslexia: A multiple case study of dyslexic children. *Developmental Science, 9*(3), 237-255.

Willcutt, E. G., Betjemann, R. S., McGrath, L. M., Chhabildas, N. A., Olson, R. K., DeFries, J. C., & Pennington, B. F. (2010). Etiology and neuropsychology of comorbidity between RD and ADHD: The case for multiple-deficit models. *Cortex, 46*(10), 1345-1361.

Willcutt, E. G., Boada, R., Riddle, M. W., Chhabildas, N., DeFries, J. C., & Pennington, B. F. (2011). Colorado Learning Difficulties Questionnaire: Validation of a parent-report screening measure. *Psychological Assessment*, *23*(3), 778-791.

Willcutt, E. G., Pennington, B. F., & DeFries, J. C. (2000a). Twin study of the etiology of comorbidity between reading disability and attention-deficit/hyperactivity disorder. *American Journal of Medical Genetics*, *96*(3), 293-301.

Willcutt, E. G., Pennington, B. F., & DeFries, J. C. (2000b). Etiology of inattention and hyperactivity/impulsivity in a community sample of twins with learning difficulties. *Journal of Abnormal Child Psychology*, *28*(2), 149-159.

Willcutt, E. G., Pennington, B. F., Olson, R. K., & DeFries, J. C. (2007). Understanding comorbidity: A twin study of reading disability and attention-deficit/hyperactivity disorder. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *144*(6), 709-714.

Willows, D. M., & Terepocki, M. (1993). The relation of reversal errors to reading disabilities. *Visual Processes in Reading and Reading Disabilities*, 31-56.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, *91*(3), 415-438.

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming-speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities*, *33*(4), 387-407.

Wolff, U., & Lundberg, I. (2003). A technique for group screening of dyslexia among adults. *Annals of Dyslexia*, *53*(1), 324-339.

World Health Organization. (2008). *International statistical classification of diseases and related health problems* (10th rev., 2nd ed.). Geneva, Switzerland.

World Health Organization (2010). *International statistical classification of diseases and related health problems – tenth revision* (Version:2010). Geneva, Switzerland.

World Health Organization (2015). The ICD-10 (Version: 2015) classification of mental and behavioural disorders: Diagnostic criteria for research. World Health Organization: Geneva.

Wright, I. C., McGuire, P. K., Poline, J. B., Travere, J. M., Murray, R. M., Frith, C. D., ... & Friston, K. J. (1995). A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. *NeuroImage*, *2*(4), 244-252.

Xia, Z., Hancock, R., & Hoeft, F. (2017). Neurobiological bases of reading disorder part I: etiological investigations. *Language and Linguistics Compass*, *11*(4), e12239.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45-57.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., ... & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, *21*(4), 551-559.

Ziegler, J. C., Castel, C., Pech-Georgel, C., George, F., Alario, F. X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition*, *107*(1), 151-178.

Ziegler, J. C., & Goswami, U. (2006). Becoming literate in different languages: Similar problems, different solutions. *Developmental Science*, *9*(5), 429-436.

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F. X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences*, *102*(39), 14110-14115.

11

Publications and author contributions

Chapter 2 is published as:

Tamboer, P., Vorst, H. C. M., & Oort, F. J. (2014). Identifying dyslexia in adults: An iterative method using the predictive value of item scores and self-report questions. *Annals of Dyslexia*, *64*(1), 34-56.

Contributions:

Peter Tamboer designed the study, created new test material, carried out the data collection and statistical analyses and wrote the paper as first author. Harrie Vorst designed the study, created the *Communication Questionnaire*, contributed to the creation of other new test material, supervised the data collection and analyses, and reviewed the paper. Frans Oort reviewed the paper.

Chapter 3 is published as:

Tamboer, P., Vorst, H. C. M., & Oort, F. J. (2016). Five describing factors of dyslexia. *Journal of Learning Disabilities*, *49*(5), 466-483.

Contributions:

Peter Tamboer designed the study, carried out the statistical analyses and wrote the paper as first author. Harrie Vorst contributed to the analyses and reviewed the paper. Frans Oort contributed to the analyses – especially analyses in LISREL – and reviewed the paper.

Chapter 4 is published as:

Tamboer, P., & Vorst, H. C. M. (2015). A new self-report inventory of dyslexia for students: Criterion and construct validity. *Dyslexia*, *21*(1), 1-34.

Contributions:

Peter Tamboer designed the study, carried out the data collection and statistical analyses and wrote the paper as first author. Harrie Vorst designed the study, supervised the analyses and reviewed the paper.

Chapter 5 is published as:

Tamboer, P., Vorst, H. C. M., & de Jong, P. F. (2017). Six factors of adult dyslexia assessed by cognitive tests and self-report questions: Very high predictive validity. *Research in Developmental Disabilities, 71*, 143-168.

Contributions:

Peter Tamboer designed the study, carried out the data collection and statistical analyses and wrote the paper as first author. Harrie Vorst supervised the analyses and reviewed the paper. Peter de Jong contributed to the designs of the study, provided test material, and contributed to the analyses and the writing of the paper.

Chapter 6 is published as:

Tamboer, P., Scholte, H. S., & Vorst, H. C. M. (2015). Dyslexia and voxel-based morphometry: Correlations between five behavioural measures of dyslexia and gray and white matter volumes. *Annals of Dyslexia*, *65*(3), 121-141.

Contributions:

Peter Tamboer designed the study, carried out the data collection and statistical analyses and wrote the paper as first author. Steven Scholte designed the study, carried out the statistical analyses in MRI and reviewed the paper. Harrie Vorst designed the study, supervised the analyses, and reviewed the paper.

Chapter 7 is published as:

Tamboer, P., Vorst, H. C. M., Ghebreab, S., & Scholte, H. S. (2016). Machine learning and dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia. *NeuroImage: Clinical*, *11*, 508-514.

Contributions:

Peter Tamboer designed the study, carried out the data collection and statistical analyses and wrote the paper as first author. Harrie Vorst designed the study, supervised the analyses, and reviewed the paper. Sennay Ghebreab performed machine classification and pattern analyses of MRI data. Steven Scholte designed the study, carried out the statistical analyses in MRI and contributed to the writing of the paper.

12

Dankwoord

In de zomer van 2010, ik was net getrouwd, zei Harrie Vorst tegen mij: “Je moet maar gaan promoveren”. Daar had ik eigenlijk nooit aan gedacht. Ik vond het gewoon leuk om onderzoekjes te doen, maar niet met een bepaald persoonlijk doel voor ogen. Uiteindelijk is het er toch van gekomen en daarom wil ik een flink aantal mensen danken voor hun hulp en steun.

Ik moet eigenlijk beginnen met Jan Berends, Jeroen Beerepoot en Arjan de Wolf. Met hen heb ik mijn allereerste onderzoek naar dyslexie uitgevoerd in mijn tweede studiejaar onder leiding van Harrie Vorst. En door een fenomenale samenwerking gedurende mijn gehele studietijd met deze vrienden voor het leven ontdekte ik hoe leuk studie, wetenschap en onderzoek is.

Ik dank ook Patrick Snellings, die mij begeleidde tijdens een onderzoek voor mijn master, en mij stimuleerde door mij een enorme autonomie te geven in de invulling van het onderzoek. Dit soort stimulans had ik nodig om geïnspireerd te blijven voor toekomstig onderzoek.

Een buitengewoon aangename samenwerking heb ik gehad met Steven Scholte. Hij heeft alle analyses in MRI uitgevoerd, was altijd positief over plannen van Harrie en mij, en is bovendien een van de meest open-minded mensen die ik op de universiteit ben tegengekomen. Maar vooral zijn expertise is van onschatbare waarde geweest voor de interpretatie van onze onderzoeken. Ook Sennay Ghebreab wil ik hierbij bedanken voor zijn waardevolle bijdrage aan de analyses met MRI. En natuurlijk Tinka Beemsterboer die ervoor gezorgd heeft dat alle MRI-scans soepel zijn verlopen.

Toen ik mij met Harrie Vorst meldde bij Frans Oort met de vraag of hij mijn promotor wilde zijn, gingen er nieuwe deuren van expertise open. Zijn hulp bij de confirmatieve factoranalyses heeft mij veel geleerd. En ook Frans dank ik voor de buitengewoon positieve manier waarop hij mij heeft gestimuleerd om door te gaan, ook als er weer eens vertraging was doordat er bijvoorbeeld voor de zoveelste keer een artikel gereviseerd moest worden. Met Frans Oort kwam ook Peter de Jong in beeld. Mijn dank aan hem betreft vooral dat hij mij tot het eind toe scherp heeft gehouden. Elke redenering, elke zin is door hem met een

vergrootglas bekeken. En elke keer was het resultaat dat mijn teksten helderder werden en wetenschappelijk beter verantwoord.

Onderzoek doen is niet alleen analyseren en schrijven. Er komt zoveel bij kijken, dat het onmogelijk is alles alleen te doen. Toen ik begon met mijn promotieonderzoek maakte ik zelf testjes en vragenlijsten. Alles werd gekeurd door Ineke van Osch. Als geen ander kon zij beoordelen of test- of vragenlijstitems te moeilijk of te makkelijk waren. Ik vertrouwde daarop en achteraf is dat terecht gebleken: alle tests en vragenlijsten hadden precies de juiste moeilijkheidsgraad. Daarna kwam het afnemen van tests en vragenlijsten in de testweek. Met een onwaarschijnlijke snelheid en precisie hebben Paul Brouwer en Nihayra Leona alles geprogrammeerd, altijd geduldig als ik op het laatste moment nog met een wijziging kwam.

Degene die ik onwaarschijnlijk veel dank verschuldigd ben is Jan Hoogeboom. Veel van mijn tests en vragenlijsten zijn schriftelijk afgenomen bij honderden proefpersonen. Alles moest ingevoerd worden in de computer, een hels karwei dat Jan door de jaren heen steeds weer op zich genomen heeft. Zonder hem zou mijn proefschrift er nooit gekomen zijn.

Ook enkele vrienden van mij zijn van grote waarde geweest. Remco van Pareren die als taaldeskundige altijd met mij meedacht over de verschillende moeilijkheden in verschillende talen, en bovendien veel van mijn Engelse teksten met een deskundig oog heeft nagekeken. Alexander Kretzschmar die met mij meedacht als het over de kern van dyslexie ging en zelfs cruciale gedachtes als eerste heeft geformuleerd. Ruud Lok en mijn broer Victor, die mij af en toe op de juiste momenten uit mijn tunnel trokken voor een biertje, mijn onzekerheden moest aanhoren (“Misschien zit ik er wel helemaal naast”), en die altijd weer wegnam (“Waarom zou je er naast zitten?”).

Het was een mooie reis die begon in 2009, toen Harrie Vorst tegen mij zei: “We gaan eerst eens een paar keer goed van gedachten wisselen”. Een decennium lang heeft hij mij vertrouwen geschonken, mij begeleid en over alles met mij meegedacht. Soms dacht ik: “Ben ik zijn vertrouwen wel waard?” Zonder Harrie was ik nu geen onderzoeker geweest. Hij was van meet af aan de reden dat ik aan de UVA bleef plakken. Onze samenwerking en vriendschap hoef ik nu alleen nog maar in te lijsten.

Al die mooie onderzoeksplannen van Harrie en mij moesten natuurlijk wel uitgevoerd worden. En dat zou ook nooit mogelijk geweest zijn zonder de steun van Jaap Mulder. Als uitgever heeft hij alle projecten ondersteund. Maar niet alleen op papier. Ook tijdens eindeloze gesprekken in de auto op weg naar Deventer, en tijdens etentjes met heerlijke wijnen, met zijn vrouw Fieke, Harrie en Ineke, en Tessie, onvergetelijke avonden.

Tot slot wil ik natuurlijk mijn familie danken. Mijn vader, Kees, die vanaf mijn allereerste college tot aan de Nederlandse samenvatting van dit proefschrift van alles op de hoogte wilde blijven en die mij altijd voorgespiegeld heeft vol te houden in de strijd. Mijn moeder, Gini, een paar maanden geleden overleden. Van haar heb ik geleerd: als je iets wil bereiken, heb je twee dingen nodig, discipline en geduld. Zij heeft gelukkig nog net de eerste versie van mijn proefschrift in handen gehad.

"Je moet maar gaan promoveren". Mijn vrouw Anastasia, die mij alleen kende als iemand die steeds nachten door moet werken om op tijd af te kunnen studeren, dacht dat na mijn afstuderen dat nachtelijke werken wel voorbij zou zijn. Zij slaakte een diepe zucht bij deze woorden van Harrie. Toch is zij altijd mijn grootste steunpilaar geweest. Ik schrijf dit dankwoord na een nacht zonder slaap. Maar ik beloof je, Tessie, dat ik mijn leven nu zal beteren.