DOCTOR OF PHILOSOPHY

Dyslexia in adulthood : screening assessment and manifestations

Reynolds, Andrea

Award date:
2014

Awarding institution:
Bangor University

School of Psychology

Bangor University

DYSLEXIA IN ADULTHOOD: SCREENING ASSESSMENT AND MANIFESTATIONS

Andrea Reynolds

Supervisors:
Dr Markéta Caravolas and Professor James Intriligator

Thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

Bangor University

August 2014

Declaration and Consent

Details of the Work

I hereby agree to deposit the following item in the digital repository maintained by Bangor University and/or in any other repository authorized for use by Bangor University.

Author Name: Andrea Reynolds

Title: Dyslexia in Adulthood: Screening Assessment and Manifestations

Supervisor/Department: Dr Markéta Caravolas (School of Psychology)

Funding body (if any): Knowledge Economy Skills Scholarships (KESS)

Qualification/Degree: PhD in Psychology

This item is a product of my own research endeavours and is covered by the agreement below in which the item is referred to as "the Work". It is identical in content to that deposited in the Library, subject to point 4 below.

Non-exclusive Rights

Rights granted to the digital repository through this agreement are entirely non-exclusive. I am free to publish the Work in its present version or future versions elsewhere.

I agree that Bangor University may electronically store, copy or translate the Work to any approved medium or format for the purpose of future preservation and accessibility. Bangor University is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Bangor University Digital Repository

I understand that work deposited in the digital repository will be accessible to a wide variety of people and institutions, including automated agents and search engines via the World Wide Web.

I understand that once the Work is deposited, the item and its metadata may be incorporated into public access catalogues or services, national databases of electronic theses and dissertations such as the British Library's EThOS or any service provided by the National Library of Wales.

I understand that the Work may be made available via the National Library of Wales Online Electronic Theses Service under the declared terms and conditions of use (http://www.llgc.org.uk/index.php?id=4676). I agree that as part of this service the National Library of Wales may electronically store, copy or convert the Work to any approved medium or format for the purpose of future preservation and accessibility. The National Library of Wales is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Statement 1:

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless as agreed by the University for approved dual awards.

Signed .... (candidate)

Date: August 15^{th}, 2014

Statement 2:

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

All other sources are acknowledged by footnotes and/or a bibliography.

Signed .... (candidate)

Date: August 15^{th}, 2014

Statement 3:

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loan and for electronic repositories, and for the title and summary to be made available to outside organisations.

Signed .... (candidate)

Date: August 15^{th}, 2014

<u>Statement 4</u>:

Choose **<u>one</u>** of the following options

a)	I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University and where necessary have gained the required permissions for the use of third party material.	√
b)	I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University when the approved **bar on access** has been lifted.	
c)	I agree to submit my thesis (the Work) electronically via Bangor University's e-submission system, however I **opt-out** of the electronic deposit to the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University, due to lack of permissions for use of third party material.	

Options B should only be used if a bar on access has been approved by the University.

In addition to the above I also agree to the following:

1. That I am the author or have the authority of the author(s) to make this agreement and do hereby give Bangor University the right to make available the Work in the way described above.
2. That the electronic copy of the Work deposited in the digital repository and covered by this agreement, is identical in content to the paper copy of the Work deposited in the Bangor University Library, subject to point 4 below.
3. That I have exercised reasonable care to ensure that the Work is original and, to the best of my knowledge, does not breach any laws – including those relating to defamation, libel and copyright.
4. That I have, in instances where the intellectual property of other authors or copyright holders is included in the Work, and where appropriate, gained explicit permission for the inclusion of that material in the Work, and in the electronic form of the Work as accessed through the open access digital repository, *or* that I have identified and removed that material for which adequate and appropriate permission has not been obtained and which will be inaccessible via the digital repository.
5. That Bangor University does not hold any obligation to take legal action on behalf of the Depositor, or other rights holders, in the event of a breach of intellectual property rights, or any other right, in the material deposited.
6. That I will indemnify and keep indemnified Bangor University and the National Library of Wales from and against any loss, liability, claim or damage, including without limitation any related legal fees and court costs (on a full indemnity bases), related to any breach by myself of any term of this agreement.

Signature: ……….........................…………………. Date: August 15th, 2014

Summary

Dyslexia affects a large portion of the population, and identification of the individuals affected is important for effective intervention/accommodation that will facilitate the best outcomes. Screening is an important first step in the identification of individuals with the disorder. The measures utilized for this purpose must be effective, in order to ensure maximum accuracy, so that individuals are correctly identified, and are able to access the assistance required to mitigate the effects of the disorder. Currently, there are only a few tests available to screen for dyslexia in adults, and research evidence of their effectiveness is limited. This thesis provides: (1) in Chapter 2, a limited review of currently available screening tools; (2) in Chapters 3 and 6, empirical studies of the psychometric properties of two dyslexia screening tests for adults currently in use in the United Kingdom, the Bangor Dyslexia Test (BDT) and the Instines; (3) in Chapters 4 and 5, empirical studies of the development of a new test, the Bangor Adult Literacy Index (BALI); and (4) in Chapter 7, a study of the predictors (including several component measures of the BALI) of literacy skills among adults with and without dyslexia. The main findings across the research programme reported in this thesis are as follows. The review of the screening tools (Chapter 2) highlighted a lack of independent research studies on the tests and little empirical evidence to support the validity and reliability of some of the tests. The results of the evaluation of the BDT (Chapter 3) revealed that it is psychometrically sound and effectively discriminates between adults with and without dyslexia. For the Instines (Chapter 6), the opposite was true, and the validity and reliability of the test was found to be questionable. For the new screening measure, the BALI (Chapters 4 and 5), the results provided evidence of the validity and reliability of the tasks selected for inclusion. This new tool proved to be an excellent discriminator of dyslexic and non-dyslexic adult groups. The results of the investigation of the profiles of cognitive predictors of literacy skills for adults with and without dyslexia (Chapter 7) provided evidence that the predictors are similar to those identified for children. Importantly, we also found evidence to suggest that the profiles of cognitive predictors of literacy skills, and the effects of IQ on these in adults with dyslexia, are the same or similar regardless of IQ level. Together, these studies have contributed to research into dyslexia in adulthood by enhancing

the literature available on dyslexia screening tests and our understanding of the manifestation of dyslexia in this population.

Acknowledgments

This thesis was a Knowledge Economy Skills Scholarship funded project and involved collaborative research with Remploy a company sponsor. This thesis would not have been possible without my supervisor Dr. Markéta Caravolas who created the opportunity for me to undertake this research. Thank you for having the confidence in me and my abilities to accomplish the work required. Thank you also for all your help, patience, support, the skills and knowledge I've gained during the course of this work. I would also like to thank my second supervisor Professor James Intriligator and the chair of my PhD. committee Dr. Christoph Kline for their feedback on my work.

Thank you to Siti Wan Aminah Wan Norudin for her valuable assistance with data collection. Special thanks to Mark Burrett my supervisor at my company sponsor Remploy for his great help in facilitating the research undertaken with the company. I am very grateful to Liz Du Pré from the Miles Dyslexia Center for help in recruiting dyslexic participants and for proofreading my work. Thank you also to Dr. Anna Samara for her companionship and friendship and especially for her help with formatting this thesis.

Special thanks to all my family and friends whose love, support, and encouragement sustained me throughout. I am especially grateful to my mother Sylvia Brown, my sister Marva Reynolds, and my friend Deidre-Ann Mills for their regular phone/Skype calls. Thank you also to my church family at Ebenezer Evangelical Church for their prayerful and emotional support.

Finally, and most importantly, I thank my heavenly Father for His unending love, amazing grace, and great faithfulness towards me.

Contents

List of Figures

List of Tables

Chapter 1: Literature Review

1.1 Definition of Dyslexia

Dyslexia is a condition that affects 4-10% of the population (Rice & Brook, 2004), and has been recognized and studied for over 100 years. Yet, despite this long-standing and intense interest in the condition, a widely accepted definition has yet to be produced. The absence of a single universally accepted definition of dyslexia has resulted in a plethora of definitions for the disorder. These definitions are developed by dyslexia advocacy groups, practitioners, researchers, and governments. Rice and Brooks (2004) offer an extensive list of some definitions as well as a more detailed analysis of 28 of these. There are several reasons why a definition of dyslexia is essential. Lyon (1995) highlighted three key reasons. Firstly, a definition is necessary for the *identification* of dyslexia, the main symptoms and characteristics needed to describe it. Secondly, *treatment* of dyslexia requires an understanding of the difficulties experienced by individuals with dyslexia. Thirdly, a definition is necessary for *research purposes*, as researchers who investigate the causes, consequences, co-morbidity and other aspects of the disorder need to have clearly defined selection criteria. Variations in the definitions of dyslexia are therefore likely to affect the identification of and interventions offered to dyslexic individuals as well as research findings. A definition is also important for the educational systems to identify and put appropriate measures in place to assist students (Siegel, 1999).

1.1.1 IQ Discrepancy Approach

Historically, a common feature of definitions of dyslexia has been the inclusion of a discrepancy criterion. The discrepancy criterion generally states that a discrepancy between an individual's IQ and his/her reading achievement or reading age and chronological age must exist for a diagnosis of dyslexia to be made. The use of this discrepancy criterion can be traced to a study of reading disability by Rutter and Yule (1975). These researchers investigated the characteristics of poor readers and concluded that there were two distinct groups, one group with reading achievement below expected level based on their IQ, and a another group whose reading achievement was consistent with their IQ. However, for over two decades, the use of an IQ discrepancy for defining dyslexia and the identification of dyslexics has been controversial, and this has become one of the most contentious aspects of the definition of dyslexia. A growing body of research with children has cast doubt on the validity of the IQ discrepancy definition.

The main arguments against the discrepancy approach are that there is an overlap between individuals with reading disabilities with and without an IQ discrepancy, and it is difficult to differentiate the groups according to their cognitive and behavioural characteristics, as well as their response to intervention and prognosis (Fletcher et al., 1992; Fletcher et al., 1994; Francis et al., 2005; Stage, Abbott, Jenkins, & Berninger, 2003; Stuebing, Barth, Molfese, Weiss, & Fletcher, 2009; Stuebing et al., 2002). For example, research by Fletcher et al. (1992) compared the performance of children with reading disability who presented with or without an IQ discrepancy on various neuropsychological tests, and found only small group differences. Similarly, Fletcher et al. (1994) found small and non-significant differences when they compared the performances of children with reading disability with and without an IQ discrepancy on tasks related to reading proficiency. Also, a recent meta-analysis of 22 studies investigating the relationship between different measures of IQ and reading intervention revealed that IQ was not an important predictor (Stuebing et al., 2009). Based on these and other similar research findings the discrepancy criterion has been eliminated from some of the more current definitions of dyslexia. For example, in the United States a discrepancy criterion is no longer required for a diagnosis of dyslexia (Taymans et al., 2009). Similarly, in the UK the recently conducted Rose Review (2009) by Sir Jim Rose, of the identification and teaching of children with dyslexia produced a definition of dyslexia without an IQ discrepancy. However, the discrepancy criterion is still widely used in the UK in the *assessment* and *identification* of dyslexia.

Although there is now a large body of research questioning the validity of the IQ discrepancy in children, very little research with adults has been undertaken. A meta-analysis conducted by Swanson and Hsieh (2009) investigated the validity of the IQ discrepancy definition in adults, and their results contradicted the research with children. This meta-analysis of 52 studies revealed that adults with reading disability with IQ and reading discrepancy had greater deficits when compared to their counterparts with reading disabilities and no IQ and reading discrepancy (Swanson & Hsieh, 2009). More research is needed with adults with dyslexia to confirm the findings of Swanson and Hsieh (2009) and to help resolve the debate about the use of the IQ discrepancy definition of dyslexia.

1.2 Manifestations of Dyslexia

Notwithstanding the lack of agreement on the definition of dyslexia, there is convergent empirical evidence of the behavioural manifestations of the disorder in children and in adults. Children with dyslexia have extreme difficulty acquiring basic reading sub-skills such as letter sound knowledge and word identification (Vellutino, Fletcher, Snowling, & Scanlon, 2004). Additionally, the disorder is characterised by difficulty in phonological processing, reading (accuracy and speed), and spelling, with dyslexic children performing less well than their non-dyslexic counterparts (Lennox & Siegel, 1993; Manis & Custodio, 1993; Snowling, Goulandris, & Defty, 1996). For phonological processing, dyslexic children often exhibit deficits on various phonological skills such as phonological awareness, phonological short term memory, and rapid naming (Gathercole & Baddeley, 1990; Griffiths & Snowling, 2002; Manis & Custodio, 1993; Snowling et al., 1996).

1.2.1 Manifestations of Dyslexia in Adults

Longitudinal studies of dyslexic children have shown that some of their difficulties persist into adolescence and adulthood (Maughan et al., 2009; Shaywitz et al., 1999; Snowling et al., 1997, Snowling, Muter, & Carroll, 2007; Svensson & Jacobson, 2006). These studies assessed the difficulties experienced by individuals who were diagnosed with dyslexia in childhood. Maughan et al. (2009) assessed the spelling of poor readers in adolescence and again 30 years later, and found that their spelling difficulties were highly persistent. As in the child population, adults with dyslexia have been reported to exhibit a wide range of deficits at the behavioural level, as well as differences at the brain and genetic levels (see Fisher et al., 2002; Vellutino et al., 2004; Wagner, 2005 for reviews). Over and above weaknesses in reading efficiency (accuracy and speed) and spelling, behavioural markers most typically include difficulties in the accuracy and/or speed of processing phonological (speech sound) information, and memory (Vellutino et al., 2004). In addition, single or multiple deficits have sometimes been reported in the domains of language use and comprehension, auditory and speech perception, visual attention, motor coordination, and associative learning (Pennington, 2006; Ramus, Pidgeon, & Frith, 2003). The prevalence of each type of difficulty and their rates of co-occurrence have not yet been clearly established in the broad dyslexic population; however, deficits in phonological processing and verbal short term memory tend to predominate in terms of both severity and frequency of occurrence relative to the other domains (Bruck, 1992; Gathercole, Alloway, Willis, & Adams, 2006; Snowling et al., 1997).

Although the underlying deficits in dyslexia may not resolve over the course of development, or not completely, some changes in the behavioural manifestations of difficulties nevertheless do occur. For example, by adulthood many English-speakers with dyslexia are able to close the gap in their reading accuracy though rarely also in fluency (Lefly & Pennington, 1991). These changes in the behavioural manifestations have been attributed to instruction and compensation (Frith, 1999). Nevertheless, it seems that spelling accuracy as well as phonological processing *speed* tend to remain impaired into adulthood, and this across languages (Callens, Tops, & Brysbaert, 2012; Hatcher, Snowling, & Griffiths, 2002; Re, Tressoldi, Cornoldi, & Lucangeli, 2011).

1.3 Comorbidity

Dyslexia often occurs with other developmental disorders such as attention deficit hyperactive disorder (ADHD), specific language impairment (SLI), developmental coordination disorder (DCD), speech sound disorder (SSD) and dyscalculia (Hulme & Snowling, 2009; Landerl & Moll, 2010; McGrath et al., 2011; Pennington, 2006, Tunick & Pennington, 2002). Epidemiological research indicates that the co-occurrence of these disorders is greater than expected based on the rate of their occurrence in the population (Caron & Rutter, 1991). Accurate estimates of the rates of comorbidity of dyslexia and other developmental disorders are difficult, as this requires representative samples of each population (Hulme & Snowling, 2009). The comorbidity rates of dyslexia and some of the developmental disorders reported across studies are: for ADHD – 25-40% (Pennington, 2006), SLP – 17-29% (Catts, Adlof, Hogan, & Weismer, 2005), and SSD – 2-13% (Pennington & Bishop, 2009). Comorbidity of developmental disorders has been attributed to shared aetiological, genetic, cognitive, and environmental risk factors that increase susceptibility to developing more than one disorder (Pennington, 2006; Pennington et al., 2009; Hulme & Snowling, 2009; Stevenson et al., 2005; Willcutt et al., 2010). Comorbidity may have implications for the manifestations of the disorders, exacerbating the impairments exhibited, or alternatively result in compensatory effects which mask some deficits resulting in more complex assessment and treatment (Pennington, 2006).

1.4 Main Causal Theories of Dyslexia

In addition to the controversies surrounding the definition of dyslexia, there is also debate about the cause or causes of the disorder, and there is no universally accepted causal theory of dyslexia. Currently, there are a number of cognitive theories of dyslexia that propose that the disorder is caused by cognitive deficits due to structural and functional brain anomalies (Vellutino et al., 2004). These theories may be divided into two main headings phonological and sensorimotor (Ramus, 2003). The main sensorimotor theories include the auditory deficit (Tallal, 1980) and the cerebellar deficit (Nicolson &Fawcett, 1999) the main phonological theory is referred to as the phonological deficit theory. A review of these theories follows.

1.4.1 Phonological Deficit Theory

The phonological deficit theory has been the most widely accepted causal cognitive theory of dyslexia. This theory states that dyslexia is caused by a phonological processing deficit (possibly from structural and functional anomalies in the brain) which results in impairment in the representation and processing of speech sounds (Reid, Szczerbinski, Iskierka-Kasperek, & Hansen, 2007; Snowling, 2001; Stanovich & Siegel, 1994; Vellutino et al., 2004). Manifestations of the phonological deficit include difficulties in phonological awareness (the production and manipulation of the sound structure of words), verbal short term memory (temporary storage of phonological representations), and phonological processing speed (the articulation of the phonological form of the word) (Snowling, 1995).

Phonological deficits provide a good explanation of the behavioural manifestations of the disorder and are consistent with theories of normal literacy development, as learning to read requires an understanding of the sound structure of words and children with dyslexia have difficulties acquiring this basic letter-sound knowledge (Hulme & Snowling, 2009; Snowling, 2001). The theory is supported by strong and convergent evidence of phonological processing deficits in children with dyslexia (Vellutino et al., 1996) which persist into adulthood (Bruck, 1992; Kemp, Parrila, & Kirby, 2009; Shaywitz et al., 1999). Furthermore, studies investigating the acquisition of reading and the causes of reading failure have provided evidence of a causal relationship between phonological awareness and reading ability. Studies such as the seminal work of Bradley and Bryant (1983) and others have demonstrated that the acquisition of reading skills is affected by the ability to learn the sound structure of words (i.e., to have phoneme awareness) (Caravolas et al., 2005; Hulme, Bowyer-Crane, Carroll, Duff, & Snowling, 2012, Hulme et al., 2002; Hulme & Snowling, 2009; Savage & Carless, 2005; Snowling, 1998). The causal relationship between phonological processing skills and reading is also supported by intervention studies which have demonstrated that phonological awareness training has a beneficial effect on literacy development (Snowling, 1998; Hulme et al., 2012). Deficits in phonological processing are also associated with difficulties in the storage and retrieval of printed words (Vellutino et al., 2004).

Neuroimaging studies of dyslexic individuals provide evidence of defective activation and abnormal connectivity in the language regions (left hemisphere) of their brains (Démonet, Taylor, Chaix, 2004; Peterson & Pennington, 2012; Richlan, Kronbichler, & Wimmer, 2009; Shaywitz et al., 1998; for a review see Sun, Lee, & Kirby, 2010). In these studies, in both children and adult dyslexics, researchers have found differences in the structure and function of the language areas of the brains when engaged in reading and reading related skills (Shaywitz et al., 1998; Shaywitz et al., 2002). For example, differences (increased activation) have been reported in the left temporoparietal cortex of dyslexic children and adults, including the Wernicke's area, the angular gyrus, striate cortex and the inferior frontal gyrus (Peterson & Pennington, 2012; Temple, 2002).

Despite the general acceptance of a phonological deficit in individuals with dyslexia, there is some opposition to the theory. The main argument against the theory is that it does not explain other deficits (e.g. sensory and motor) sometimes exhibited by dyslexic individuals and therefore cannot be the only cause of reading disability (Ramus et al., 2003b; Valdois, Bosse, & Tainturier, 2004). Consequently, a number of alternative causal theories have developed that posit that dyslexia may be caused by more general deficits in (auditory, visual, and motor areas) (Nicolson & Fawcett, 1999; Tallal, 1980; Tallal, Miller, & Fitch, 1993; Valdois, et al., 2004). Additionally, the theory does not explain the existence of individuals with reading disability without a phonological deficit (Valdois, et al., 2003). While the other theories challenge the phonological deficit theory, on the basis that a phonological deficit is not the only cause of dyslexia, they are all compatible with its main premise that the deficit is a direct cause of dyslexia. A review of some of the contending theories of dyslexia follows.

1.4.2 Auditory Temporal Deficit Theory of Dyslexia

This theory states that dyslexia is caused by an underlying general deficit in auditory temporal perception of short or rapidly varying sounds that impairs speech perception (Tallal, et al., 1993; Tallal, 1980). The impairment in speech perception may lead to difficulty with speech at the phonemic level and the skills required for reading acquisition. The theory is based on research conducted by Tallal and Piercy (1975) which found that children with specific language impairments were impaired on temporal order judgment tasks presented at long but not short inter-stimulus intervals (Tallal & Piercy, 1975; Tallal & Piercy, 1973). In a subsequent study, Tallal (1980) compared the performance of children with reading difficulties and controls on a battery of auditory perception tasks and found the dyslexic children were impaired (made more errors) in their performance on stimuli presented rapidly but not at slower rates of presentation. The researcher also found that the scores of the children on auditory tasks were significantly positively correlated with their scores on a non-word reading task. The researcher used this correlation to support her hypothesis that the auditory perceptual deficit is associated with difficulty in learning the sound-symbol relationship.

The theory is supported by several research studies that have also found auditory processing deficit on a range of auditory perception tasks in children (Bishop et al., 1999; Farmer & Klein, 1995; Goswami et al., 2002; Heiervang, Stevenson, & Hugdahl, 2002; Muneaux, Ziegler, Truc, Thomson, & Goswami, 2004; Tallal et al., 1993) and adults with dyslexia (Hämäläinen, Salminen, & Leppänen, 2013; Pasquini, Corriveau, & Goswami, 2007). However, research studies conducted by Goswami and colleagues investigating auditory perception deficits in individuals with dyslexia (Goswami et al., 2002; Muneauxi et al., 2004; Pasquini et al., 2007; Richardson, Thomson, Scott, & Goswami, 2004) suggest that amplitude envelope rise time may be more important as a causal deficit in dyslexia than a temporal perception or other auditory deficits. The rise time may be described as the point at which a vowel or a syllable is perceived and is designed the Perceptual center (Muneauxi et al., 2004). The Perceptual center may be important for onset and rime awareness and a deficit may cause impairment in phonological awareness and literacy. Goswami and colleagues in their research found that compared to controls both children and adults with dyslexia are deficient on tasks assessing amplitude envelope rise time as well as other auditory deficits including temporal order judgment and rapid frequency discrimination. Additional, they found scores on the amplitude envelope rise time tasks are significant unique predictors of performance on reading and phonological processing tasks even after controlling for other auditory deficits such as rapid temporal processing (Muneauxi et al., 2004; Pasquini et al., 2007). From their research these researchers have concluded that an auditory deficit in amplitude envelope rise time may be a more important causal deficit for dyslexia than a temporal order deficit.

One of the main arguments against the auditory deficit theory is that there is no evidence that the reading deficit experienced by dyslexics is caused by an auditory deficit (Hämäläinen et al., 2013; Ramus et al., 2003; Vellutino et al., 2004). Also, while phonological deficits have been found in most individuals with dyslexia, only a small sub-group of individuals with dyslexia exhibit auditory deficits, suggesting that the phonological deficit is not caused by an auditory deficit (Ramus et al., 2003b; Rosen & Manganari, 2001). Additionally, there is little evidence to support the theory's position that the phonological deficit is secondary to impairments in rapid auditory processing (Marshall, Snowling, & Bailey, 2001). Although there is evidence of a relationship between performance on tasks assessing auditory deficit and phonological abilities (as indicated by their correlations) the nature of this has not been clearly explained by the theory (Marshall et al., 2001). A study by Marshall, Snowling, and Bailey (2001) found that auditory deficit does not predict phonological deficits as would be expected based the theory. Furthermore, research studies investigating a rapid auditory processing deficit in children with dyslexia have suggested that this may be due to a speech discrimination deficit and not a temporal processing deficit (Mody et al., 1997; Waber et al., 2001). To summarise the evidence against the auditory processing deficit theory suggests that although some individuals with dyslexia may exhibit this deficit, which may be associated with a phonological deficit, it has no causal role in dyslexia. Tallal’s theory can be grouped with other theories that propose that dyslexia is caused by temporal processing deficits due to magnocellular impairments one of these is described below.

1.4.3 Magnocellular Theory

This theory states that the temporal processing deficit exhibited by individuals with dyslexia is caused by a general deficit in the processing of rapidly changing sensory information in any domain, including auditory, visual, and motor (Stein, 2001; Stein & Walsh, 1997). This deficit is attributed to impairment in the magnocellular system which processes fast temporal information (Stein, 2001; Stein & Walsh, 1997). This impairment results in reduced motion sensitivity, resulting in unsteady binocular fixation (control of eye movement) and poor visual localization which affects the appearance and order of letters in a word (Stein, 2001). It also impairs sensitivity to auditory transients which determines the ability to distinguish acoustic cues in frequency (FM) and amplitude modulation (AM) of tone, resulting in deficient phonological skills (Stein, 2001). Support for the theory is provided by research evidence of impairment in motion sensitivity and auditory transients in individuals with dyslexia (Stein & Walsh, 1997; Talcott, Hansen, Assoku, & Stein 2000). As with the auditory temporal deficit theory, the main argument against the magnocellular theory is the failure to identify deficits in individuals with dyslexia, and where a deficit is found, only a small sub-group exhibit it (Ramus et al., 2003b; Skottun, 2000).

1.4.4 Cerebellar/Automatization Deficit Theory

The cerebellar deficit theory of dyslexia evolved from the automatization deficit theory. Automatization may be defined as “the process by which skilled performance becomes smoother and smoother, requiring less and less effort, following extensive practice” (Nicolson & Fawcett, 1999 p. 159). Therefore, when a skill becomes automatized it is performed automatically without conscious monitoring (Ramus et al., 2003). The theory states that individuals with dyslexia have difficulty automatizing all skills, motor and cognitive, including reading, spelling, and phonological (Nicolson & Fawcett, 1999). The theory was based on research conducted by the proponents of the theory which found impaired automatization of balance in children with dyslexia. Using a dual task paradigm, the researchers found that children with dyslexia exhibited impairments in motor skills in the dual task conditions, where motor balance and backward counting were assessed simultaneously, but not on single task conditions (Nicolson & Fawcett, 1990). Nicolson and Fawcett, (1999) attributed the automatization difficulties observed in children with dyslexia to a deficit in the cerebellum and proposed the cerebellar theory as an explanation for the automatization deficit and the main deficits (reading, spelling, and writing) observed in individuals with dyslexia. The theory posits that the cerebellar deficit can account for the literacy related difficulties exhibited by dyslexics as follows: (1) it results in poor general motor skill causing poor handwriting quality, (2) it affects articulatory skills resulting in phonological deficits, and hence reading difficulty, (3) it causes an automaticity impairment resulting in difficulty in automatizing the sub-skills of reading and spelling, the acquisition of visual word forms affecting both reading and spelling (Nicolson & Fawcett, 1999; Ramus et al., 2003). The proponents of the theory argue that unlike the other theories of dyslexia this theory provides a parsimonious account of the range of difficulties exhibited by individuals with dyslexia (Nicolson & Fawcett, 1999).

The cerebellar deficit theory is supported by a number of research studies conducted by the proponents of the theory and their colleagues as well as other researchers. Studies have found that children (Fawcett & Nicolson 1999; Fawcett, Nicolson, & Dean 1996; Quercia, Demougeot, Santos, & Bonnetblanc, 2011; Ramus et al., 2003; Wolff, 2002) and adults (Needle, Fawcett, & Nicolson 2006; Stoodley, Fawcett, Nicolson, & Stein, 2006) with dyslexia perform less well than their counterparts without dyslexia on behavioural tasks (e.g. heel-to-toe balance, maintenance of posture, muscle tone) assessing the cerebellar deficit. In addition, anatomical and functional differences in the brains of adults with dyslexia compared to adults without dyslexia have also been found (Nicolson, Fawcett, & Dean 2001; Finch, Nicolson, & Fawcett 2002; Rae et al., 1998). In one study, Nicolson and Fawcett along with other colleagues conducted a number of studies which investigated the cerebellar deficit by comparing the performance of children with and without dyslexia on various tasks assessing difficulties associated with the deficit. The researcher found that children with dyslexia performed less well than their counterparts without dyslexia, and that their performance was consistent with a cerebellar impairment, thus providing empirical support for the theory.

The cerebellar theory is challenged by independent studies that have not found differences in the performance of children (Irannejad & Savage 2012; White et al., 2006, Viholainen et al., 2010; Wimmer, Mayringer, & Landerl, 1998; Wimmer, Mayringer, & Raberger, 1999) and adults (Ramus, et al., 2003b) with and without dyslexia on cerebellar tasks. Additionally, in some studies where group differences have been found, only a minority of the dyslexics exhibited the deficit, which precludes the cerebellar deficit as a causal deficit (Ramus, et al., 2003a; Ramus, et al., 2003b; Wimmer, Mayringer, & Raberger, 1999). Also, research indicates that the existence of the cerebellar deficits in some individuals with dyslexia may be attributed to the comorbidity of dyslexia with developmental disorders such as attention deficit hyperactive disorder (ADHD) and dyspraxia (Rochelle & Talcott, 2006; Wimmer et al., 1999). For example, Wimmer et al., (1999), compared the balancing performance of children with and without dyslexia and found impairment only in the performance of the children with dyslexia who also obtained high rating on a teacher rating of ADHD symptoms. The performance of the children with dyslexia and low rating on the ADHD measures was similar to that of the controls. Further evidence is provided in a meta-analysis of 15 studies conducted by Rochelle and Talcott (2006) that compared the balance performance of individuals with dyslexia and controls. From their analysis, these researchers found that although balance deficits may be associated with dyslexia there are not strongly related to reading ability. The final argument against the cerebellar theory is that there is little research evidence of a causal relationship between motor difficulties and literacy skills (Barth et al., 2010; Ramus et al., 2003a). In a study investigating the cerebellar deficits and reading intervention with children with reading difficulties Barth et al. (2010) found no correlation between the children's scores on two cerebellar tasks (postural stability and bead threading) and literacy measures (reading, spelling and phonological awareness). Similar results were obtained in another study investigating the relationship between motor control and phonology in children with dyslexia. In this study, Ramus et al. (2003a) compared the performance of children with and without dyslexia on tasks assessing phonological skills and cerebellar function and also found no correlation between the children's scores on these tasks.

More recently, the proponents of the cerebellar theory have further updated their causal theory of dyslexia providing an explanation of the disorder at the neural system level, with their proposal that the disorder is due to a dysfunction in the procedural learning system (Nicolson & Fawcett, 2011; Nicolson & Fawcett, 2007; Nicolson, Fawcett, Brookes, & Needle, 2010). The procedural learning system (also called the procedural memory system) is involved in: (1) the learning of rules governing language, (2) the learning of new skills, and (3) the control of sensorimotor and cognitive habits (Nicolson & Fawcett, 2007). The theory states that impairment in the procedural learning system resulting from dysfunction in the cortico-cerebellar circuits in the brain results in impaired acquisition of skills including language.

1.4.5 Visual Attention Span Deficit Hypothesis

The visual attention span (VAS) deficit theory proposes that in addition to the well established phonological deficit, a VAS impairment is a second core deficit in dyslexia (Bosse, Tainturier, & Valdois, 2007). The theory is theoretically grounded in the connectionist multitrace memory model of polysyllabic word reading, a model of reading which posits two routes, lexical and non-lexical for reading, and explains how damage to these can result in reading difficulty (Ans, Carbonnel, & Valdois, 1998). Bosse et al. (2007) define VAS as "the amount of distinct visual elements which can be processed in parallel in a multi-element array" (p. 1). The theory states that dyslexia is caused by two independent cognitive impairments, a phonological or a VAS deficit, with both resulting from damage to the visual attention window (VAW) through which information about the orthography of words is obtained (Valdois, Bosse, & Tainturier, 2004). The extent of the damage to the VAW determines the type of deficit exhibited. Moderate damage to the VAW would result in a VAS deficit and the pattern of reading difficulty associated with surface dyslexia, with intact phonological skills but difficulties with irregular word reading. More severe damage to the VAW results in a nonword deficit and a pattern of reading difficulty associated with phonological dyslexia with impaired phonological skills and difficulty with both regular and irregular words. The theory therefore explains the phonological and surface subtypes of dyslexia.

The theory is supported by research undertaken by Valdois and colleagues and others who have found that children (Bosse et al., 2007; Bosse & Valdois, 2009; Hawelka & Wimmer, 2005; Lobier, Zoubrinetzky, & Valdois, 2012; Valdois et al., 2003) and adults (Hawelka, Huber, & Wimmer, 2006) with dyslexia exhibit a visual attention span deficit. Bosse et al. (2007) conducted two studies; the first examined the performance of French speaking and the second English speaking, dyslexic and control children, on tasks assessing their reading, phonological processing and VAS skills. For the study with the French speaking children, the dyslexic children performed less well than the controls on all the tasks and the children's scores on the phonological and VAS tasks correlated with their reading scores. Also, the children's scores on the phonological and VAS tasks were significant independent predictors of reading, suggesting that the VAS deficit was independent of the phonological deficit. More importantly using principal component analysis, the research found that a high proportion of the dyslexics (44%) exhibited a VAS deficit only. These results were replicated in the second study with the English speaking dyslexic children where the researchers controlled for nonverbal IQ, vocabulary, and letter identification skills. The theory is challenged by research studies that have not found a VAS deficit in adults with dyslexia (Hawelka & Wimmer, 2008; Shovman & Ahissar, 2006). Additionally, there is some debate about the nature of the deficit demonstrated with the VAS task, and some researchers contend that this may be phonological and not visual because of the use of verbal material and oral report in the task (Hawelka & Wimmer, 2008; Shovman & Ahissar, 2006; Ziegler, Pech-Georgel, Dufau, & Grainger, 2010). This view is supported by research studies using a VAS task without oral report that failed to find the deficit (Hawelka & Wimmer, 2008; Shovman & Ahissar, 2006). However, the proponents of the theory recently conducted a study that provides evidence of the deficit in dyslexic children on VAS tasks as well as categorization tasks, using both verbal and non-verbal stimuli (Lobier et al., 2012).

Additionally, other studies that have used VAS tasks without verbal involvement have also found the deficit (Pammer, Lavis, Hansen, & Cornelissen, 2004).

1.5 Current Theoretical View of Dyslexia

Increasingly dyslexia is being viewed as a disorder with a multifactorial/multiple deficits aetiology. This view acknowledges that the disorder does not have clearly identified boundaries and that it may share aetiological risk factors with other developmental disorders such as ADHD (Norton & Wolf, 2012; Pennington, 2006; Snowling, 2012). This new conceptualization of dyslexia does not negate the widely accepted phonological deficit theory but acknowledges the fact that there are aspects of the disorder that we do not currently fully understand.

1.6 Identification of Adults with Dyslexia

There is general agreement that the identification of children with dyslexia should be undertaken as early as possible, as this results in better outcomes (Rose Review, 2009). However, there is some debate about the method to be used to identify children with dyslexia, whether school based assessments (such as comparison of progress with peers without dyslexia) or screening tests are more appropriate. Research indicates that screening tests may be unreliable producing large numbers of false positives and false negatives and the Rose Review recommended the comparison with peers without dyslexia as the first step in identification of children with dyslexia (Rose, 2009; Simpson & Everatt, 2005). In addition to identifying dyslexia in children, there is also a need to be able to identify adults with dyslexia as not all individuals affected by the disorder are identified before adulthood. The report of the National Working Party on Dyslexia in Higher Education (1999), which indicates that approximately 43%, of students with dyslexia are only identified after admission to post-secondary institutions, suggests that a large proportion of individuals are not identified in childhood. It has also been suggested that in the wider work community the number of adults with undiagnosed dyslexia may be high (Nicolson, Fawcett, & Miles, 1993). This has implications for not only these individuals but also society as a whole.

In addition, several acts of legislation (Disability Discrimination Act 1995, Special Education Needs and Disability Act (SENDA), 2001, Disability Discrimination Act 2005), culminating in the passing of the recent Equality Act 2010, have highlighted the need for awareness-raising about dyslexia in the higher /further education sector and in the workplace. These acts were designed, on the one hand, to prevent discrimination against people with disability (including dyslexia) in education, employment and access to services, and on the other hand, for enabling education providers and employers to identify and support adult students and employees with dyslexia. Additionally they aim to ensure for those with disabilities, fair treatment and access to opportunities on a par with counterparts without disabilities. Similar legislation has been passed in the European Union (The Framework Directive on Equal Treatment in Employment and Occupation 2000) and the United States (Americans with Disability Act 1990). Thus, there is an ongoing need for the identification of adults with dyslexia, to allow them to access the resources required to assist them in achieving their potential, and to assist HEIs, employers and others in fulfilling their legal and professional obligations.

Dyslexia (in both children and adults) can only be diagnosed by a fully trained psychologist or other professionals such as specialist teachers who have received the appropriate training. The limited number of these professionals, and the high cost associated with diagnosis, has resulted in the practice of screening individuals to identify those most at risk of the disorder, and the referral of these individuals for a full assessment by a trained psychologist or other professionals. The use of this two stage process received the endorsement of the majority of the professionals working with dyslexics in a feasibility study into adult dyslexia screening conducted in 1992 (Nicolson, Fawcett, & Miles, 1993).

Screening tests are broadly designed to be quick, cost effective, and easy to administer tools for identifying individuals most at-risk of a disease or disorders (Singleton, 2009; Evans, Galen, & Britt, 2005). As such, screening tests do not need to be administered by fully qualified psychologists, and thus may comprise useful tools for HEIs, employers, and others to use in the identification of individuals who may be at-risk of dyslexia. In contrast with in-depth evaluations, however, screening tools have certain limitations. Their brevity may increase the risk of misclassifications (of false negatives and false positives) being made.

Moreover, many screening tools currently on the market lack adequate validation and standardization studies to attest their appropriateness, sensitivity and specificity for the reliable detection of dyslexia risk in adults (see Chapter 2 for details).

In comparison with the availability of dyslexia screening tests for children, very few are available for adults, and, of these, there is no generally accepted or "gold standard" measure. The tests fall into two main categories: paper based and computerised assessments. Two frequently used paper based tests are the Bangor Dyslexia Test (BDT) (Miles, 1997) and the Dyslexia Adult Screening Test (DAST) (Fawcett & Nicolson, 1998). The main computerised screening tests include the Instines (Teare, 2001), Lucid Adult Dyslexia Screening Plus (LADS Plus) (Lucid Research Limited, 2010), and QuickScan (Zdzienski, 1998). There are a number of advantages as well as disadvantages for both types of screening tests. For paper based tests, advantages include: (1) they allow the collection of qualitative data; (2) they are suitable for individuals with low literacy skills; (3) they facilitate comprehension of task requirements. Disadvantages include: (1) they maybe require more time to administer; (2) the cost, including material and labour, may be greater; (3) scoring may be less objective. For computer based tests, advantages include: (1) more accurate and consistent administration; (2) reduction in labour, time and cost; (3) more objective and accurate scoring. Disadvantages include: (1) more costly development process (2) possible technology failure, and (3) frequent lack of transparency about the causes of poor performance on computerised (frequently self-administered) tasks.

Despite the fact that the screening tests currently used in the United Kingdom, have been in use for several years (the BDT for almost three decades), there is a still a paucity of published research on these tests, and this includes developmental as well as independent research. Additionally, where independent research studies do exist, their findings may contradict those reported by the test developers. Thus, there is a need for further research on these screening tests, in order to determine their capacity to effectively discriminate between adults with and without dyslexia. A limited review of the more popular screening tools currently being used in the United Kingdom is undertaken in Chapter 2.

Chapter 2: Review of Dyslexia Screening Tests for Adults

2.1 Introduction

In this chapter, we review six of the dyslexia screening tests for adults that are available in the United Kingdom. Three of the tests are paper based: Bangor Dyslexia Test, Dyslexia Adult Screening Test, and York Adult Assessment Battery-Revised; and three are computer based: Instines, Lucid Adult Dyslexia Screening Plus, and QuickScan. The tests are assessed on the established characteristics of a good test. A test may be described as good, if it is valid, reliable, discriminating, and has adequate norms (Cohen & Swerdlik, 2005; Kline, 1986).

2.2 Bangor Dyslexia Test

2.2.1 Test Description

The Bangor Dyslexia Test (BDT) was one of the first dyslexia screening tests to be developed in the United Kingdom and was created by the late Professor Emeritus of Bangor University Tim R. Miles. It is a paper-based dyslexia screening test for individuals aged from 7 years to adulthood. To develop the test, Professor Miles was mainly informed by anecdotal evidence from his clinical experience and work with dyslexic individuals. Although, as Professor Miles stated in his book Dyslexia: The Pattern of Difficulties (1993), there was no theoretical basis for the *specific items* selected for the test, in his theoretical perspective, he believed dyslexia to be a syndrome with a distinctive pattern of symptoms/difficulties resulting from a lexical or verbal labelling deficit (Miles, 1993; Payne, Miles, & Wheeler, 2007). Over a period of approximately six years, he experimented with a number of different tasks before selecting 12 that were included in the original version of the test published in 1983. The test was subsequently revised and its current format was published in 1997 and has been translated into several languages including Greek, German, Japanese and Welsh (Miles, 1993). It is intended for use by professional as well as non- professional assessors. The items selected for inclusion on the test were all tasks on which most dyslexic individuals had difficulties and which Miles believed could identify the pattern of difficulties associated with the disorder. The test consists of 10 subtests: 8 skill-based tasks, and 2 questions, which are anecdotal queries on persisting confusion of the letters *b* and *d*, and on the incidence of dyslexia in the family. The subtests are Left-Right, Polysyllabic Words, Subtraction, Tables, Months Forwards, Months Reversed, Digits Forwards, Digits Reversed, B-D Confusions and Familial Incidence. Descriptions of these subtests are provided in Table 2.1. Administration time for the test is approximately 30 minutes.

Table 2.1. Description of the Subtests of the Bangor Dyslexia Test

Subtests	Descriptions	Skills Assessed
Left – Right	Tests the awareness of left and right using body parts.	Verbal working memory, spatial awareness, and mental rotation.
Polysyllabic Words	Tests the ability to repeat polysyllabic words such as ‘preliminary’ and ‘philosophical’.	Verbal/phonological short term memory and articulatory accuracy.
Subtraction	Tests the ability to complete verbally presented subtraction problems.	Verbal working memory and arithmetic skills.
Tables	Tests the ability to recite 6, 7 & 8 times tables.	Rote and verbal working memory, arithmetic skill, and executive functions for sequencing.
Months Forwards	Tests the ability to recite the months of year in the correct order.	Rote recall and executive function for sequencing.
Months Reversed	Tests the ability to recite the months of the year in reverse order.	Verbal working memory and executive function for sequencing.
Digits Forwards	Tests the ability to repeat digits in the order in which they were presented.	Verbal short-term memory.
Digits Reversed	Tests the ability to repeat digits in the reverse order of presentation.	Verbal working memory.
B – D Confusion	Question on the confusion of the letters *b* – *d* beyond the age of eight.	Not applicable
Familial Incidence	Question on the incidence of learning difficulty in the family.	Not applicable

The scoring system for the screening process is deliberately simple. There are three possible scores on each subtest: + (plus), a dyslexia positive response; – (minus), a dyslexia negative response; and 0 (zero) an ambiguous response that cannot be scored as either dyslexia positive or negative. In addition to the correctness of the response for 8 of the subtests (excluding B-D Confusion and Familial Incidence), scoring is also based on the clinical judgement of the assessor. The assessor is required to take into account any manifest difficulty experienced or explicit strategies used by the individual being assessed to achieve the correct response. Indications of difficulties experienced by the test taker (to be noted by the assessor) include hesitations, requests for repetitions of the question, repeating the question before answering, and other strategies indicating difficulties. Therefore, a + (plus) score would be given not only for incorrect responses, but also for correct responses that satisfy other criteria such as the use of strategies or other evidence of difficulty. A – (minus) score is awarded for correct responses (with no indication of difficulties or strategies) and a score of 0, (zero) is awarded for correct responses that are ambiguous and cannot be scored as either dyslexia positive or negative. These scores are then assigned numerical values in order to calculate the individual's total score on the scale such that plus (+) = 1, zero (0) = .5, and minus (–) = 0; thus, the total score on the scale overall ranges from a minimum of 0 to a maximum of 10.

2.2.2 Test Development

Norms. Norms for the BDT were established from data collected in 1980, as a part of the longitudinal Child Health and Education Study involving all children in England, Scotland and Wales born during the week of April 5^{th} -11^{th}, 1970 (Miles, 1993). Data were collected from 12,905 children at age 10 years, 6685 boys and 6,220 girls (Miles, Haslum, & Wheeler, 1998). The children were assessed on three of the BDT subtests: Left-Right, Months Forwards, and Months Reversed. For each subtest, only a minority of the children obtained positive scores as follows: Left-Right 1408 (11%), Months Forwards 971 (7.85%) and Months Reversed 1615 (13.2%). Norms for two subtests (Digits Forwards and Digits Reversed) were already established by other recognised measures such as the British Ability Scales and the Wechsler Intelligence Scale for Children (Miles, 1993). Norms for adults were not established. The test has a manual that provides information on administration, scoring, and interpretation of test scores, along with information on the validation studies and the establishment of norms. Further details about the development of the test are also available in the book Dyslexia: The Pattern of Difficulty by Professor Miles.

Validity. In addition to establishing norms for children, data collected from the Child Health and Education Study were used to validate some of the subtests of the BDT (Miles, 1993). The relationship between three subtests (Left-Right, Months Forwards, and Months Reversed) of the BDT and literacy measures (word recognition, spelling, mathematics and comprehension) was examined using stepwise regression. The results indicated that scores on the subtests were significant predictors of scores on the literacy measures such that children with positive scores on the subtests obtained lower scores on the literacy measures than children with negative scores. Similar results were also obtained for the Recall of Digits subtest from the British Ability Scales that was also administered to the children and this test is similar to the Digits Forwards and the Digits Reversed subtests of the BDT.

Further validation was obtained by comparing the performance of 264 children and adolescents, 132 with dyslexia and 132 without dyslexia (matched for intelligence) on 7 of the subtests: Left-Right, Polysyllabic Words, Subtraction, Tables, Months Forwards, Months Reversed, and Digits Reversed (Miles, 1993). The participants were aged between 7 years and 18 years and were divided into three age groups, 7-8 (42 children), 9-12 (160 children), and 13-18 (62 adolescents), with equal numbers with and without dyslexia in each group. For all the subtest, the children with dyslexia performed significantly less well (obtaining higher scores) than the children without dyslexia and this was true for all three age groups. Additionally, on all the subtests, the percentage of children with dyslexia obtaining a plus score or a positive indicator was greater than those without dyslexia. Finally, the BDT was administered as part of a study which investigated the speed of multiplication in boys with dyslexia. The performance of the boys with dyslexia (age 13-14 years-old), was compared to that of chronological age-matched peers, and to that of spelling age-matched boys (10-11 years of age) (Miles, 1993). The performance of the groups differed significantly, with the dyslexic group performing less well (obtaining higher scores) than the others. The results of these studies indicated that children with dyslexia are more likely to obtain higher scores on the BDT than children without dyslexia and provided empirical evidence of the capacity of the BDT to discriminate between these such groups.

2.2.3 Independent Research

Sutherland and Smith (1991) compared the BDT to two other dyslexia screening tests, the Boder Test of Reading-Spelling Patterns and the Aston Index. The Boder is an American test and was designed for use with individuals aged 5 to adulthood. The Aston Index, a British test, was designed for individuals aged 5 to 14 years. Participants were 20 children with literacy problems and average non-verbal IQ mean age 11.5 years, the researchers compared the screening outcome of the tests, their ability to group participants into sub-types of dyslexia (auditory, visual and mixed), and how easily teachers could administer the tests. The researchers found that the screening outcomes of the three measures were inconsistent, and coincided for only six of the 20 participants. Additionally the classification of the children without dyslexia into different sub-types was inconsistent, with only three participants given the same sub-type classification by all three tests. For the BDT, its total score as well as the score on the Polysyllabic Words subtest were significant predictors of screening outcomes. However, the researchers concluded that the BDT was too general to be useful for teachers. Additionally they were critical of the BDT manual which they believed provided limited guidance on how to interpret the test scores.

In another study conducted by Nichols, McLeod, Holder, and McLeod (2009), the researchers compared the effectiveness of a tutor screening battery and the computerized screening test, Lucid Adult Dyslexia Screening (LADS) for identifying university students with learning disability (dyslexia, dyspraxia, and Meares-Irlen Syndrome). The ‘tutor screening battery’ (a compilation of published tests used for screening students with disabilities at the University of Worcester) consisted of seven subtests of the BDT (Left-Right, Polysyllabic Words, Subtraction, Tables, Months Forwards, Months Reversed, B-D Confusion, and Familial Incidence) and eight subtests of the Dyslexia Adult Screening Test (DAST). Participants in the study were 74 students of whom 46 screened positive for at least one of the disabilities and were referred for full assessment by a qualified professional. Of these 46 students given a full assessment, 35 were diagnosed with at least one disability and the majority 30, with dyslexia. The ‘tutor screening battery’ had a sensitivity rate of 91% and a specificity rate of 79%, compared to 66% and 90% respectively for the LADS. The researchers concluded that the ‘tutor screening battery’, with its higher sensitivity rate, was more effective, as the lower sensitivity rate of the LADS would result in a higher proportion of individuals with disabilities not being identified. The researchers also used logistic regression analysis to examine the capacity of the subtests of the BDT (eight subtests) and the DAST (seven subtests) to predict the three disabilities assessed, and both tests were significant predictors. They further examined the relationship of the individual subtests of the BDT and DAST and indications of the three disabilities. This analysis revealed that of the eight subtests of the BDT, two (Left-Right and Months Reversed) were significant predictors, while Polysyllables (p = 0.076) approached significance. Additionally, the Months Forward subtest was excluded from the analysis as only one student obtained a positive score on it. It is not possible to properly evaluate the effectiveness of the BDT from this study, as it was combined with another screening test, and not all the subtests were included; nevertheless, the results suggest that it may correctly identify adults with dyslexia.

2.2.4 Summary and Conclusion

The BDT was developed as a quick and easily administered screening test for dyslexia. Norms for the test were established from a large and widely representative sample of the population of children in the United Kingdom but were restricted to 10 year olds. Additionally norms were only established for three subtests (norms already existed for the Digits Forwards and Digits Reversed subtests) and no norms were established for adults. The validation studies conducted by the developer, Professor Miles, indicate that the BDT is capable of discriminating between children with and without dyslexia. However, no validation study with an adult population was conducted. Independent research, (Nichols et al., 2009) suggests that the BDT may be capable of identifying adults with dyslexia. However, the results of this study also suggest that the specificity rate of the BDT may be below the minimum acceptable (90%), and this would result in an unacceptably large number of individuals without dyslexia being incorrectly classified as dyslexic. No research studies have examined the reliability of the BDT, and given the limitations highlighted above in the research undertaken for the development of the test, and in the study by Nichols et al. (2009), more research is needed (particularly with adults) is order to properly evaluate the BDT.

2.3 Dyslexia Adult Screening Test

2.3.1 Test Description

The Dyslexia Adult Screening Test (DAST) is a paper-based dyslexia screening test developed by Professors A. J. Fawcett and R. I. Nicolson to screen individuals aged 16 years 6 months to age 75 years. The developers had previously created the Dyslexia Early Screening Test (DEST) and the Dyslexia Screening Test (DST) designed for children between 4 years and 16.5 years. The DST, the prototype for the DAST, was modified to facilitate screening for dyslexia in adults. The DAST manual indicates that several years of research and testing contributed to its development and modifications were made based on feedback from teachers in further education. The developers indicate that the theoretical basis of the DAST is derived from clearly established difficulties experienced by individuals with dyslexia, including phonological, cerebellar (balance), working memory, and speed of processing difficulties, and tasks assessing these difficulties are included in the test (Nicolson & Fawcett, 1998). The tasks included in the test also reflect the developers' view that dyslexia is caused by a cereballar deficit. They indicated that two main types of tasks were included in the test, diagnostic (capable of giving positive indication of dyslexia) and attainment tests. The DAST consists of 11 subtests: and includes tests of attainment: One Minute Reading, Two Minute Spelling, and One Minute Writing: and diagnostic tests measures - Rapid Naming, Postural Stability, Phonemic Segmentation, Backwards Digit Span, Nonsense Passage Reading, Nonverbal Reasoning, Verbal Fluency, and Semantic Fluency. Further details of the subtests are provided in Table 2.2 below. The tests of attainment of reading (efficiency), spelling, and writing assess skills with which the developers believe adults with dyslexia experience difficulty and which are unlikely to be easily remediated (Nicolson & Fawcett, 1998). The developers support the discrepancy definition of dyslexia, that is the disorder is defined by a discrepancy between literacy achievement and IQ and the Non-verbal Reasoning subtest was designed to be a measure of fluid intelligence. However, the discrepancy definition of dyslexia is not supported by the majority of current research studies (see Chapter 1 Literature review). The Non-verbal Reasoning subtest along with the Semantic Fluency subtest was included, as a test of possible strengths for individuals with dyslexia (Nicolson & Fawcett, 1997). A raw score is obtained on each subtest which is converted to an ‘At Risk Quotient’ (ARQ) based on the age appropriate norms for the subtest. The ARQ for each subtest is then averaged to obtain an overall ARQ which gives an indication of the risk of dyslexia. An overall ARQ of 1.0 or greater indicates high risk of dyslexia while an ARQ between 0.7 and 1.0 indicates low risk of dyslexia. The DAST is intended for use by employment or education professionals such as teachers and trainers, as a dyslexia screening test and to provide a profile of strengths and weaknesses, which can be used to guide support in education or in employment. Administration time is approximately 20 - 30 minutes.

Table 2.2. Description of Subtests of the Dyslexia Adult Screening Test

Subtests	Description	Skills Assessed
Rapid Naming	Tests the time taken to name all the simple pictures on a card	Speed of lexical access and articulation
One Minute Reading	Tests the number of words read correctly in one minute	Reading Efficiency
Postural Stability	Tests the disturbance in balance caused by a calibrated push in the back	Balance
Phonemic Segmentation	Tests the ability to split words into constituent phonemes (syllables)	Phonological awareness
Two Minute Spelling	Tests the number of words spelt correctly in two minutes	Spelling Fluency
Backwards Digits Span	Tests the ability to correctly repeat a number of digits in the reverse order of presentation	Working memory
Nonsense Passage Reading	Tests the number of words correctly read in a short passage with real and nonsense words	Grapheme/phoneme translation fluency
Nonverbal Reasoning	Test of nonverbal IQ requiring the completion of a sequence, an analogy, and identification of similarities and differences	Nonverbal reasoning
One Minute Writing	Tests the number of words of a sentence that can be correctly copied in one minute	Transcription fluency
Verbal Fluency	Tests the number of words beginning with S that can be orally generated in one minute	Verbal fluency
Semantic Fluency	Tests the number of names of animals that can be orally generated in one minute	Semantic fluency

Note. Adapted from Nicolson & Fawcett, 1997

2.3.2 Test Development

Norms. With the exception of the Nonverbal Reasoning test, all tests were taken from the DST, and were adapted mainly by increasing the difficulty and the length, to make them age appropriate. The manual indicates that professionals from the Further Education and Higher Education sectors were also consulted and their feedback was used to refine the tests. Two sets of norms were established, one for the general population and another for the student population (including norms for different age groups). Norms for students were established by testing 550 students from Universities in Sheffield, Leeds and Bristol. The sample included at least 100 first, second, and third year undergraduates, and smaller numbers of postgraduates and mature students; no other information on the characteristics of the students is given. Norms for the general population were established from data collected from 600 adults (whose geographic locations were not reported) with an equal number of males and females. For this sample, 65% had finished school at ages 16-17, 19% at ages 18-19, and 21% had engaged in higher or further education after age 21; no further information on the characteristics of sample is given.

Reliability. The test-retest reliability of the DAST was assessed by administering the test a week apart to 33 adults (the manual gives no further details of this sample). The correlations for the subtests of the DAST ranged from a low of r = .64 for Backwards Digit Span to a high of r = .93 for Two Minute Spelling (see all reliabilities in Table 2.3). Six of the 11 subtests (One Minute Reading, Phonemic Segmentation, Two Minute Spelling, Nonsense Passage Reading, One Minute Writing, and Verbal Fluency) had very good test-retest reliability, with correlations of r =.81 to r =.93 which are above the minimum (.70) recommended (Kline, 2000). The reliability of the remaining subtests was adequate.

Table 2.3. Test-retest Reliability of the Subtests of the Dyslexia Adult Screening Test (N = 33)

Subtests	Correlation
Rapid Naming	.68
One Minute Reading	.90
Postural Stability	.72
Phonemic Segmentation	.90
Two Minute Spelling	.93
Backwards Digit Span	.64
Nonsense Passage Reading	.92
Nonverbal Reasoning	.75
One Minute Writing	.87
Verbal Fluency	.81
Semantic Fluency	.76

Note. Adapted from Nicolson & Fawcett, 1997

Validity. The DAST was validated by comparing the performance of students with dyslexia (15) and without dyslexia (150). The overall mean ARQ for the dyslexics was 1.72 and for the non-dyslexics 0.18. Using the high risk criterion (ARQ 1.0 and above), the manual reports that 14 of the 15 students with dyslexia (sensitivity rate of 93%) and all of the 150 students without dyslexia (specificity rate 100%) were correctly classified, for an overall hit rate of 99%. However, if the low risk criterion is used (ARQ 0.07 - 1.0), two students without dyslexia were misclassified which would lower the specificity rate slightly to 98.6% and the overall hit rate to 98.2%. These rates are above the minimum recommended for a good screening test, 80% and 90% respectively (Glascoe & Byrne (1993). The Two Minute Spelling subtest was the most discriminating, identifying all the students with dyslexia. Conversely, the Postural Stability subtest was the least discriminating identifying 63% of the students with dyslexia. The manual indicates that the subtests with the lowest hit rates (Postural Stability, Phonemic Segmentation, and Backwards Digit Span) were redesigned to make them more challenging; however, no details of this were given. The results of the validation study indicated that the DAST is capable of discriminating between adults with and without dyslexia, with dyslexics obtaining higher scores on the subtests as well as the overall score.

2.3.3 Independent Research

Harrison and Nichols (2005) investigated the capacity of the DAST to discriminate between adult students with and without specific learning disabilities (dyslexia) in post-secondary education by comparing their performance on the measure. The DAST was administered to 238 students, 117 with diagnosed specific learning disabilities (SLD) and 121 controls with no history of SLD. Based on their total score on the DAST, 87 of the 117 students with SLD were correctly classified, for a sensitivity rate of 74%, and 102 of the 121 student without SLD for a specificity rate of 84%. Both rates are below acceptable levels (sensitivity 80% and specificity 90%), and would result in unacceptable levels of false negatives 26% and false positives 16%. Also, both rates are below those obtained by the developers in their validation study which indicated that the DAST had a sensitivity rate of 93% and a specificity rate of 100%. Harrison and Nichols (2005) also assessed the capacity of each of the subtests of the DAST to discriminate between the groups by examining the scores obtained by the students on them. The results indicated that the sensitivity rates of three of the subtests were higher than that of the test overall. The Nonsense Passage Reading subtest was the most discriminating, identifying 91% of the students with SLD, followed by Phonemic Segmentation and Two Minute Spelling, with 85% each. Three of the subtests identified fewer than 50% of the students with SLD; Rapid Naming 47%, Verbal Fluency 27 %, and Semantic Fluency 15%. Interestingly, the Postural Stability subtest correctly identified 66% of the students with SLD. This is similar to the percentage of students with dyslexia correctly classified in the validation study for the DAST suggesting that the redesign of the task (as reported in the manual) may not have improved its sensitivity. From the results, Harrison and Nichols (2005) concluded that the DAST is limited as a screening test in its present form.

In another study, (highlighted earlier), Nichols et al. (2009) compared the effectiveness of the Lucid Adult Dyslexia Screening (LADS) test to a tutor screening battery consisting of eight subtests of the Bangor Dyslexia Test and seven subtests from the DAST. This study found that the tutor screening battery had a higher sensitivity rate than the LADS, 91% compared to 66%, but its specificity rate was lower at 79% compared to 90% for the LADS. The nature of this study makes it impossible to determine the effectiveness of the DAST, as it was not the only screening test used, however, the results suggest that total scores on the DAST may be able to discriminate between adults with and without dyslexia. Nichols et al. (2009) also used logistic regression to examine the discriminatory capacity of the subtests of the DAST. They found that only three of the seven subtests (One Minute Reading, Phonemic Segmentation, and Two Minute Spelling) included in the analysis were significant predictors of the participants' group membership for the disabilities examined. For these subtests, the results of this study are similar to those of Harrison and Nichols (2005), and the validation study undertaken by the developers which found that these subtests had high sensitivity rates and effectively discriminated between adults with and without dyslexia. The results also indicate the need for further research in order to establish the effectiveness of the DAST.

2.3.4 Summary and Conclusion

The DAST was designed to screen for dyslexia in adults by assessing skills on which adults with dyslexia exhibit deficits, as well as to identify strengths and weaknesses. The manual for the DAST is fairly comprehensive and provides information on the development of the test, administration, scoring, and interpretation of results. Norms for the test appear to have been adequately established with a sample of 1150 adults with separate norms for students and the general population. However, not enough detail is given about the characteristics of the samples and this is particularly true of the sample used to establish norms for the general population. The test-retest reliability reported for the subtests of the DAST was within or above the minimum, .70 considered acceptable, indicating that they are consistent over time. However, these figures should be treated with caution as the short time period between testing (one week) may have artificially inflated the reliability coefficients. The findings of the research studies undertaken to date suggest that the test overall is capable of discriminating between adults with and without dyslexia. However, these results (including the validation study by the developers) also suggest that this capacity may be based on only some of the subtests included in the DAST. Furthermore, how effectively the DAST discriminates, that is its sensitivity and specificity rates, is questionable, as research findings are inconsistent across studies. There is therefore a need for further research in order to determine the true effectiveness of the DAST as a dyslexia screening test for adults and to confirm the adequacy of its psychometric properties.

2.4 Instines

2.4.1 Test Description

The Instines is a computerized dyslexia screening tool designed to recognise characteristics associated with dyslexia in adolescents (aged 12 years and older) and adults (James, 2004). It was developed by Philip Teare, is now being distributed by the Dyslexia Foundation, and has recently been renamed *Dyslexia-Check*. There is no published manual for the test and we were not able to find any published empirical research studies. Information for this review was obtained mainly from the internet and from using the software. The theoretical basis for the test is not known but an examination of the tasks included in the test suggests that it may have been designed to assess deficits associated with dyslexia such as word recognition, spelling, and reading fluency. The inclusion of a verbal reasoning task suggests support for the discrepancy definition of dyslexia, and James (2004) indicates that the test includes an IQ test which is used to detect discrepancy in performance. It consists of eight subtests, namely Spatial Awareness and Recognition, Homophone/Spelling Recognition, Verbal Reasoning, Digit Span, Right and Left Awareness, Directional Awareness, Reading Speed, and Comprehension. Further details of the subtests are provided in Table 2.4 (also please see Chapter 6 for more information). The test provides a screening outcome which states whether there are indications of dyslexia and this is determined by comparing the scores of individuals with and without dyslexia (James, 2004; Teare, 2003). It also provides a percentile score for each subtest but no overall score is generated. This information is also presented graphically and is considered useful for assessing support required for individuals with positive screening outcomes (James, 2004). The test is self-administered, instructions for all the subtests are presented verbally and/or visually, and the test taker is required to either select the correct answer from a number of options or type a response. Administration time is approximately 30 minutes.

Table 2.4. Description of the Subtests of the Instines

Subtests	Description	Skills Assessed
Spatial Awareness and Recognition	Tests the ability to complete pattern sequences	Visual processing abilities
Homophone/Spelling Recognition	Tests the ability to recognise and distinguish real words from nonwords	Orthographic, word recognition, and spelling abilities
Verbal Reasoning	Tests the ability to identify similarity in meaning between words	Conceptual understanding of words
Digit Span	Tests the ability to recall digits in the order of presentation and reverse order of presentation	Verbal short-term and working memory
Right and Left Awareness	Tests knowledge of left and right	Directional awareness
Directional Awareness	Tests knowledge of the four cardinal points and the ability to navigate a 3D maze	Directional awareness
Reading Speed	Tests the time taken to read a passage of text	Reading speed
Comprehension	Tests the ability to correctly answer questions on passage read	Reading comprehension

2.4.2 Test Development

Norms. Information on the development of the Instines is extremely limited. The following was taken from information provided for administrators of the test in the software. Norms for the Instines were established in the United Kingdom on a stratified sample of over 100 participants per age group, with two groups, 12 - 15 years and 16 years and over. Participants were from independent and state secondary schools, further and higher educational institutions, employed, and unemployed adults. No further details about the sample or the procedures used are given.

Validity. James (2004) in his descriptive study of the Instines, indicated that the test was validated on a sample of over 700 of which 170 were individuals with dyslexia. However, no further details of this validation were provided. Information in the software that is available to the administrator states that the test was validated by comparing the screening outcome of the Instines with the diagnosis made by psychologists in three trials. The screening outcomes of the Instines were correct on 94 – 98% of screenings; however, no further details of these trials are provided.

2.4.3 Independent Research

We were able to identify only one published study of the Instines by James (2004). The researcher made a presentation at the sixth international conference of the British Dyslexia Association comparing screening tests for dyslexia for use in further and higher educational institutions. In his presentation, James reported information on several adult dyslexia screening tests: three computerised – Instines, Lucid Adults Dyslexia Screening, and QuickScan, and two paper based – BDT, DAST, and a check list – the Adult Dyslexia Checklist. He provided information on the screening tests under the following headings: (1) how dyslexia is indicated; (2) recommendations for needs and learning style; (3) inclusion of an IQ test; (4) literacy skills tested; (5) validation. In his description of the validation study for the Instines (highlighted above) the researcher was critical suggesting that the Instines had not been tested in a rigorous scientific manner.

2.4.4 Summary and Conclusion

The Instines is a computerised dyslexia screening test designed to detect characteristics associated with dyslexia in adolescents and adults. It is not possible to adequately assess the Instines, as there is insufficient or no information on its development and psychometric properties (reliability and validity). Although there is some indication that norms for the test were established, details of this process are not available and therefore no assessment of it is possible. The same is true of the validation study reported by James (2004), and of the information provided in the software for the test. Several other aspects of the Instines raise concern. The test, unlike most dyslexia screening tests, does not include a task assessing phonological processing (while it does include several unusual tests, such as

Directional Awareness). As a phonological deficit is the main deficit associated with the disorder, the lack of a task assessing it is likely to adversely affect the capacity of the test to identify individuals with dyslexia. In addition, the criteria used to determine the screening outcome of the Instines is not known. However, if the reported 94% – 98% agreement of the screening outcome of the test with a psychologist diagnosis were correct, the test would be highly sensitive. There is, therefore, an urgent need for research into the capacity of the Instines to discriminate between individuals with and without dyslexia and an examination of its psychometric properties. This is required especially because despite the lack of research, and knowledge about how the test discriminates between individuals with and without dyslexia, a promotional flyer advertising the test states that approximately 800 schools, further and higher educational institutions and other organizations in the United Kingdom are using the Instines.

2.5 Lucid Adult Dyslexia Screening Plus

2.5.1 Test Description

The Lucid Adult Dyslexia Screening Plus (LADS Plus) is an adaptive computerised dyslexia screening test designed to screen for dyslexia in individuals aged 15 years and older and was developed by Lucid Innovations Limited (Singleton, Horne, Thomas, & Leedale, 2004). It was developed from its predecessor the Lucid Adult Dyslexia Screening (LADS) and was designed to facilitate group as well as individual screening. For group screening, up to 40 persons can be screened simultaneously on a computer network. The manual indicates that the LADS was developed after three years of research at the University of Hull into computer-based dyslexia screening test. LADS was first released in 2002, and was revised in order to improve it capacity to identify individuals with dyslexia in the wider population including individuals with disadvantaged educational background and second language English speakers; this led to the development of the LADS Plus (Lucid Adult Dyslexia Screening Plus Administrator's Manual, 2010). The developers indicate that the design of the LADS Plus was based on strong scientific research and focused on the "core cognitive deficits" associated with dyslexia in adults. They identify these core deficits as phonological processing, lexical access, and working memory and tasks assessing skills in these areas are included in the test.

The test consists of five subtests: Word Recognition, Word Construction, and Working Memory (dyslexia-sensitive measures designed to identify deficits associated with the disorder), and Nonverbal Reasoning and Verbal Reasoning (designed as quasi measures of nonverbal and verbal intelligence). Further details of the subtests are provided in Table 2.5 below. Although the skills assessed by the tasks included in the test are in line with the developers' conceptualisation of dyslexia, the capacity of this combination of tasks to effectively identify adults with dyslexia is questionable. The tasks included in the test assess reading, memory, and IQ. There is an ongoing debate about the usefulness of IQ for the identification of individuals with dyslexia (see literature review Chapter 1) but the general consensus among most researchers is that an assessment of IQ is not required (Fletcher et al., 1994; Francis et al., 2005; Stage, Abbott, Jenkins, & Berninger, 2003; Stuebing, Barth, Molfese, Weiss, & Fletcher, 2009). Also, although deficits in reading are well documented in adults with dyslexia, it may not be the most prominent deficits exhibited and therefore may not be the most effective measure for identification of adults with the disorder (Kemp, Parrila, & Kirby, 2009; Lefly & Pennington, 1991).

Scores are obtained on each subtest. For the three dyslexia-sensitive subtests, scores range from one to nine as follows: 1 – 3 (no indications of dyslexia), 4 – 5 (weak indications of dyslexia), and 7 – 8 (strong indications of dyslexia), with higher scores indicating increased probability. The test also provides an overall classification of the probability of dyslexia which includes four categories: (1) *low probability of dyslexia*, (2) *borderline*, (3) *moderate probability of dyslexia*, and (4) *high probability of dyslexia*. For the Nonverbal Reasoning and Verbal Reasoning subtests, scores are classified into five categories: (1) *low* – lowest 10% of adult population, (2) *below average* – next 15% of adult population above low, (3) *average* – middle 50% of adult population, (4) *above average* – next 15% of adult population, and (5) *high* – top 10%. The test is self-administered with each subtest taking approximately 5 minutes to complete, totalling approximately 25 minutes for the entire screening process. An internet version of this test called *Spot your Potential* has also been developed by Lucid Innovations Limited.

Table 2.5. Description of the Subtests of the Lucid Adult Dyslexia Screening Plus

Subtests	Description	Skills Assessed
Nonverbal Reasoning	Tests the ability to complete matrix puzzles	Intellectual ability (nonverbal IQ)
Verbal Reasoning	Test the ability to identify conceptual relationship between two items (pictures).	Intellectual ability (verbal IQ)
Word Recognition	Tests the speeded recognition of real words from non-words	Phonological processing (decoding), lexical access and working memory
Word Construction	Tests the ability to create (encode) non-words from syllables.	Phonological processing (awareness and encoding), lexical encoding and auditory short-term memory
Working Memory	Tests the recall of a sequence of digits in the reverse order of presentation	Short-term working memory

2.5.2 Test Development

Norms. The manual indicates that the Verbal and Nonverbal Reasoning subtests were adapted from other assessment tests created by the developers. The Verbal Reasoning subtest was adapted from the Verbal Reasoning subtest of the Lucid Ability test, which is a verbal and non-verbal reasoning test for children aged 4–16 years. Norms for the subtest were established on a sample of 2,216 children aged 4 to 16 years from several schools and a range of socio-economic groups in the United Kingdom (Lucid Ability Administrator’s Manual, 2011). No further information about the characteristics of the normative sample is given. The Nonverbal Reasoning subtest was adapted from the Reasoning subtest of the Lucid Assessment System for School (LASS). Norms for the LASS were established from a sample of 505 students, of which 300 boys and 205 girls from 14 schools in different areas of the United Kingdom (Lucid Assessment System for School Teachers Manual, 2010). The students ranged in age from 11 years to 15 years 11 months with an average age of 13 years 2 months. No further information about the characteristics of the normative sample is given.

The manuals for the LADS and the LADS Plus contain no information about how norms for the other three subtests were developed nor about the norms for an adult population.

Reliability. The internal consistency reliability coefficient reported for the Word Recognition subtest is $\alpha = .95$ and Word Construction subtest $\alpha = .96$ (Lucid Adult Dyslexia Screening Plus Administrator's Manual, 2010) which are well above the minimum acceptable $\alpha = .70$ (Kline, 2000). The internal consistency of the Working Memory, Verbal Reasoning, Non-verbal Reasoning subtests and the overall test were not reported nor was any other assessment of reliability.

Validity. The three dyslexia- sensitive subtests (Word Recognition, Word Construction, and Working Memory) were validated in three different studies. The results of first validation study were published in the *Journal of Research in Reading* (Singleton, Horne, & Simmons, 2009). In this first study, the performance of adults with and without dyslexia was compared on the subtests and the Reading and Spelling subtests of the Wide Range Achievement Test (WRAT-3). The participants' performance was assessed on the full pool of items available for each subtest as well as on the adaptive versions. Participants were 139 adults recruited from eight adult educational institutions, two universities, three colleges of further education, and three basic skills centres, and included 70 dyslexics and 69 controls, of which 58 were males and 82 females. The number of participants across educational institutions was: universities 42, further education colleges 48, and basic skills centres 49. The results revealed statistically significant differences between the performances of the groups on all the tasks including both versions of the LADS Plus subtests. The dyslexic group obtained higher scores (indicating greater risk of dyslexia) than the controls. There were also significant differences between the performances of the groups when the scores from all three subtests were combined, thus indicating that the subtests were capable of discriminating between dyslexics and non-dyslexics. Using the composite scores for the three subtests (adaptive forms), the capacity of the subtests to discriminate between the dyslexic and the control participants was assessed using discriminant function analysis. The results indicated that the subtests had a sensitivity rate of 90.6% and a specificity rate of 90% both of which are above the minimum required for a good screening test.

The second study evaluated the validity of the adaptive version of the LADS Plus by examining the correlations between the test and other established assessment and attainment measures. Scores on the subtests, as well as the combined scores, were correlated with participants' scores on the Woodcock-Johnson Word Attack Test (a test of phonological decoding skills using non-words), the Spelling subtest from the Wide Range Achievement Test III (WRAT III), and the Digit Span subtest from the Wechsler Adult Intelligence Scale III (WAIS III). Participants were 48 university students with no known history of dyslexia, 19 males and 29 females. The Word Recognition subtest did not correlate with any of the established measures but was correlated to the combined scores for the test $r = .62$. The other two subtests and the composite scores (the amalgamation of all scores on the three subtests) were significantly negatively correlated with the established measures, with correlations ranging from $r = -.33$ to $r = -.58$. Higher scores (greater risk of dyslexia) on the Word Construction and Working Memory subtests as well as the composite score were associated with low scores on the established measures. The Word Construction subtest had its highest correlation with the Digit Span subtest of the WAIS and not with the measures assessing similar constructs (such as the Wood-cock-Johnson Word Attack Test) suggesting that it may have more in common with this test than the other established measures. The correlation of the Working Memory subtest of the LADS Plus with the established measure of a similar construct, the Digit Span subtest of the WAIS III, was $r = -.58$, which is lower than the level required to establish concurrent validity $r = .75$ (Kline, 2000).

In the final study, analysis of variance, was used to compared the performance of university students matched on intelligence (19 with dyslexia and 19 without dyslexia) on the LADS Plus subtests and the combined scores. There were significant differences in the performances of the groups on the Word Recognition and Word Construction subtests, and on the composite scores, with the dyslexic group obtaining higher scores. The performance of the groups did not differ on the Working Memory subtest. For this subtest, these results are contrary to the results of the first validation study reported earlier and suggest that the subtest may not be a reliable discriminator of dyslexia in adults.

2.5.3 Independent Research

As previously reported in the sections reviewing the DAST and the BDT, Nichols et al. (2008) compared the effectiveness of the LADS and a *tutor screening battery* in detecting learning disability (Dyslexia, Dyspraxia and Meares-Irlen Syndrome). The tutor screening battery consisted of eight subtests of the BDT and seven subtests from the DAST. The researchers concluded that the tutor screening battery was more effective with a higher sensitivity rate (91%) than the LADS (66%) as this meant that the battery would be able to identify a greater proportion of individuals with learning disabilities than would the LADS. The LADS, however, had a higher specificity rate (90%) than the tutor screening battery (79%), indicating that it was better at identifying individuals without disabilities. The specificity rate of the LADS in this study (90%) is similar to that reported by the developers of the LADS (91%); however, the sensitivity rate (66%) is much lower than that reported by the developers (90%). The results of this study indicate that the LADS may be able to correctly identify individuals without dyslexia, but its ability to correctly identify individuals with the disorder is below an acceptable level.

2.5.4 Summary and Conclusion

The LADS Plus is an adaptive computerised dyslexia screening test designed to screen for dyslexia in individuals aged 15 years and older. The adaptive nature of the test, which can significantly reduce administration time, is very useful as long as the battery is reliable and valid. The manual for the LADS Plus is comprehensive and provides information on the development of the test, administration, scoring, and interpretation of results. Norms were established for both the Verbal and Non-verbal Reasoning subtests; however, this was on a sample of children, and no norms for adults were reported. Additionally, no information on norms for the other dyslexia-sensitive subtests was provided and therefore the adequacy or otherwise of these cannot be assessed. The overall reliability of the LADS Plus was not reported; however, the internal consistency reliability of two of the five subtests, Word Recognition and the Word Construction, was excellent. Overall, the validation studies conducted by the developers indicated that the subtests, as well as their combined scores, discriminated between adults with and without dyslexia. However, for the Working Memory subtest, the results of the validation studies were conflicting, with the first study indicating that it could discriminate between adults with and without dyslexia, while the second study indicated that it could not. The sensitivity and specificity rates reported from the validation study for the overall test are good and above the minimum required for a good screening test. Also, independent research (Nichols et al., 2009) concurred with the specificity rate of the LADS Plus reported in the validation study, but not with the sensitivity rate, which was lower than reported. Although current studies indicate that the LADS Plus may be able to discriminate between adults with and without dyslexia more research is required to determine if its capacity is at an acceptable level. In addition, more research is required to effectively evaluate its psychometric properties.

2.6 QuickScan

2.6.1 Test Description

QuickScan is part of the StudyScan Suite developed by Dr. Dorota Zdzienski as a computerised multi-functioning screening questionnaire and assessment test designed to identify individuals at risk of dyslexia, assess for dyslexia, identify individual learning styles, and assess attainment in cognitive and educational areas (Pico Educational Systems Limited; Zdzienski, 1998). The StudyScan Suite is based on the Scholastic Abilities Test for Adults (designed to measure the scholastic competence of individuals 16 to 70 years) and on the research conducted by Dr. Zdzienski for her doctoral thesis and it consists of two parts, QuickScan and StudyScan (Zdzienski, 1998). StudyScan is the assessment element of the test and QuickScan is the screening. QuickScan is designed as an individual learning styles, study skills, and dyslexia questionnaire (Zdzienski, 1998). It is designed to identify indicators of dyslexia, preferred learning style, and appropriate study skills, and to indicate whether a full dyslexia assessment is required. It was designed for use with individuals of 14 years and older, consists of 110 items, and is based on the Adult Dyslexia Checklist (Vinegrad, 1994). QuickScan, although it was designed to identify individuals at risk of dyslexia, unlike the other tests reviewed, is not a test but a questionnaire. The developer in her thesis gave no indication of the theoretical basis for the questionnaire. It includes questions on learning style, assessment history, laterality, incidence of dyslexia in the family, spelling difficulties, sequencing, and memory. It produces a report that details learning preferences, study styles, and personalised study guidelines (Haslum & Kiziewicz, 2001). The test distinguishes five categories of risk of dyslexia as follows: (1) no indicators, (2) borderline indicators, (3) some indicators, (4) many indicators, and (5) most indicators. It is self-administered, and is said to take approximately 10 to 20 minutes to complete.

2.6.2 Test Development

Norms. No information on the development of norms for the QuickScan was reported in Dr. Zdzienski's thesis, nor were we able to find this elsewhere.

Reliability. Dr. Zdzienski in her thesis reported that the administration of QuickScan by the computer was reliable $r = .90$. However no information on how this was evaluated was given. In addition, no other information on the reliability of the QuickScan was included in her thesis nor were we able to find any information elsewhere on the reliability the test.

Validity. The validity of QuickScan was assessed by comparing the scores of adults with and without dyslexia (Zdzienski, 1998). Sixty students from the universities of Leicester and Ulster (30 with dyslexia and 30 without dyslexia) participated in the study. The scores obtained by the groups differed significantly, with the dyslexic group obtaining higher scores. The results suggest that scores on QuickScan were capable of discriminating between adults with and without dyslexia.

2.6.3 Independent Research

Research conducted by Sanderson (2000) examined the test re-test reliability of QuickScan and StudyScan. A small sample of 10 undergraduate students from the University of Nottingham participated in the study. Two participants were assessed using QuickScan three times, after intervals of a few weeks, three were assessed twice and the remaining five once. The results of the assessment of learning style for all five participants (who repeated the test) produced by QuickScan were inconsistent, with different styles identified at each testing. Also, and more worryingly, for four of the five participants, their dyslexia categorisation and the severity of their difficulties were evaluated differently between tests. Sanderson (2000) concluded, on the basis of this very small study, that improvements were required to the StudyScan Suite, but noted that some improvements had already been made to a revised version of the StudyScan Suite after the completion of the study. Although the sample size of the study was too small to evaluate the reliability of the QuickScan, the results are nevertheless cause for concern about its reliability.

In another study, Haslum and Kiziewicz (2001) validated the StudyScan Suite by examining the relationship between QuickScan and StudyScan results and their ability to predict the results of subsequent educational psychologist assessments. The study was part of a collaborative research project between the Universities of Bath, Bristol, and the West of England, and was funded by the Higher Education Funding Council. Data for the study were collected over an 18-month period from 126 students of the universities. Students were first assessed on QuickScan and then on StudyScan if this was recommended by the first assessment. Additionally, some students, who were not recommended for further assessment based on their QuickScan results, were asked to complete the StudyScan assessment if their personal learning history indicated difficulties associated with dyslexia. Of the 126 students 91 (72.2%) were classified as showing some indicators of dyslexia and were recommended to undertake the StudyScan assessment. Of this 91, 66 agreed to also undertake the StudyScan assessment and, in addition, a further 19 students whose screening outcome indicated no risk of dyslexia were also asked to complete it. These 19 students were similarly classified by StudyScan as having no indicators of dyslexia. However, only 28 (42%) of the 66 students classified by QuickScan as showing indicators of dyslexia were given the same classification by StudyScan. Forty participants (of the 66) were assessed by six educational psychologists. Of these participants, 16 were assessed as showing indicators of dyslexia by StudyScan, and only 9 (56%) were given a similar assessment by the educational psychologist, indicating that the sensitivity and specificity rates may not be acceptable. The results suggest that the screening and assessment outcomes of QuickScan and StudyScan may not be reliable or valid.

The website for Pico Educational Systems Limited, the distributors of the StudyScan Suite, provides details of a number of reviews of the tests by individuals working with organizations that use the tests, including Nottinghamshire Dyslexia Association, Park Lane College, St Dominic’s Sixth Form College, St George’s Hospital Medical School, and West Kent College. Some of the reported advantages of the StudyScan Suite were: it is useful for assessing dyslexic students’ strengths and weaknesses; it enables tailored teaching programmes; it enables the organization to provide better guidance to students; and it is student friendly. Additionally, Isabel Martin, a Dyslexia Tutor at West Kent College reported that the StudyScan assessment was very similar to that of an educational psychologist, although this is at odds with the results of the Haslum and Kiziewicz (2001) study. Disadvantages reported were: it requires a high comprehension level; it contains ambiguous questions; it produces lengthy outcome reports; and that StudyScan was too long. On balance, the reviews were favourable, but did not provide any empirical evidence to support the views expressed (http://www.quickscn.com/studyscn/reviews.htm).

2.6.4 Summary and Conclusion

QuickScan is a computerised multi-functioning screening questionnaire designed to identify individuals at risk of dyslexia, individual learning needs and study styles. There is no indication that norms for the questionnaire were established. The one validation study reported indicates that it can discriminate between adults with and without dyslexia. However, independent research indicates that its screening outcomes may be unreliable (Haslum & Kiziewicz, 2001; Sanderson, 2000). StudyScan has also received a number of favourable reviews from individuals/organizations; however, these are based mainly on anecdotal evidence and no empirical evidence has been presented to support the reviews. Although there has been some limited research into using questionnaires to identify individuals with dyslexia and the results of these studies, indicate that this method of identification may be valid (Kasler & Fawcett, 2009; Snowling, Dawes, Nash, & Hulme, 2012; Wolff & Lundberg, 2003), much more research is needed to evaluate the validity of this method. With the exception of the studies undertaken by the developer, we found no independent studies that have provided empirical evidence to support the validity of the StudyScan Suite as dyslexia screening test.

2.7 York Adult Assessment Battery-Revised

2.7.1 Test Description

The York Adult Assessment Battery (YAA-R) is a paper-based assessment battery designed as a screening test for young adults with dyslexia (Warmington, Stothard, &

Snowling, 2013). It was also designed to assess the skills required by students involved in study at further and higher education levels. It was developed by Warmington et al. (2013) and is a revision of the York Adult Assessment (YAA) previously created by Hatcher and Snowling (2002). The YAA consists of six subtests namely: Nonsense Passage Reading, Spoonerisms, Writing Speed, the Spelling subtest of the WRAT III, Timed Précis, and Proof Reading. It was based on an assessment protocol developed to assess the difficulties experienced by dyslexic adults and was evaluated in a study conducted by Snowling, Nation, Moxham, Gallagher, and Frith (1997).

For the YAA-R, three of the tasks from the YAA were excluded: Nonsense Passage Reading, Spelling subtest of the WRAT III, and Proof Reading, and two tasks assessing reading comprehension and rapid automatized naming (RAN) were added. The revised battery therefore consists of five subtests namely: Reading Comprehension, Written Précis (formerly Timed Précis), Writing Speed, Spoonerisms, and RAN. Further details of the subtests are provided in Table 2.6 below. Two versions of RAN digits and objects were included. The tasks from the YAA that were retained (Written Précis, Writing Speed, and Spoonerisms) were revised. The Writing Speed task was revised by including a typing speed portion in addition to the writing speed task; also, the sentences used in the task were different from the one used in the previous Writing Speed task. The Spoonerisms task, which used the names of well known British and international personalities, was revised by changing the names to contemporary ones. For example, Joanna Lumley was removed and Wayne Rooney added. Four different scores were calculated from the Reading Comprehension task: (1) comprehension – the total of the correct responses to the 15 items on the tasks; (2) reading accuracy – the total number of words read correctly from the reading comprehension passage; (3) reading time – the total time taken to read the passage; and (4) reading rate – the total number of words read per minute. Similarly, four different scores were calculated from the Written Précis tasks: (1) content – the total score for summarising the passage read; (2) time – the time taken to complete the task; (3) précis rate – writing speed, the number of words written per seconds; (4) spelling error rate – the total number of words spelt incorrectly as a proportion of the total number of words written. For the spoonerism task, an accuracy score was calculated as well as a rate score, the total time for correct items divided by total correct items. The developers do not explicitly state the theoretical basis for the battery; however, it includes tasks assessing literacy and phonological processing skills on which adults with dyslexia are known to exhibit deficits compared to their counterparts without dyslexia.

Table 2.6. Description of the Subtests of the York Adult Assessment Battery-Revised

Subtests	Description	Skills Assessed
Reading Comprehension	Tests three components of reading comprehension: knowledge, vocabulary, and inference making. Also assesses reading accuracy and reading rate	Comprehension, reading accuracy and rate
Written Précis	Tests the ability to summarise and write with time constrains, and also assesses writing rate and spelling error rate	Reading comprehension, ability to summarise, writing rate, and spelling
Writing Speed	Tests hand writing and typing speeds	Writing and typing
Spoonerisms	Tests the manipulation and production of phoneme sounds	Phonological awareness
Rapid Automatised Naming	Tests phonological processing speed	Speed of lexical access and articulation

Details of how scores on the tasks should be used to assess an individual's risk of dyslexia are not stated in the battery which can be freely downloaded from the internet or in the published study which reports the validation of the battery (Warmington et al., 2013). However, normative percentile scores are provided from a sample of 106 university students with no history of reading difficulties. Norms for the 5th to 95th percentile, along with means, standard deviations and 95% confidence intervals are included in the battery. Administration time for the test is approximately 30 minutes.

2.7.2 Test Development

Norms. Norms were established for both the YAA and the YAA-R. For the YAA, norms were established from a sample of 50 undergraduate students at York University with no reported history of literacy difficulty (Hatcher, Snowling & Griffiths, 2002). The sample consisted of 16 males and 34 females, mean age 21 years 8 months and range 18 to 41 years. The norms are percentiles for the 10th to 90th percentile, along with means and standard deviation scores. Participants were also assessed on the Reading and Spelling subtests of the Wide Range Achievement Test III (WRAT III) and obtained scores in the average or above average range. Norms for the YAA-R were established from a sample of 106 adult students from several universities and sixth form colleges in the United Kingdom. The participants in the sample reported no history of reading difficulties and consisted of 75 males and 31 females, for a gender ratio approaching three male to one female. This gender ratio was selected by the developers to match with that of the dyslexic sample (details below) (Warmington et al., 2013). Mean age of the sample was 21 years 10 months, and range 18 years to 36 years. The ethnicity of the sample was White = 92 (87%), Asian = 11(10%), and Black = 3 (3%). Participants were also assessed on the Reading and Spelling subtests of the Wide Range Achievement Test III (WRAT III), the Block Design, and Vocabulary subtests of the Wechsler Abbreviated Scale of Intelligence (WASI), and the Brown Attention Deficit Disorder Scales (Brown ADD). The Brown ADD is a self-report scale that measures symptoms of attention deficit disorder and the mean score of participants on it was below a clinically significant level. For the standardised measures, participants performed at the upper end of the average range. No manual for the batteries is available; however, they both contain instructions for the administration and scoring of the tasks included.

Reliability. The internal consistency reliability for two of the tasks in the battery, Reading Comprehension and Spoonerisms, was assessed using the normative sample (Warmington et al., 2013). For the Reading Comprehension task, reliability was reported for the total score on the task as well as a reading accuracy score calculated from the task by subtracting the total number of errors made from the total number of words in the passage read. The reliability for reading accuracy $\alpha = .81$ and the Spoonerisms tasks $\alpha = .76$ was good and above the minimum recommended .70 (Kline, 2000). For the Reading Comprehension task, reliability was α = .53, which is well below the minimum recommended and suggests that the items may not be consistent and may not be measuring the same underlying construct. The developers attributed the low reliability of this task to the small number of items in the tasks (15) and the heterogeneity of the construct assessed. The reliability of the other tasks in the battery was not assessed (Warmington et al., 2013).

Validity. The validity of the YAA-R was assessed by: (1) correlating the scores on the tasks to scores on standardised measures of literacy and vocabulary; (2) comparing the performance of adults with and without dyslexia on the tasks in the battery; and (3) examining its sensitive and specificity rates. The concurrent validity of the YAA-R was assessed by correlating scores of the participants (normative sample) on the tasks in the battery to their scores on the Reading and Spelling subtests of the WRAT III and the Vocabulary subtest of the WASI. With the exception of the RAN tasks, scores on all the other tasks in the battery were significantly correlated with the standardized scores with low, r = .24, to moderate, r = .55, correlations. Twenty adults without dyslexia were selected from the normative sample of 106 adults and their performance on the tasks in the YAA-R was compared to that of 20 adolescents and adults with dyslexia. The dyslexic sample consisted of 16 males and 4 females, mean age 22 years 5 months, range 15 to 31 years. Participants in the dyslexic sample were recruited from the University of York and York College; all had a formal diagnosis of dyslexia, and were matched on general cognitive abilities with the adults without dyslexia. The performances of the groups differed on all the measures with the exception of the Written Précis time, Spoonerism rate and RAN object rate, with the dyslexic group performing significantly less well than the controls. Cohen's d effect sizes were calculated and indicated moderate d = 0.70 to large d = 1.61group differences. Discriminant analysis (logistic regression) with all the participants (126, normative and dyslexic samples) indicated that the YAA-R correctly classified 119 participants for an overall classification of 94%. It correctly classified 16 of the 20 dyslexics for a sensitivity rate of 80%, and 103 of the 106 controls for a specificity rate of 97%, which are both above the minimum required 80% and 90% respectively for a good screening test.

2.7.3 Independent Research

The YAA-R is a newly developed screening test and we did not find any published independent study on the battery. We also did not find any published independent study of its predecessor the YAA.

2.7.4 Summary and Conclusion

The YAA-R is a paper-based assessment battery designed to screen for dyslexia in adults. There is no manual, and instructions for the administration of its tasks are included in the battery. Norms for the battery were established with a sample of 126 adults from different educational institutions; however, the adequacy of the norms is questionable. Norms should be obtained from a representative sample and the sample should be sufficiently large to minimize standard errors and provide stable values (Anastasi & Urbina, 1997; Kline, 2000). A sample size of approximately 500 is recommended by Kline, (2000) as adequate for minimizing standard errors. The small size of the normative sample for the YAA-R suggests that it may lack in representativeness, and its scores may not be stable. In addition, the instructions for the battery do not indicate how the normative scores should be used to identify individuals at risk of dyslexia, as no clinical cut-off points are suggested. The internal consistency reliability of the Reading Comprehension task is below an acceptable level and the reliability of three of the tasks was not assessed. The results of the validation study indicated that the tasks in the battery could discriminate between adults with and without dyslexia. Also, its sensitivity rate (80%) is at the minimum recommended and its specificity rate is excellent (97%) and well above the minimum recommended. Although the results of the validation study for the YAA-R indicated that the tasks in the battery could discriminate between adults with and without dyslexia, the reliability of the battery as a whole and three of the tasks included was not assessed. As reliability is essential for validity, it is not possible to assess the validity of the battery as a whole.

2.8 General Overview of Dyslexia Screening Tests for Adults

One very important issue highlighted by this review is the limited number of independent published studies evaluating the adult dyslexia screening tests available in the

United Kingdom. This is despite the fact that most have been around for several years and are being widely used in a number of educational institutions and other organizations. In addition, for some of the tests, empirical evidence of their validity and reliability is weak. There is therefore a need for further research on these tests, particularly on their psychometric properties, that will provide information to enable adequate assessment of them. This should include comparative studies, where two or more screening tests are examined, and their effectiveness is compared. This will provide educational institutions, service providers and other institutions with the information necessary to make better and more informed decisions regarding adult dyslexia screening tests.

2.9 Aims of the Thesis

Our limited review of six dyslexia screening tests for adults currently used in the United Kingdom highlighted the need for more research on these tests, specifically research to determine their validity and effectiveness. One of the aims of this thesis is to help fill this gap in existing research by conducting an evaluation of the psychometric properties of two of the tests, the BDT and the Instines, and to examine their suitability to screen for dyslexia in adults. This will provide current and potential users of the tests with empirical evidence that can be used to objectively evaluate the tests. Also, given the limited number of available dyslexia screening tests for adults, and the inadequacies revealed in some of the measures, another aim of the thesis is to create a new measure, with a strong theoretical framework, based on existing research evidence that is psychometrically sound and effective. The final aim of the thesis is to contribute to our current understanding of dyslexia in adults by investigating the profile of cognitive predictors of literacy skills for adults with and without the disorder. Additionally, to contribute to and assist in resolving the long standing debate on the role of IQ in the definition of dyslexia and the identification of individuals with the disorder, we will specifically examine the effects of IQ on the literacy skills of two groups of adults with dyslexia with high and low abilities.

Chapter 3: Evaluation of the Bangor Dyslexia Test (BDT) for use with Adults (Study 1)

3.1 Introduction

The BDT has been in use in the United Kingdom for over 30 years and was one of the first dyslexia screening tests for adults. Although information on the current usage of the BDT in clinical and educational settings is not available, we are aware that the test is still being used to screen adults for dyslexia. It is an integral part of the dyslexia screening process for students studying at Bangor University and is also used by the University of Worcester (Nichols, McLeod, Holder, & McLeod, 2009). A unique feature of the BDT, setting it apart from other dyslexia screening tools, is its emphasis on quick and easy-to-administer tasks that do not directly assess reading and spelling skills. The BDT subtests were conceived as more distal markers of the array of difficulties in the oral language domain that may underlie dyslexic's literacy difficulties.

In our review of the test in Chapter 2, we indicated that the development studies for the test were conducted with children only. Additionally, we have been able to identify only two independent research studies of the BDT, one of which included adult participants. However, in the study which included adults conducted by Nichols et al., 2009), a proper evaluation of the BDT was not possible as the researchers combined the BDT with the DAST (Nicolson & Fawcett, 1998). In the second independent study, Sutherland and Smith (1991) compared the BDT to two other dyslexia screening tests, the Boder Test of Reading-Spelling Patterns and the Aston Index, on their ability to group participants into sub-types of dyslexics (auditory, visual and mixed), and on ease of administration for teachers assessing secondary school children (mean age 11.5 years) with literacy difficulties. Worryingly, the screening outcomes across the three batteries were inconsistent, and coincided for only six of the 20 participants. The BDT was deemed insufficiently proscriptive about the interpretation of test scores and hence too general to be useful for teachers. Cognizant of the current paucity of psychometric information about the battery, in particular for use with adult populations, in the present study, we assessed whether its psychometric properties are adequate for use with an adult student population. We made use of the archival database of screening and full assessment outcomes of students at Bangor University. Specifically, we investigated its (construct and predictive) validity and its internal consistency reliability. We expected that if the non-literacy subtests of the BDT are valid indicators of literacy difficulties in dyslexia, then we should find performance on them to: (1) discriminate between dyslexic and non-dyslexic populations, and (2) predict individual and group differences in performance on measures of literacy skills, and (3) predict dyslexia status (i.e., dyslexic versus non-dyslexic group membership). Moreover, if the BDT is an adequate assessment tool for use with adults, we expected the students' data on each subtest and on the battery as a whole to yield robust estimates of reliability.

3.2 Method

3.2.1 Participants

Three groups of participants were included in the study. Two groups of participants were selected from the Miles Dyslexia Centre's archived data: a dyslexic and an at-risk sample. A third control group was recruited from Bangor University students who had no history of learning difficulties. Each group is described below.

3.2.1.1 Archival data

Data were obtained from the archived records of students who were screened and assessed at the Miles Dyslexia Centre of Bangor University during the period September 2004 to October 2008. Students accessed the Centre's services on a voluntary self-referral basis and data were collected and stored electronically for those who had given written consent for their data to be used for research purposes. During this period, data were collected from 373 students. They were registered in degree courses across the five academic colleges within the University and were engaged in studies in a wide range of disciplines including Psychology, Nursing, Sports Science, Zoology, Marine Biology and Social Work.

The majority of students were undergraduates 327 (90%) and 37 (10%) were postgraduates and 325 (87%) were first language English speakers.

3.2.1.1.1 *Dyslexic sample***.**

Of the original sample, 348 students were referred for a full assessment as their screening outcomes indicated that they were at risk of dyslexia or other learning disorders. Of this referred group, 230 undertook the full assessment, and 193 (i.e., 52% of the initial screened sample, and 84% of the referred-and-followed-up sample) were diagnosed as dyslexic and were included in the dyslexic group. Thirty seven (16%) were not diagnosed with dyslexia; however of this number most (33) were diagnosed with other learning disorders and 4 with no disorders. The demographic characteristics of the dyslexic group are detailed in Table 3.1, and their scores on general ability and literacy measures are presented in Table 3.2.

Table 3.1. Participants Characteristics by Sub-samples

Characteristics	Sub-samples for Total Data Set				Sub-samples for At Risk Group		
	Full Archival	Dyslexic	At Risk	Control	Screened Only	Assessed with Other Difficulties	Assessed Not Dyslexic
N	373	193	180	40	118	25	37
Age							
M	25.41	24.5	26.39	19.5	26.53	24.76	27.05
SD	9.44	8.76	10.06	1.81	9.36	9.30	12.58
Gender							
% Male	31.9	31.6	32.2	25	35.6	28	24.3
% Female	68.1	68.4	67.8	75	64.4	72	75.7

3.2.1.1.2 *At-risk sample*.

This group was included in the study to assist with the assessment of the capacity of the BDT to accurately identify individuals with dyslexia. Also, to increase the heterogeneity of the sample and to reduce restricted range effects. It consisted of 180 students (of the original 373) and included students whose screening indicated a risk of dyslexia or other learning disability, but who did not undertake a full assessment (n = 118), students whose screening indicated no risk of dyslexia or other learning disability (n = 25), and students who on screening were assessed as being at risk of dyslexia or other learning disabilities and were later diagnosed with dyspraxia, other learning and memory difficulties, or no learning difficulties (n = 37) (see Table 3.1). We expected that the students screened at risk would perform similarly to the confirmed dyslexics and those screened not at risk to be comparable to the controls.

3.2.1.2 Control sample.

Control participants were 40 undergraduate psychology students recruited through a student participation panel. All had English as their first language and no history of learning difficulties; they were compensated with course and printer credits. The characteristics of this sample and their scores on nonverbal and verbal ability and literacy measures are detailed in Tables 3.1 and 3.2, respectively. The age difference between the groups was statistically significant $F(2, 409) = 9.91$, $p < .001$, with the control group differing from the other two groups.

3.3 Procedure

3.3.1 Screening Procedure for the Dyslexic (Archival) Sample

All students were screened individually by staff at the Miles Dyslexia Centre. All assessors were trained professionals with qualifications in assessing and teaching individuals with Specific Learning Difficulty (SpLD). The screening procedure included: (1) a *semi-structured interview* probing information about prior and current difficulty experienced in their studies, as well as general background information including, medical history, primary and secondary school experience, academic achievements and any post-secondary experience including educational or work activities; (2) the Bangor Dyslexia Test, administered and scored in accordance with the instructions in the manual; (3) a timed (3-minute) free writing test used to assess the students' writing speed; and (4) four subtests of the Dyslexia Adult Screening Test (Nicolson & Fawcett, 1998), as follows: Nonsense Passage Reading - a short passage of real and nonsense words to be read aloud for a maximum of 3 minutes. Scoring was based on reading accuracy and speed. The reported test-retest reliability is $r = .92$. Two Minute Spelling - up to 32 words graded in difficulty, which are spelled to dictation for 2 minutes. The reported test-retest reliability is $r = .93$ and is used to assess spelling skills, accuracy, automaticity and fluency. Phonemic Segmentation - includes 12 segmentation items (requiring deletion of syllables or consonants from words) and 3 spoonerism items (requiring transposition of initial phonemes in pairs of words). The reported test-retest reliability is $r = .90$ and is used to assess phonological skills. Verbal and Semantic Fluency - require the rapid generation of words on the basis of either alliteration or meaning for a duration of one minute. The reported test-retest reliability is $r = .81$ for Verbal Fluency and $r = .76$ for Semantic Fluency. The DAST subtests were administered and scored according to published guidelines. Note that the present study aims to evaluate only the psychometric properties of the BDT, and not the validity of the full procedure undertaken as part of the screening process.

3.3.2 Full assessment Procedure for the Dyslexic (Archival) Sample

Students whose screening outcomes indicated that they were at risk of dyslexia or other learning disabilities were referred for further assessment. For the period of data collection (September 2004 to October 2008), with the exception of two full assessments that were undertaken by qualified Specialist Teachers, all were carried out by Educational Psychologists. Students were assessed on a battery of tests that included subtests of the Wechsler Adult Intelligence Measure III (WAIS III) (Wechsler, 1997), Wechsler Individual Achievement Test II (WIAT II) (Wechsler, 2005), Wide Range Achievement Test III (WRAT III) (Wilkinson & Robertson, 2004), and the Wechsler Objective Reading Dimensions (WORD) (Wechsler, 1993). The WAIS III subtests usually administered were Vocabulary, Block Design, Verbal Comprehension, Perceptual Organization, Working

Memory and Processing speed. Literacy attainment (reading, spelling and reading comprehension) was assessed using the WIAT II, WORD, or WRAT III.

3.3.3 Testing Procedure for Control Participants

All students were tested individually in a quiet room in a session lasting approximately 60 minutes. Students were assessed on the BDT as well as on measures of literacy attainment using the Word Reading, Spelling, and Sentence Completion subtests of the Wide Range Achievement Test (WRAT IV) (Wilkinson & Robertson, 2004) and, verbal and non-verbal ability using the Vocabulary and Matrices subtests of the Wide Range Intelligence Test (WRIT) (Glutting, Adams, & Shelow, 2000). Although the dyslexic and control groups were assessed on different background measures, comparison of their performances was possible as all are well established and standardised, and are widely used for research and assessment purposes. In addition, the manuals of the WRAT IV and the WRIT report significant moderate to high correlations of the subtests in their measures and subtests of the WIAT II and WAIS III that were used with the dyslexic group, providing evidence of acceptable convergent validity. Ethical approval for the study was granted by the School of Psychology, Bangor University.

3.4 Data Analysis

Prior to analyses, the data set was checked for outliers and outliers above or below 2.5 standard deviations from the mean were adjusted to scores corresponding to 2.5 standard deviations. To identify outliers the mean for each group was calculated separately. For the control group only one score was adjusted on the spelling task and for the at risk group one score on the nonverbal task. For the dyslexic group, two scores on the BDT, and one on the reading and the spelling tasks were adjusted. We used multivariate analysis of variance (MANOVA) followed up with univariate analyses (ANOVA) with Bonferroni-adjusted alpha level to compare the groups on the background measures. The data set was checked for appropriateness for multivariate analysis (MANOVA) and this revealed that the assumption of equality of error variances was violated; therefore, a more conservative alpha of .01 was used for significance as recommended by Tabachnick and Fidell (2007). To assess the capacity of the BDT to discriminate between adults with and without dyslexia, we compared the performance of the groups on the subtests using Mann Whitney *U*, as the scores were not normally distributed and the battery as a whole using *t*-test. Effect sizes, using Cohen's *d* formula, were calculated to determine the magnitude of any group difference. Additionally, logistic regression analysis was also used to further assess the capacity of the BDT to discriminate between adults with and without dyslexia. The data were checked for the assumptions of logistic regression and these were satisfied. In the next section, we first report the results of the group comparison on the background measures. This is followed by the results of the assessment of the internal consistency reliability of the BDT. Finally, we report the validity assessment of the BDT, firstly comparing the performances of the groups on its subtests and the battery as a whole, and secondly examining its capacity to predict group membership.

3.5 Results

We examined the group characteristics on background measures of verbal and non-verbal abilities and literacy abilities using MANOVA, where *groups* was the independent variable and the *background measures* were the dependent variables. For this analysis, the sample size for the dyslexic group was reduced from 193 to 97 because of missing data from the other 96 dyslexic participants. This missing data were due to differences in the number of tests administered to the participants during the assessment procedure. The dyslexic participants were assessed by different individuals who did not always administer the full battery of tests available. Therefore, only those dyslexic participants who were administered the full battery of tests are included.

3.5.1 Comparisons of Performance of Dyslexic and Control Groups on Background Measures.

The descriptive statistics for the background measures along with Cohen's *d* effect size are reported in Table 3.2. The results indicated a significant difference between the groups on the set of measures $F(5, 131) = 27.47$, $p < .001$; Pillai's Trace = 0.51. Follow up ANOVA (with Bonferroni adjusted alpha level of .01) revealed that the groups' performances were significantly different on all the measures except verbal ability $F(1, 135) = 5.26$, $p = .023$ measured by the relevant WAIS or WRIT subtests. For all the measures, the dyslexic group achieved lower mean scores than the control group (although only marginally so on verbal ability), however, all scores were well within the average range, as would be expected with a population of university students. On the literacy measures, although the standard scores for both groups fell within the average range, the effect sizes for reading accuracy and spelling were very large d = 1.86 and d = 1.59 respectively. The exact magnitudes of these effects should be interpreted with some caution as performance estimates were derived from different (standardized) batteries. Nevertheless, it is clear that the dyslexic sample experienced significant literacy difficulties relative to their cognitive abilities and relative to the control group.

Table 3.2. Mean Standard Scores (Standard Deviation in Parenthesis) of Dyslexic and Control Groups on Background Measures (N =137)

Measures	Dyslexics (n = 97)	Controls (n = 40)	Cohen's d
Nonverbal Ability	103.04 (13.44)[a]	110.20 (8.89)[b]	0.64
Verbal Ability	105.44 (12.28)[c]	110.40 (8.93)[d]	0.47
Reading	95.94 (10.75)	114.82 (9.57)	1.86
Spelling	94.67 (12.68)	114.72 (12.59)	1.59
Comprehension	95.94 (13.93)	105.55 (7.82)	0.88

Note. For the dyslexic group literacy was assessed with one of the following: WIAT II, WRAT III, or WORD. For the control group literacy was assessed with the WRAT IV.

[a]Score derived from Block Design subtest of the WAIS III.

[b]Score derived from Matrices subtest of the WRIT.

[c]Score derived from Vocabulary subtest of the WAIS III.

[d]Score derived from Vocabulary subtest of the WRIT.

As reported earlier the age difference between the groups was statistically significant with the control group differing (i.e., being slightly younger) from the other two groups. Age is considered in the scoring for the standardised measures and therefore it is unlikely that the age difference affected the results. For the BDT, we examined the correlations of age with the participants' scores on each subtest as well as the battery as a whole. There were no significant correlations, for the battery as a whole $r(412) = .08$, $p = .100$. It is therefore, unlikely that the age differences between the control group, and the other groups, will have affected the results obtained in the analyses conducted.

3.5.2 A Validation of the BDT

3.5.2.1 Comparison of performance of dyslexic and control groups on the BDT.

Our first a priori hypothesis was that the dyslexics would perform less well than the controls on the BDT (obtaining higher total scores) overall as well as on each subtest. Descriptive analysis revealed that the BDT total scores for the groups were normally distributed. However, the subtests because of the scoring method used, were not. The groups' scores on the overall battery were compared using independent sample *t*-test and scores on the subtests using Mann-Whitney *U* test. Beginning with the BDT total scores, the groups differed significantly with the dyslexics (M = 6.17, SD = 1.44) attaining higher index scores than the controls (M = 2.09, SD = 1.23), $t(231)$ -16.69, $p < .001$. Moreover, the groups differed significantly on each subtest of the BDT, with the dyslexics performing less well (higher scores) than the controls (see details in Table 3.3). The magnitude of the between-group effects was estimated by Cohen's *d*. For the total measure, the effect size d = 3.06, indicated that the performance difference between groups was not only significant but was also very large. The effect sizes for most subtests were also large, ranging from d = 0.87 for Digits Forwards to d = 1.70 on the Subtraction and Familial Incidence subtests. No effect size calculation was done for the Months Forwards and B-D Confusion subtests as the control group performed at ceiling on these subtests with mean scores of zero. The performance of the control group was also near ceiling on the Polysyllables, Subtraction, Months Reversed, and Familial Incidence subtests.

Table 3.3. Results of Mann-Whitney U Test of Difference in Mean Ranks of Performance of Dyslexic and Control Groups on the BDT Subtests, and Effect Sizes (N = 233).

	Dyslexics (n = 193)		Controls (n = 40)				
BDT Subtests	Mean Rank	Mean score (SD)	Mean Rank	Mean score (SD)	Mann-Whitney U	Z	Cohen's d
Left/Right	125.88	.76 (.33)	74.16	.45 (.37)	2146.50	-4.94***	0.89
Polysyllabic Words	126.20	.47 (.41)	72.63	.13 (.22)	2085.00	-4.90***	1.08
Subtraction	128.79	.49 (.43)	60.10	.03 (.11)	1584.00	-6.35***	1.70
Tables	128.25	.85 (.30)	62.73	.41 (.41)	1689.00	-6.76***	1.24
Months Forwards	119.38	.08 (.22)	105.5	.00 (.00)	3400.00	-2.29*	-
Months Reversed	125.81	.43 (.43)	74.48	.09 (.22)	2159.00	-4.79***	1.03
Digits Forwards	124.38	.85 (.35)	81.41	.48 (.50)	2436.50	-5.00***	0.87
Digits Reversed	126.86	.83 (.35)	69.43	.35 (.48)	1957.00	-6.16***	1.16
B-D Confusion	131.20	.64 (.44)	48.50	.00 (.00)	1120.00	-7.80***	-
Familial Incidence	130.60	.77 (.37)	51.38	.15 (.36)	1235.00	-7.73***	1.70

* $p < .05$. *** $p < .001$.

For a further insight into the nature of the between-group differences, the percentage of participants in the dyslexic and control groups who obtained BDT scores respectively of: 1 (i.e., positive indicator, at risk), 0 (i.e., negative indictor, not at risk), and .5 (i.e., marginal, doubtful) was also examined (Table 3.4). Overall, the percentage obtaining positive scores was higher for the dyslexic than the control group. In fact, no control participant obtained a positive score on four subtests: Polysyllables, Subtraction, Months Forwards and B-D Confusion, consistent with the group’s very low mean scores on these measures. In contrast, the dyslexic group obtained positive scores on all the subtests. For the controls, the subtests with the highest percentages of positive scores were Digits Forwards, Digits Reversed, and Tables. The dyslexic group also obtained the highest percentages of positive scores on these subtests; however, the prevalence percentages were two to three times higher. The majority (68.9%) of the dyslexic participants reported that other members of their family might be affected by similar difficulties compared to 14.5% for the controls. This response is in line with research confirming that dyslexia is highly heritable, with increased risk for individuals with a history of dyslexia among first-order family members (Byrne et al., 2009). Generally, the performance of the control group indicated minimal difficulty with the BDT, while the opposite was true for the dyslexic group.

Table 3.4. Percentages of Dyslexic and Control Participants Falling in each of the Outcome Categories of the Subtests of the BDT (N = 233)

	Outcome Categories		
Subtests	Positive	Marginal	Negative
		Dyslexics	
Left-Right	61.7	29.0	9.3
Polysyllabic Words	31.1	32.1	36.8
Subtraction	35.8	26.4	37.8
Tables	76.2	17.1	6.7
Months Forwards	3.1	8.8	88.1
Months Reversed	30.1	25.9	44.0
Digits Forwards	82.9	3.6	13.5
Digits Reversed	78.2	9.3	12.5
B-D Confusion	57.5	13.5	29.0
Familial Incidence	68.9	16.6	14.5
		Controls	
Left-Right	22.5	45.0	32.5
Polysyllabic Words	0.0	25.0	75.0
Subtraction	0.0	5.0	95.0
Tables	25.0	32.5	42.5
Months Forwards	0.0	0.0	100.0
Months Reversed	2.5	12.5	85.0
Digits Forwards	47.5	2.5	50.0
Digits Reversed	35.0	0.0	65.0
B-D Confusion	0.0	0.0	100.0
Familial Incidence	15.0	0.0	85.0

Note. Dyslexics n = 193. Controls n = 40.

Despite the differences in the performance of the dyslexic and control groups, there were some similarities. For both groups, the subtests associated with the highest number of positive indicators were the same, such that the Digits Forwards, followed by Digits Reversed and Tables subtests were most difficult regardless of group membership. In addition, both groups had the lowest mean scores and the lowest percentage of plus scores on the Months Forwards subtest. Notwithstanding these similarities, the quantitative pattern of the groups' performance on the subtests was different. While the dyslexics obtained high mean scores and high percentages of positive indicators on most of the subtest items, the opposite was true of the controls. To summarise, the Mann-Whitney *U* analysis indicated significant differences between the performances of the dyslexic and control groups on the BDT with the dyslexics obtaining consistently higher scores than the controls.

3.5.2.2 Correlational analysis.

The construct validity of the BDT was further examined by correlating the BDT total scores and the standardised measures of literacy for the dyslexic and control samples individually and collectively. Across analyses, there were significant negative correlations between the BDT scores and literacy measures, with high scores on the BDT being associated with lower scores on the measures for literacy. For the dyslexic group, correlations were low to moderate Reading $r = -.24$ $p = .001$, Comprehension $r = -.26$ $p = .002$, and Spelling $r = -.33$ $p < .001$. For the two groups combined (dyslexic and control) correlations were moderate Reading $r = -.51$ $p < .001$, Comprehension $r = -.39$ $p < .001$, and Spelling $r = -.55$ $p < .001$. Conversely, there were no significant correlations between the BDT scores and the literacy measures for control group. This may be attributed to the restricted range of the control group scores on the measures; in contrast, range effects were likely to account for the stronger correlations of the combined groups. Nevertheless, the moderate results of the dyslexic group indicate that although the BDT does not have items that explicitly assess literacy difficulty, performance on the BDT is associated with literacy difficulties and this further supports the construct validity of the BDT.

3.5.2.3 Discriminant analysis.

To further, assess the capacity of the BDT to discriminate between dyslexic adults and control participants and its validity as a dyslexia screening tool, logistic regression analysis was conducted. For the logistic regression, the total score obtained by participants on the BDT was the predictor (independent variable) and group membership (dyslexics or controls) the dependent variable. For this analysis the total scores on the BDT was used as predictor instead of scores on the subtests (categorical variables) as the sample size could not accommodate this, as the ratio of cases to predictors was inadequate. The results indicated that the model was statistically significant χ^2 (1, N = 233) = 147.34, p < .001, indicating that the BDT score distinguished between the dyslexic and control participants. The model explained a large amount of the variance in the groups .47 (Cox and Snell R^2) and .78 (Nagelkerke R^2), indicating that a significant proportion of the difference between the groups can be explained by the BDT scores. Also, the Hosmer-Lemeshow Goodness of Fit Test indicated that the model was a good fit χ^2 (8, N = 233) = 4.15, p = .843 for the data. This suggests that the model prediction was consistent with the data. As detailed in Table 3.5, the classification of the dyslexic group was more accurate than that of the control group. For the dyslexic group 96.4% were correctly classified (true positives) and 3.6% incorrectly classified (false negatives) while for the control group, the figures were 82.5% and 17.5%, respectively. The results indicate that the BDT can effectively identify dyslexic individuals with an excellent sensitivity rate of 96.4%, however its ability to identify non-dyslexic individuals (specificity rate 82.5%) was lower than the 90% minimum recommend by Glascoe and Byrne (1993). Despite the less than ideal specificity rate, the above analyses demonstrate that the BDT's overall ability to discriminate between adult dyslexics and controls is very good.

Table 3.5. Classification Result of Logistic Regression for Dyslexics and Controls Groups

Participants	Predicted Group Membership		
	Dyslexics	Controls	% Correct
Dyslexics	186	7	96.4
Controls	7	33	82.5
Overall %			94.0

The capacity of the BDT to correctly identify individuals with dyslexia was further explored by examining the classification of students who were screened for dyslexia and other learning disabilities, but not assessed. The at-risk sample (*n* = 180) consisted mainly of these students, that is students whose screenings indicated a risk of dyslexia or other learning disability but who had not undertaken a full assessment (*n* = 118) and students whose screening indicated no risk of dyslexia or other learning disability (*n* = 25). The mean scores of these subsamples on the BDT were *M* = 5.36 (*SD* = 1.77) for those screened dyslexic (screened dyslexic group), and *M* = 3.40 (*SD* =1.63) for those screened not-dyslexic (screened not-dyslexic group). Details of their scores on the BDT subtests are listed in Table 3.6.

Table 3.6. Results of Mann-Whitney U Test of Difference in Mean Ranks of Performance of Screened Dyslexic and Screened Not-dyslexic Groups on the BDT Subtests with Effect Sizes, Mean Scores and Standard Deviation in Parentheses (N = 143).

	Screened Dyslexic (n = 118)		Screened Not-Dyslexic (n = 25)				
BDT Subtests	Mean Rank	Mean (*SD*)	Mean Rank	Mean *SD*	Mann-Whitney *U*	*Z*	Cohen's *d*
Left/Right	77.66	.73	45.28	.38	807.00	-3.92***	0.91
Polysyllabic Words	76.13	.36	52.52	.12	988.00	-2.90**	0.68
Subtraction	73.17	.40	66.48	.34	1337.00	-.79†	0.15
Tables	74.64	.78	59.54	.62	1163.50	-1.96*	0.41
Months Forwards	72.69	.07	68.74	.02	1393.5	-.90†	0.29
Months Reversed	74.12	.30	62.00	.16	1225.00	-1.56†	0.39
Digits Forwards	74.30	.73	61.14	.56	1203.50	-1.75†	0.37
Digits Reversed	74.19	.73	61.68	.58	1217.00	-1.65†	0.34
B-D Confusion	76.32	.54	51.62	.24	965.50	-2.96**	0.69
Familial Incidence	76.69	.71	49.84	.38	921.00	-3.33**	0.76

†not significant. * $p < .05$. ** $p < .01$. *** $p < .001$.

Based on the results of the logistic regression, of the 143 participants in these sub-samples, 113 (79%) were classified as dyslexic (At-risk Classified Dyslexic (ACD) and 30 (21%) as not dyslexic (At-risk Classified Control (ACC). Given that these two groups of students did not proceed to have a full educational psychology assessment, we had no further indication of their dyslexia status, and hence we examined the validity of their classification according to the BDT, ACD (n = 113) and ACC (n = 30) by comparing their performances on the two literacy measures that were available for them, namely, the Two Minute Spelling and Non-sense Passage Reading subtests of the DAST. The groups differed significantly with the ACD group obtaining lower scores than those ACC group. The mean Spelling score for the ACD group was M = 27.38, SD = 4.50 and for the ACC group it was M = 30.97, SD = 4.12, $t(140) = -3.94$, $p < .001$. The mean Reading score for the ACD group was M = 79.49 (SD = 11.59) and for the ACC group it was M = 88.59, SD = 7.18, $t(70.69) = -4.03$, $p < .001$. Effect sizes calculated using Cohen's d, indicated large differences with d = 0.97 for the Reading and d = 0.83 for the Spelling. These results provide support for the accuracy of the classification of the groups based on their BDT scores.

In addition, a comparison of the dyslexic group (confirmed by full EP assessment) with the screened *at-risk* (dyslexia positive) and screened *not-at-risk* (dyslexia negative) groups indicated that, the screened *at-risk* group performance was more similar to the dyslexic group than the screened *not-at-risk* group. Although the performance of the dyslexic group differed significantly from both groups on the literacy measures, the magnitude of the difference was much greater for the screened *not-at-risk* group, than for the screened *at-risk* group. On the Spelling test dyslexic group: M = 25.90, SD = 4.64, screened *at-risk* group: M = 27.28, SD = 4.34, $t(307) = -2.60$, $p = .010$, Cohen's d = 0.31, and screened *not-at-risk* group attained M = 32.16, SD = 3.94, $t(215) = -6.44$, $p < .001$, Cohen's d size = 1.46. A similar pattern emerged for Reading, for the confirmed dyslexic group: M = 76.12, SD = 12.80, the screened *at-risk* group: M = 80.00, SD = 11.69, $t(304) = -2.66$, $p = .008$, Cohen's d = 0.32, and the screened *not-at-risk* group: M = 87.68 SD = 7.48, $t(45.11) = -6.57$, Cohen's d = 1.14. Thus, the BDT scores, which do not include estimates of literacy skills, reliably categorized students with less well developed literacy skills as dyslexia-positive and those with better developed skills as dyslexia-negative.

The results provided empirical evidence of the construct validity of the BDT as the performances of dyslexics and controls differed on the measure overall, as well as on all subtests. Additionally it indicated that scores on the BDT were predictive of group membership, and of literacy performance providing evidence of predictive validity. The results therefore provided strong empirical support for the construct and predictive validity of the BDT.

3.5.3 An Examination of the Reliability of the BDT

3.5.3.1 Internal consistency reliability of the BDT.

In order to evaluate the reliability of the measure as a whole, as well as each of its component subtests, we estimated the internal consistency (reliability) with Cronbach's coefficient Alpha using the entire sample of dyslexic, control, and at-risk participants who had undergone screening with the BDT (N = 413). The results of the validation of the BDT suggested that the performance differences between groups were primarily quantitative, and the relative patterns of difficulty across subtests were similar for dyslexics and controls thus, pooling across the three groups seemed defensible, and allowed us to maximise the heterogeneity of the sample and to reduce restricted range effects.

The reliability analysis yielded an overall coefficient $\alpha = .69$, which may be described as adequate indicating that the subtests are consistent and are likely measuring the same underlying construct (see Table 3.7). Although adequate its reliability is, lower than what is considered ideal (Anastasi & Urbina 1997; Field, 2009; Kline, 2000). The internal consistency of the measure may have been influenced by the heterogeneity of the subtests as indicated by its low mean inter-item correlation of $r = .18$. Item-total correlations ranged from .16 to .45 with Months Forwards and Tables having the lowest and highest correlations, respectively. The item-total correlations of all the subtests, except Months Forwards, were greater than .30 indicating that they were correlated to the total score on the test and contributed to the reliability of the measure (Field, 2009). The ceiling scores coupled with a low correlation of the Months Forwards subtest indicates that it is not contributing to the reliability of the measure and is insensitive for this age group. The squared multiple correlation figures indicated that the subtest with the greatest contribution to the internal consistency of the measure was Digits Reversed R^2 = .29, while the subtest with the least contribution, not surprisingly, was Months Forward R^2 = .10. The Cronbach's alpha-if-item-deleted figure for each subtest indicated that the reliability of the measure could not be improved by deleting any of the subtests. However, the deletion of the Months Forwards subtest left alpha for the measure unchanged, providing further evidence of the redundancy of this subtest. The three subtests that had the greatest statistical bearing on the overall reliability of the measure were Digits Reversed, Digits Forwards, and Tables, with Subtraction a close fourth. This pattern of results is largely consistent with that obtained for the comparison of the groups on the BDT where Digits Forwards, Digit Reversed, and Tables were also the most difficult for most participants. The results of the reliability analysis indicated that the BDT is a measure with adequate reliability, and with the exception of Months Forwards, all subtests were contributing to its reliability. Additionally the reliability of the BDT cannot be improved by deleting subtests.

Table 3.7. Item-total Statistics for BDT Subtests (N = 413)

BDT Subtests	Corrected item-total correlation	Squared multiple correlation	Cronbach's alpha if item deleted
Left-right	.32	.12	.67
Polysyllables	.31	.13	.67
Subtraction	.43	.24	.65
Tables	.45	.25	.65
Months Forwards	.16	.10	.69
Months Reversed	.33	.18	.67
Digits Forwards	.37	.26	.66
Digits Reversed	.43	.29	.65
B-D Confusion	.31	.13	.68
Familial Incidence	.37	.17	.66

3.6 Discussion

The main purpose of this study was to evaluate the BDT by examining its psychometric properties, especially its ability to discriminate between adult students with dyslexia and non-dyslexic controls using data from a large sample of university students. We investigated the data of 413 university students consisting of 3 groups, 193 diagnosed with dyslexia (dyslexic group), 40 with no history of learning difficulties (control group), and 180 screened at-risk of dyslexia or assessed with other learning difficulties. The cognitive profiles of the two main groups of interest, the confirmed dyslexic and control groups were found to be in the average to above average range on the standardised tests of ability and literacy. However, with the exception of the measure of verbal ability, where the groups were comparable, the controls attained significantly higher scores than the dyslexics, and this most notably on measures of reading (d = 1.86) and spelling (d = 1.59). The groups were assessed on different batteries however, and therefore, any direct comparisons have to be interpreted with caution. The group that underwent screening only was not assessed on general cognitive ability, however, their screening scores on DAST literacy measures were greater than those of the dyslexic group but lower than the control group, and there was no reason to suspect that their general abilities deviated systematically from the other two groups.

3.6.1 Validity

The performance of the dyslexic and control groups provided evidence for the construct and predictive validity of the BDT. The dyslexic group performed significantly less well than controls on the measure as a whole as well as on each subtest. Thus, the subtests of the BDT reliably differentiated between dyslexic and non-dyslexic groups, on eight basic skills as well as on two self-report anecdotal questions. Additionally the effect size calculations revealed that these differences were not only significant but also very large. Similar results were reported in Miles (1993) in a study investigating the speed of multiplication of boys aged between 7 and 14 years with and without dyslexia, who also completed the BDT. Miles found that the dyslexics performed less well (obtained higher scores) than both chronological and spelling age-matched controls. In our study of adults, on most of the subtests of the BDT, the dyslexics (57.5% to 82.9%) obtained positive dyslexia indices and their performance contrasted greatly with that of the controls of whom only a minority obtained positive scores (2.5% to 47.5%). In fact, on four of the subtests Polysyllabic Words, Subtraction, Months Forwards, and B-D Confusion none of the controls obtained a positive score. These results indicate that overall, the BDT sensitively differentiates between dyslexic and non-dyslexic adult students.

This global sensitivity, however, obscures the weakness of several subtests which clearly are less appropriate for inclusion in assessing adults. The least sensitive subtest was Months Forwards, on which both groups were at ceiling. In addition, relatively few dyslexic participants obtained positive scores on the subtests Months Reversed (30.1%), Polysyllabic Words (31.1%), and Subtraction (35.8%). The low sensitivity of these subtests suggests that they might add relatively little value to the battery as a whole (see reliability analysis); however, their inclusion in no way damaged the reliability of the battery.

Further evidence of the capacity of the BDT to discriminate between dyslexics and controls was provided by the results of the logistic regression analysis. The BDT correctly classified 94% of the participants, which is an excellent hit rate. It also had an excellent sensitivity rate, correctly classifying 96.4% of the dyslexic group, well above the minimum recommended 80% (Glascoe & Byrne, 1993). Only 7 of the 193 dyslexic participants were incorrectly classified (false negatives). This high sensitivity rate will ensure that the number of false negatives is kept to a minimum and that most adults who are at risk of dyslexia will be correctly identified. The BDT also correctly classified 82.5% of the controls making its specificity rate lower than the 90% minimum recommended. This may result in a larger than acceptable proportion of non-dyslexics being incorrectly identified as being at risk of dyslexia (false positives).

An examination of the BDT scores of the 14 participants misclassified (7 dyslexics and 7 controls) revealed that the scores of these individuals varied considerably from the mean scores of their respective groups. For the dyslexic participants classified as controls (false negatives), their scores on the BDT, which ranged from 2.5 to 3, were much lower than the mean score 6.17 of the dyslexic group and outside the average variance in the performance of this group $SD = 1.44$. The opposite pattern held for the misclassified (false positives) control participants whose scores ranged from a low of 3.5 to a high of 5. These scores were much higher than the mean score of 2.09 for the control group and outside the average variance in the performance of this group, *SD* = 1.23. These 14 participants had scores outside the norms for their respective groups, suggesting that setting a cut off point for identifying adults at risk of dyslexia may help to improve the specificity rate (proportion of individuals without dyslexia correctly classified) of the BDT while not adversely affecting its sensitivity rate (proportion of individuals with dyslexia correctly classified). Using a cut off point of 4 as the minimum required for a positive indication of risk at screening (as suggested by Miles (1997) for research purposes) would have increased the specificity rate of the BDT in the present study from 82.5% to 92.5% which is above the minimum recommended (Glascoe & Byrne, 1993). This cut off could also be used for general screening purposes.

The classification of the at-risk group provided further evidence of the capacity of the BDT to discriminate between dyslexics and non-dyslexic individuals. Based on their BDT scores 79% of this group were classified as dyslexics and 21% as controls. A comparison of the performance of the two at-risk groups on literacy tasks, indicated that the individuals classified as dyslexic performed significantly worse than those classified as controls, thus supporting their classification. Furthermore, the performance of the individuals in the at-risk classified as dyslexic group, was very similar to dyslexic group. Sutherland and Smith (1991) in their study which compared the BDT to two other two other dyslexia screening tests, the Boder Test of Reading-Spelling Patterns and the Aston Index were critical of the scoring guidance provided by the manual, which they believed was inadequate. The results of this study suggest that the researchers' criticism of the BDT may not be justified. In this study, participants in the archival sample were screened for dyslexia by several different assessors. The accuracy of the screening outcome of these participants when compared to full assessment (84% of these screened positive were diagnosed with dyslexia) suggest that the scoring criteria as detailed in the manual are adequate.

Despite the less than desirable specificity rate, the classification rate of the BDT compares favourably with other adult dyslexia screening tests. The DAST manual indicates a sensitivity rate of 93% and a specificity rate of 100%; however, only 15 dyslexics were included in the validation study for the test (Nicolson & Fawcett, 1998). Additionally independent research has reported lower rates for sensitivity (85%) and specificity (74%) for the DAST (Harrison & Nichols, 2005). The sensitivity rate of the BDT also compares favourably with other adult dyslexia screening tests, the Lucid Adult Dyslexia Screening Plus (91%), (Singleton, Horne, & Simmons, 2009), and the York Adult Assessment-Revised (80%), (Warmington et al., 2013). However, the specificity rate of the BDT is lower than the rates reported for these test 90% and 97% respectively. The higher sensitivity rate of the BDT means it is able to more accurately identify adults with dyslexia and minimise the numbers who are misclassified. This arguably makes the BDT a more effective screening test. What is more, as shown above, the specificity rate of the test could be improved by clearly stipulating a cut-off score of 4 for classification decisions. From these results, we can conclude that the BDT is a valid dyslexia screening tool for adults with the capacity to discriminate between dyslexics and controls.

3.6.2 Reliability

The results of the reliability analysis indicated that the internal consistency reliability of the BDT (α = .69) is adequate, but not optimal (Anastasi & Urbina 1997; Field, 2009; Kline, 2000). This suggests that the items on the BDT are consistent and are likely measuring the same underlying construct. It is likely that the internal consistency was affected by the heterogeneity of the subtests included in the test. The creator of the test Professor Miles believed dyslexia to be a syndrome with a distinctive pattern of symptoms/difficulties and the test was designed to reflect this.

An examination of the alpha-if-deleted figures indicated that the internal consistency of the measure could not be improved by deleting any of the subtests. However, the deletion of the Months Forwards subtest leaves alpha unchanged, indicating that this subtest may be deleted as it is not contributing to the internal consistency of the measure. This coincides with the results of validity of the BDT which indicated that the Months Forwards subtest produced ceiling effects even among the dyslexic group. Miles (1993) reached a similar conclusion after assessing 48 dyslexic adults on the BDT with only 10.4% obtaining positive scores on this subtest. The difference between the percentage of dyslexic participants that obtained positive scores in Miles' study and in this study (3%) may be attributed to the composition of the samples in the studies. For this study the dyslexic sample consisted of only university students (highly compensated dyslexics) while the Miles sample was more diverse with only 17 (35%) university students. The reliability of the BDT is also comparable to that of the

DAST which reports test-retest reliability for its subtest items ranging from a low of $r = .64$ to a high of $r = .93$ (Nicolson & Fawcett, 1998) and also with the YAA-R, which reports internal consistency reliability ranging from $\alpha = .53$ to $\alpha = .81$ on its subtests.

3.7 Conclusion

There is a need for more empirical evidence on the effectiveness of the dyslexia screening tests that are currently in use. This study provides empirical evidence of the reliability, construct and predictive validity of the BDT using a large sample of dyslexic adult students. The results indicate that the BDT is a reliable and valid measure capable of discriminating between dyslexic and non-dyslexic adults. The study also highlighted some weaknesses or areas for improvement as well as suggestions for how the measure may be enhanced. The BDT has been used effectively to screen for dyslexia in children and adults for almost three decades, and the results of this study suggest that it deserves its place as a quick, engaging and sensitive dyslexia screener.

Chapter 4: Development of the Bangor Adult Literacy Index: Pilot Study (Study 2a)

4.1 Introduction

The review of current adult dyslexia screening tests undertaken in Chapter 2 highlighted weaknesses of the tests, as well as the ongoing need for effective screening tests, which are adequately researched and psychometrically sound, to identify adults with dyslexia. The main objective of the present study was to evaluate a number of tasks selected for inclusion in a new screening test for adults with dyslexia, with a view to developing a test, based on current research evidence, that is psychometrically sound, cost-effective, quick and easy to administer (by trained non-professionals), and suitable for use with adults in the general population including compensated dyslexics. In short, the new tool is assumed to possess all the characteristics of a good screening test.

4.1.1 Theoretical Framework

The first step in the development of this new test was to establish a suitable theoretical framework on which the test could be based. As stated in the literature review, the phonological deficit theory of dyslexia is the most widely accepted causal theory of dyslexia and is supported by strong and converging research evidence in different areas (behavioural, intervention, neurobiological, neuroimaging, and genetics) of research on dyslexia. Although current theorizing about dyslexia suggests multiple deficits may be very typical among people with dyslexia, instead of single deficit aetiology for the disorder, a phonological deficit as a core causal deficit is still compatible with this view (Norton & Wolf, 2012; Pennington, 2006, Snowling, 2012). Additionally, the behavioural manifestations of the phonological deficit in adults with dyslexia are well documented and researched (Bruck, 1992; Felton, Naylor, & Wood, 1990; Snowling, Nation, Moxham, Gallagher, & Frith, 1997).

Further, it is recognised that in the absence of physiological criteria that can be used to assess dyslexia, behavioural deficits are our best alternative (Reid, Szczerbinski, Iskiera-Kasperek, & Hansen, 2007; Tonnessen, 1997; Turner, 2008). Such deficits offer an objective way of identifying individuals with the disorder (Turner, 2008). Existing dyslexia screening tests such as the Bangor Dyslexia Test (BDT) and the Dyslexia Adult Screening Test (DAST) indeed include tasks assessing the behavioural manifestations of the disorder, including items assessing phonological processing.

Researchers have created a number of tasks that have proven to be very effective in detecting the phonological difficulties experienced by adults with dyslexia and which effectively discriminate between individuals with and without the disorder. Thus, there are strong theoretical and practical justifications for using the phonological deficit theory as the basis of the new screening test for adults with dyslexia. We therefore developed tasks appropriate for detecting behavioural deficits that were also in line with the theoretical framework of the test, and were effective in detecting dyslexia in the general adult population, including adults who have compensated for their reading difficulties (i.e. high functioning adults such as university students). Based on prior research and a review of the literature, six tasks were included in the new screening test; the tasks and the rationales for their inclusion are detailed below.

4.1.2 Rationale for Task Selection

4.1.2.1 Literacy tasks.

Dyslexia is a learning disability, manifested in difficulty acquiring literacy skills in childhood, and dyslexic adults typically perform less well (more slowly and less accurately) on literacy tasks when compared to their non-dyslexic counterparts. There is strong and consistent research evidence of deficits in the literacy skills, particularly reading and spelling, of adults with dyslexia, and these tasks have been shown to discriminate between adults with and without dyslexia (Gottardo, Siegel, & Stanovich, 1997; Hatcher, Snowling, & Griffiths, 2002; Reid et al., 2007; Swanson & Hsieh, 2009). However, some dyslexics can improve their reading accuracy and are able to perform in the average range on tests of attainment, but continue to lack fluency (Kemp, Parrila, & Kirby 2009; Lefly & Pennington, 1991; Parrila, Georgiou, & Corkett, 2007). Further, measures of reading fluency have been found to be sensitive and effective for discriminating between adults with and without dyslexia including compensated dyslexics (Re, Tressoldi, Cornoldi, & Lucangeli, 2011). It therefore seemed prudent to include a measure of reading fluency in our new screening test. A similar measure is included in the Dyslexia Adult Screening Test (DAST) (Nicolson & Fawcett, 1998), in the Test of Word Reading Efficiency (Wagner, Torgesen, Rashotte, 1999), and it is used in research studies investigating dyslexia in both children (Pennington, Cardoso-Martins, Green, & Lefly, 2001) and adults (Reid et al., 2007; Svensson & Jacobson, 2006). However, our measure is unique in its design as a measure of fluency only, consisting of words with high familiarity and frequency and hence low difficulty.

As with reading, there is strong and consistent research evidence that adults with dyslexia exhibit a deficit in spelling (Hatcher, et al., 2002; Reid et al., 2007; Swanson & Hsieh, 2009). Research with children suggests that reading and spelling skills develop concurrently and are correlated (Bruck & Waters, 1990). However, spelling is considered a more demanding task than reading and a spelling test is likely to be a more sensitive test which can effectively discriminate between adults with and without dyslexia (Caravolas, Hulme, & Snowling, 2001; Frith, 1999). Additionally, sometimes a spelling deficit is the only deficit exhibited by adults with dyslexia (Romani, Olson, & Di Betta, 2005). Consequently, measures of spelling are included in the screening and assessment of dyslexia. Therefore, a task assessing spelling skills was considered appropriate for inclusion in the new screening test. Furthermore, researchers have found that speeded measures may be more effective for discriminating between adults with and without learning disabilities (Lesaux, Pearson, & Siegel, 2006; Ofiesh, Mather, & Russell, 2005; Re et al., 2011). For example, Re et al. (2011) compared the performance of university students with and without dyslexia on a number of literacy and cognitive measures and found that speeded measures were better (larger effect sizes) for discriminating between the groups, thus corroborating similar earlier findings across a range of tasks (Bruck, 1993; Miller-Shaul, 2005; Ramus et. al., 2003b). Not surprisingly, additional time for the completion of tasks is one of the most requested and provided accommodations for postsecondary students with learning disabilities (Ofiesh et al., 2005). Therefore, timed tasks are likely to be more sensitive and effective for identifying adults with dyslexia, and the spelling task, as well as all the other tasks included in the screening test, was timed. A similar measure of spelling is included in the DAST (Nicolson & Fawcett, 1998). However, our task differs from that of the DAST because of the careful selection of words with inconsistent grapheme included. The timed element of the tasks also facilitated one of our main objectives, that of developing a screening test that requires relatively little time to administer each task.

4.1.2.2 Phonological processing tasks.

With the phonological deficit theory as the framework for the new screening test, and the strong and converging research evidence of the persistence of a phonological deficit in adults with dyslexia, tasks assessing this deficit were included (Bruck, 1992; Hatcher et al., 2002; Reid et, al., 2007; Wilson & Lesaux, 2001). Three main types of tasks are generally used to measure phonological processing skills and these are: phonological awareness, verbal short-term memory, and rapid automatized naming (RAN) (Hulme & Snowling, 2009; Szenkovits & Ramus, 2005). Tasks designed to assess phonological awareness measure the ability to perceive, manipulate, and produce the speech sounds of a language (Hulme & Snowling, 2009), and include tasks such as phoneme isolation, spoonerisms, and phoneme deletion. Research evidence of a deficit in children (Sucena, Castro, & Seymour, 2009; Vellutino, Fletcher, Snowling, & Scanlon 2004) and adults (Bruck, 1992; Kemp et al., 2009; Snowling et al., 1997) on tasks assessing phonological awareness is long standing and consistent. In addition, the literature indicates that for adults, tasks with nonwords, such as nonword reading are more effective for detecting phonological deficits in adults than real words and this was true for even compensated dyslexic adults (Elbro, Nielsen, & Petersen, 1994; Gross-Glenn, Jallad, Novoa, Helgren-Lempesis, & Lubs, 1990). For example, Elbro et al., (1994) compared the performances of 102 Danish adults with reading difficulties to 56 controls on several measures including word and nonword reading. They found that the dyslexics exhibited deficits on measures of phonological processing, and word and nonword reading. However, the largest deficit for the adults with reading difficulties was on the nonword reading task. Tasks assessing phonological awareness are included in existing dyslexia screening test such as the DAST, Lucid Adult Dyslexia Screening Plus, and the York Adult Assessment Battery-Revised (YAA-R). The measure of phonological awareness selected for inclusion in our test differed from those currently in use as it was designed to reliably assess both accuracy and speed.

Tasks designed to assess verbal short term memory measure the ability to accurately repeat a series of digits or words in the order of presentation (Hulme & Snowling, 2009) and often include nonword repetition and digit span tasks. For nonword repetition tasks, there is consistent research evidence that the performance of children (Barbosa, Miranda, Santos, & Bueno, 2009; Melby-Lervåg & Lervåg, 2012; Messbauer & de Jong, 2003; Taylor, Lean, & Schwartz, 1989) and adults (Elbro et al., 1994; Parrila et al., 2007; Szenkovits & Ramus, 2005) with dyslexia is deficient (less accurate) when compared to their counterparts without dyslexia, and the tasks can effectively discriminate between them (Barbosa et al, 2009; Szenkovits & Ramus, 2005; Taylor et al., 1989). For example, in a meta-analysis that included 34 studies, Melby-Lervåg and Lervåg (2012) compared the performance on nonword repetition measures of 800 children with dyslexia and 1021 chronological age matched controls. They found that the dyslexic children performed less well than the controls with overall large effect size $d = 0.92$. Similarly, Szenkovits and Ramus (2005) found that the performance of adults with dyslexia on word and nonword repetition tasks was impaired compared to chronological age and IQ matched controls. Currently there are assessment batteries (the Children's Test of Nonword Repetition and the Comprehensive Test of Phonological Processing) that include nonword repetition tasks, and the BDT includes a similar word repetition task. A nonword repetition task was created for inclusion in the test.

RAN tasks involve the speeded naming of a series of familiar items (digits, letters, colours, or objects) presented in random order in a grid array (Denckla & Cutting, 1999; Norton & Wolf, 2012). The task was originally created by Denckla and Rudel (1976) and is also known as rapid serial naming, rapid naming, and naming speed. There is considerable debate about exactly what RAN measures with several contending views which are all supported by research evidence (for review see Kail & Hall, 1994; Kirby et al., 2010; Norton & Wolf, 2012; Savage, Pillay, & Melidona, 2007). In brief, first, RAN is regarded as a phonological processing task measuring the rate of access to and retrieval of phonological information (Savage et al., 2007). Second, RAN is seen by some as a measure of one part of a general processing speed that is the speed associated with cognitive processing (Kail & Hall, 1994). Third, RAN is proposed to represent a microcosm of the reading process as it involves the same processes such as eye saccades, working memory, automaticity, orthographic and phonological representations (Kirby et al., 2010; Norton & Wolf, 2012). Despite the controversy around precisely what RAN measures, there is considerable research evidence that it is related to literacy skills in children and adults (see Chapter 7 for more details), (Mellard, Fall, & Woods, 2010; Moll et al., 2014; Ransby & Swanson, 2003; Savage et al., 2005). Additionally, individuals with dyslexia exhibit deficits on RAN tasks when compared to their counterparts without dyslexia (Catts, Gillispie, Leonard, Kail, & Miller, 2002; Vukovic, Wilson, & Nash, 2004; Wimmer, Mayringer, & Landerl, 1998). RAN tasks are included in two of the current dyslexia screening tests for adults, the DAST and YAA-R. Two standardized measures have also been developed the Rapid Automatized Naming-Rapid Alternating Stimulus (Wolf & Denckla, 2005) and the rapid naming subtest of the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999a). Two versions of RAN digits and objects similar to the original created by Denckla and Rudel (1976) were used in the study and items included were comparable for speakers of both Welsh and English.

All three measures of phonological processing were included in the new screening test as the use of multiple measures to assess the construct is likely to improve the psychometric properties and sensitivity of the screening test (Chester, 2005). Additional advantages of using these tasks are that they are quick and easy to administer which therefore facilitates the development of a quick screening test.

4.1.2.3 Executive function task.

Taking into account the multidimensional view of dyslexia and the possibility that additional skills impairments might be present among adults with dyslexia, we wanted to include a task that tapped skills other than literacy and phonological processing. This would widen the range of potential difficulties assessed and hopefully the battery's capacity to effectively discriminate between adults with and without dyslexia. Researchers have reported the existence of a co-occurring executive function deficit in children (Brosnan et al. 2002; Helland & Asbjørnsen, 2000; Jeffries & Everatt, 2004; Protopapas, Archonti, &

Skaloumbakas, 2007; Reiter, Tucha, & Lange, 2005) as well as in adults with dyslexia (Brosnan, et al., 2002; Weyandt, Rice, Linterman, Mitzlaff, & Emert, 1998). Executive function is regarded as a multidimensional process used to guide behaviour toward a goal and includes several cognitive abilities (Banich, 2009; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003) such as: (1) shifting - the ability to switch between tasks; (2) problem solving; (3) sequencing – the ability to perform a task in the required order; (4) organization; and (5) inhibition – the ability to deliberately suppress dominant responses (Banich, 2009; Brosnan et al., 2002; Lehto et al., 2003). Although research on executive function in dyslexia is limited and the particular aspects of executive functioning examined have varied (especially for the adults), most of the studies with children and one with adults have reported deficits in the executive function of inhibition (Brosnan et al., 2002; Helland & Asbjørnsen, 2000; Jeffries, & Everatt, 2004; Reiter et al., 2005). Additionally, there is anecdotal evidence from the literature on adults with dyslexia in the workplace that appears to support some type of executive function difficulties in adults with dyslexia. Some of the difficulties reported in this literature include sequencing, organization, and concentration and attention (Bartlett & Moody, 2000; Moody, 2009; Reid & Kirk, 2001). In light of the above, we included a task assessing inhibition in order to investigate this deficit and to widen the range of deficits assessed in the new screening test.

Having identified tasks that are sensitive for detecting the behavioural manifestation of dyslexia in adults, and that are capable of discriminating between adults with and without the disorder, we selected similar tasks for inclusion in the new screening test. We adapted two existing tasks (phoneme deletion and RAN) and created four new tasks (reading, spelling, phonological short term memory and inhibition) for the new screening test which we named the Bangor Adult Literacy Index (BALI). We then conducted a pilot study with the tasks, to investigate their capacity to effectively discriminate between adults with and without dyslexia, individually and collectively, and to evaluate their validity and reliability.

4.2 Method

4.2.1 Participants

Participants were 36 native English speaking undergraduate students, 18 dyslexics and 18 controls, recruited from Bangor University. All the participants in the dyslexic group had a formal diagnosis of dyslexia and were registered with the Dyslexia Centre at the university. The dyslexic group comprised 11(61%) females and 7 (39%) males, with a mean age of M = 22.72 (SD = 6.79). The participants in the control group reported no learning or literacy difficulty and additionally satisfied the minimum criteria of standard scores of 85 and above on the general cognitive and background literacy measures. The group comprised 13 (72%) females and 5 (28%) males, with a mean age of M = 20.67 (SD = 5.24). The age difference between the groups was not statistically significant $t(34) = -1.02$, $p = .316$. The control participants were compensated with £4 of printer credits for their participation. Dyslexic participants were compensated with £5 as they were not all eligible for the printer credit scheme, coming from a variety of departments.

4.2.2 Measures

4.2.2.1 Background measures.

Participants' nonverbal and verbal IQ was assessed with the Matrices and Vocabulary subtests, respectively, of the Wide Range Intelligence Test (WRIT) (Glutting, Adams, & Shelow, 2000). The WRIT was developed to assess cognitive abilities in individuals aged 4 – 85 years. It consists of four subtests, two measuring nonverbal IQ (Matrices and Diamonds) and two measuring verbal IQ (Vocabulary and Verbal Analogies). The psychometric properties for the battery as reported in the manual are good. Its overall correlation with the Wechsler Adult Intelligence Scale III (WAIS III) (Wechsler, 1997) is $r = .97$: and for reliability: Matrices subtest - internal consistency $\alpha = .90$ and test-retest reliability $r = .77$, and Vocabulary subtest - internal consistency reliability $\alpha = .91$ and test-retest reliability $r = .91$. Literacy attainment was assessed using the Word Reading (single word reading aloud), Spelling (single word spelling to dictation), and Sentence Comprehension (a modified close format) subtests of the Wide Range Achievement Test (WRAT IV) (Wilkinson & Robertson, 2004). The WRAT IV was developed to measure achievements in literacy and numeracy in individuals aged 5 – 94 years. It consists of 4 subtests: Word Reading, Spelling, Sentence Comprehension, and Math Computation. The psychometric properties for the battery as reported in the manual are good. Its subtests are correlated with similar subtests of the Wechsler Individual Achievement Test II (WIAT II) (Wechsler, 2005) as follows: Word Reading $r = .71$, Spelling $r = .64$ and Sentence Comprehension $r = .61$. Internal consistency reliability reported: Word Reading $\alpha = .92$, Spelling $\alpha = .90$ and Sentence Comprehension $\alpha = .93$ and test-retest reliability $r = .88$, $r = .85$ and $r = .83$ respectively.

4.2.2.2 Experimental tasks.

Participants were also assessed on the six experimental tasks designed for inclusion in the new Bangor Adult Literacy Index (BALI). These included: two tasks assessing literacy skills; one assessing executive function, inhibition ability; and three phonological processing skills (phonological awareness, verbal phonological short term memory, and phonological processing speed).

4.2.2.2.1 Literacy tasks

4.2.2.2.1.1 1 Minute Reading task.

This task was created to measure reading fluency by assessing the number of words read accurately in one minute. The words included in the tasks were selected using the MRC Psycholinguistic Database (http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm) to identify words of different syllable lengths with a minimum print familiarity rating of 500 (the scale ranges from 100 to 700 points) and minimum Kucera-Francis written frequency rating of 100 occurrences per million. This was to ensure that participants had minimal difficulty with word recognition in the task, as it is primarily a measure of reading fluency and not accuracy. This focus of the test on fluency (rather than efficiency, which includes accuracy and speed) is novel and distinct from existing speeded reading tests in English language screeners and test batteries. One hundred and forty four words were included in the task, with mean print familiarity $M = 574.06$, and mean Kucera-Francis written frequency $M = 691.32$. An example of the stimuli is shown in Appendix A. The number of syllables in the selected words ranged from one to six, and the numbers of words of each length were as follows: 28 one-syllable words, 55 two-syllable words, 45 three-syllable words, 12 four-syllable words, 3 five-syllable words, and 1 six-syllable word. The words were arranged into three columns on each side of an A-4 sheet and were printed in lower case letters, 16-point Arial bold font, ordered by the number of syllables with one syllable words first. Participants were required to read the words aloud as quickly as possible for one minute. The task was timed and audio recorded, and the total number of words read correctly as well as the total number read incorrectly were recorded. The main score was the number of words read correctly in one minute.

4.2.2.2.1.2 2 Minute Spelling task.

This task was created to measure spelling efficiency (i.e. accuracy and speed) and assessed the number of words spelt accurately in two minutes. The words were selected using the MRC Psycholinguistic Database (http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm) to sample across words with common phonological, orthographic and morphological patterns in English. The task consisted of 44 words, of which10 were phoneme-to-grapheme inconsistent (the spelling of the word is not consistent with its pronunciation), and 34 were phoneme-to-grapheme consistent (the spelling of the word is consistent with its pronunciation). The systematic manipulation of the inconsistent graphemes and their basis in the words is novel and distinct from the similar measure of this sort (i.e., the DAST two-minute spelling test). An example of the stimuli is shown in Appendix B. The words ranged from one to five syllables and were not arranged in any particular order. Participants were required to spell as many words as possible for two minutes. The words were dictated to the participants, and the researcher pronounced the next word as soon as the participant finished writing the previously dictated word. The task was timed and the number of words spelt correctly, as well as the number spelt incorrectly were recorded. The main score was the number of words spelled correctly in 2 minutes.

4.2.2.2.2 Executive Function

4.2.2.2.2.1 Inhibition task.

This is an adaptation of the *Same World Opposite World* subtest which assesses inhibition in the Test of Everyday Attention for Children (Manly et al., 2001). In our version of the test, the stimuli were arrow heads pointing either up or down. Arrows were selected because the task requires highly automatized stimuli, for which the correct response should be difficult to inhibit. Arrows are highly iconic, overlearned symbols, and the “up” / “down” orientations are easy to pronounce. The task required the participants to name, as quickly as possible, the same or the opposite direction of arrows in a long quasi-random sequence of arrows. It consisted of three trials, one Same Direction and two Opposite Direction. Examples of the stimuli are shown in Appendices C and D. In the Same Direction trial, the participants were required to say the direction in which each arrow was pointing. For the Opposite Direction trials, the participants were required to say the opposite of the direction in which each arrow was pointing (“up” for a down arrow and “down” for an up arrow). The order of presentation of the items in the second of the Opposite Direction trial was the reverse of that of the first. The Same Direction trial was administered first followed by the two Opposite Direction trials. The researcher demonstrated how the task should be completed by saying the direction of the first three arrows during the administration of instructions for the task for the Same Direction and the first Opposite Direction trials. In the Opposite Direction trials, for the first 12 items, if the participant made 4 consecutive errors (saying the direction of the arrow instead of the opposite direction) the researcher stopped the task and reminded the participant of the instructions, after which the task was restarted. This was done only once during the administration of the task. The entire task was audio recorded and the time taken to complete each trial as well as the number of errors were recorded. To determine the inhibition score (the extent to which the participant could inhibit the tendency to name the real orientation of the arrow), the time taken to complete the two Opposite Direction trials was averaged, and the time taken to complete the Same Direction trial subtracted from this figure. A score reflecting good inhibition ability would be close to zero, indicating virtually no cost to producing the opposite response.

4.2.2.2.3 Phonological Processing tasks

4.2.2.2.3.1 Phoneme Deletion task.

This task was adapted from Judge, Caravolas, and Knox (2006) and is a measure of phoneme awareness, assessing the ability to produce and manipulate speech sounds. The task consisted of 24 monosyllabic nonwords, and was divided into two blocks; the first block contained 12 consonant-consonant-vowel-consonant (CCVC) (e.g., *stek*) nonwords, and the second block 12 consonant-vowel-consonant-consonant (CVCC) (e.g., *semp*) nonwords. An example of the stimuli is shown in Appendix E. The task required participants to delete a phoneme in a nonword to produce a new nonword. For the CCVC items, participants were required to delete the second phoneme of the word. For example, for the nonword *stek* they were required to delete the /t/ to produce the new nonword *sek.* Similarly, for the CVCC items, participants were required to delete the penultimate phoneme and produce the new nonword. For example, for the nonword *sont* they were required to delete the /n/ to produce the new nonword *sot*. Block one always preceded block two. Each block was preceded by three practice trials and the main task was administered only if the participant was able to produce the correct response on at least one of the practice items. If the participant failed all the practice items in block one, the task was not administered. The researcher was trained in the accurate and fluent pronunciation of the nonwords. During administration the researcher said the nonword aloud, the participant repeated it, silently deleted the targeted phoneme, and then said aloud the new nonword as accurately and quickly as possible. If the participant incorrectly pronounced the stimulus item but deleted the correct phoneme and the nonword produced was accurate, this was also scored as correct. For example, for the CCVC nonword *treap*, if instead the participant repeated it as *fleap* and then proceeded to delete the /l/ and produced the new nonword *feap*, this was accepted as correct. Similarly, for the CVCC nonword *thoist*, if the participant said *foist* and then proceeded to correctly delete the /s/ and produce the new nonword *foit*, this would be scored as correct. The task was recorded and the recordings were used to score the accuracy of the participants' response. A score of one was assigned to each correct response and a score of zero to an incorrect response, for a maximum accuracy score of 24 (12 for each block). The time taken to complete each block was also recorded and was considered an estimate of phoneme awareness speed.

This measure of phoneme awareness was specifically selected because it was designed to measure both accuracy and speed reliably. For block 1, the level of difficulty of the items included was relatively low and it is expected that most adults would perform at ceiling and therefore speed would be the main discriminator. For block 2, the items were relatively more difficulty (compared to block 1) and both accuracy and speed should discriminate. It was anticipated that speed, much more than accuracy, would be a good indicator of dyslexia in adults. Therefore this task would be advantageous especially where participants perform at ceiling in accuracy.

4.2.2.2.3.2 Nonword Repetition task.

This task was created to assess verbal phonological short term memory and articulation fluency. The words were created by manipulating syllables and changing the sounds of real words. For example, the real word "*personnel*" was manipulated to [ʻmɛlpɚʼsaɪ] "*melpersai*" by replacing the first syllable with the last and changing the letters *n – m*, and *o – ai*. The pronunciation of the nonwords in terms of their lexical stress pattern, and the number of syllables was the same as the real words from which they were created. This facilitated more accurate and consistent pronunciation of each word during administration. Examples of the stimuli are shown in Appendix F. The task consisted of 10 nonwords of three to six syllables as follows: one three-syllable word, three four-syllable words, four five-syllable words, and two six-syllable words. The researcher (a native English speaker) was trained in the accurate and fluent pronunciation of the nonwords. Participants were required to repeat each nonword spoken by the researcher. The task was recorded and the recordings used to score the accuracy of the participants' response. A score of one was assigned to each correct response and a score of zero to an incorrect response, for a maximum score of 10.

4.2.2.2.3.3 Rapid Automatized Naming (RAN) task.

This task was originally created by Denckla and Rudel (1976) and is included as a test of phonological processing speed. It typically consists of a set of highly familiar items such as digits, letters, colours, or objects presented in a grid array and requires the rapid recognition and naming of visually presented stimuli. The two versions of RAN in the present study were adapted from Caravolas et al. (2012) and assessed Digits and Objects naming; the stimuli were 40 items (8 x 5 grids) comprised of five item types (e.g., 5 digits, 5 object pictures). Examples of the stimuli are shown in Appendices G and H. Each pool of items was presented eight times in pseudorandom order. Participants were required to say each item in the array as quickly as possible. There were two trials for each task and the second trial was administered directly after completion of the first. The order of presentation of the items on the second trial differed from that of the first. The entire task was audio recorded, and the recordings used to score the accuracy of the participants' response. The time taken to complete each trial, as well as the number of errors made on each trial, was recorded.

4.2.3 Procedure

All participants were tested individually, in a session lasting approximately 60 - 70 minutes, on the background measures as well as the experimental tasks included in the BALI. Tasks were administered in a fixed order with first the Reading, Spelling, and Sentence Completion subtests of the WRAT IV, second the Matrices and Vocabulary subtests of the WRIT, and finally the experimental tasks in the following order: RAN Objects, Phoneme Deletion, 1 Minute Reading, RAN Digits, 2 Minute Spelling, Inhibition, and Nonword Repetition. The standardised measures were all administered in accordance with administration instructions and the experimental tasks in line with established instructions. Ethical approval for the study was granted by the School of Psychology, Bangor University.

4.3 Data Reduction and Preliminary Analyses

We created composite scores for the two RAN tasks (Digits and Objects) in order to increase construct reliability. This was calculated by averaging the time taken to complete each trial of the task. Both trials assessed the same construct and were highly correlated, RAN Digits $r(36) = .96$, $p < .001$and RAN Objects $r(36) = .88$, $p < .001$. Additionally there was no significant difference in the (negligible) number of errors made by the groups on the RAN tasks, for RAN Digits (controls $M = .00$, $SD = .00$, dyslexics $M = .17$, $SD = .51$), $t(17)$ -1.37, $p = .187$ and for RAN Objects (controls $M = .11$, $SD = .32$, dyslexics $M = .56$, $SD = 1.04$), $t(20.25)$ -1.73, $p = .099$. Scores for the Inhibition task were calculated by averaging the times taken to complete the two Opposite Direction trials, and subtracting the time taken to complete the Same Direction trial. Prior to calculating the scores the correlation of the two Opposite Direction trials was examined and they were highly correlated $r(36) = .90$, $p <$ .001. Also, there were no significant differences in the number of errors made by the groups on any of the trials of the task, Same Direction (controls $M = .17$, $SD = .38$, dyslexics M = .00, $SD = .00$), $t(17)$ 1.84, $p = .083$, Opposite Direction trial one (controls $M = 1.67$, SD = 3.31, dyslexics $M = .89$, $SD = 1.49$), $t(34)$ .91, $p = .369$, and Opposite Direction trial two (controls $M = .22$, $SD = .73$, dyslexics $M = .33$, $SD = .69$), $t(34)$ -.47, $p = .641$. For the Nonword Repetition task the groups did not differ on the time taken to complete the task $t(34)$ -.56, $p = .579$, however, they differed in terms of accuracy scores, which were used for the data analyses. The 1 Minute Reading task was designed as a measure of reading fluency and not accuracy, and thus very few errors were expected, even from the dyslexic group. To determine if this objective was achieved, we examined the error performance of the groups on the task. As expected, both groups made very few reading errors (dyslexics $M = .33$ SD = .69, controls $M = .11$ $SD = .32$), and did not differ significantly in this respect ($t(24.2)$ -1.24 p = .226). This suggests that the task was measuring reading fluency (as intended) rather than word recognition accuracy. Based on these preliminary analyses, the following scores from the experimental tasks were included in the BALI measure: 1 Minute Reading fluency (number of words correctly read per minute) scores – to assess reading fluency; 2 Minute Spelling efficiency (number of words spelled correctly per 2 minutes) scores – to assess spelling efficiency; Inhibition scores – to assess the executive function of inhibition; Phoneme Deletion accuracy scores – to assess phoneme awareness accuracy; Phoneme Deletion speed scores – to assess phoneme awareness fluency; Nonword Repetition accuracy scores – to assess verbal phonological short term memory; and RAN Digits and Objects time scores – to assess phonological processing speed.

After the preliminary analyses and prior to the main data analyses, the data were checked for outliers and the assumptions of the multivariate analysis (MANOVA). To identify outliers the mean for each group was calculated separately. Outliers above or below 2.5 standard deviations from the mean were adjusted to the score corresponding to 2.5 standard deviations from the mean. There were no outliers for either group on the background measures. For the experimental tasks, only one score was adjusted for each group on the Phoneme Deletion Accuracy task. The Shapiro-Wilk test for normality indicated that scores on the RAN Digits, Phoneme Deletion Accuracy, and Nonword Repetition tasks were not normally distributed and these were transformed. However, the results of the MANOVA with the untransformed variables are reported as analyses with the transformed variables produced the same pattern of results. To analyse the data, we assessed the capacity of the experimental tasks collectively (the BALI battery) and individually to discriminate between dyslexics and controls. This was done with MANOVA, and followed up with univariate analyses (ANOVA) with Bonferroni adjusted alpha level of .006. Cohen's *d* effect size is also reported as an indication of the magnitude of group differences. The validity of the experimental tasks was evaluated by examining their correlation with the standardised measures of literacy (WRAT IV subtests) and the internal consistency reliability of two of the experimental tasks, Phoneme Deletion and Nonword Repetition was assessed.

4.4 Results

4.4.1 Comparisons of Performance of Dyslexic and Control Groups on the Background Measures

The descriptive statistics of the groups on the background measures along with Cohen's *d* effect size are reported in Table 4.1. The MANOVA analysis with groups as the independent variables and the background measures as the dependent variables was used to assess whether the groups differed on the set of measures. The results indicated a significant difference between the dyslexics and controls on the set of measures F (5, 30) = 6.46 $p < .001$; Pillai's Trace = 0.52. Follow up ANOVA (with Bonferroni adjusted alpha level of .01) revealed that the group performances were significantly different on the measures of WRAT IV Word Reading, F (1, 34) = 19.12, $p < .001$ and WRAT IV Spelling, F (1, 34) = 15.32, $p < .001$, with the dyslexic group performing less well than the control group, but not on the subtests of the WRIT and the WRAT IV Sentence Comprehension. Cohen's *d* effect sizes indicated that the performance differences were very large, with WRAT IV Reading $d = 1.46$ and WRAT IV Spelling $d = 1.32$. Despite these differences, the scores obtained by the dyslexic group on both tasks were well within the average range, as can be seen in Table 4.1. This is not unusual with a university population. These results suggest that the groups were similar in terms of their verbal and nonverbal IQ and only differed on their literacy skills, specifically reading and spelling.

Table 4.1 Mean Standard Scores (Standard Deviation in Parenthesis) of Dyslexic and Control Groups on Background Measures and Effect Size (N = 36)

Measures	Dyslexics (*n* = 18)	Controls (*n* = 18)	Cohen's *d*
WRIT Matrices	105.50 (12.47)	104.50 (6.85)	0.10
WRIT Vocabulary	102.11 (13.53)	105.39 (9.42)	0.29
WRAT IV Word Reading	97.56 (7.01)	107.44 (6.55)	1.46
WRAT IV Spelling	97.44 (13.14)	112.78 (10.18)	1.32
WRAT IV Sentence Comprehension	108.89 (14.66)	106.28 (8.98)	0.22

4.4.2 An Examination of the Reliability of the Tasks in the BALI

4.4.2.1 Internal consistency reliability of the Phoneme Deletion task.

We examined the internal consistency reliability (Cronbach's alpha) of two of the tasks included in the BALI, Phoneme Deletion and Nonword Repetition using the total sample. For the Phoneme Deletion task, we examined the reliability for each block as well as the task as a whole. The internal consistency of the task as a whole was good, $\alpha = .88$, indicating that the items in the task were consistent and measuring the same construct. The item-total correlations for most of the items were above .30 indicating that they were consistent with the total score, and contributing to the reliability of the task (see Table 4.2). However, for five items (*Stek, Friss, Twish, Drak,* and *Dwib*) correlations were below .30, indicating that the items were not consistent with the total score and were not contributing to the reliability of the task. The low correlations of these items may be due to restriction of range as the participants overall, as well as in each group performed at ceiling on them. However, the alpha-if-item-deleted figure for the items indicated that the reliability of the scale could not be improved by deleting them.

Table 4.2 Item-total Statistics for Phoneme Deletion Task (N = 36)

Phoneme Deletion Items	Corrected item-total correlation	Cronbach's alpha if item deleted
Stek	.22	.88
Proosh	.52	.88
Friss	.12	.89
Twish	-.00	.89
Drak	.09	.89
Spol	.56	.88
Treap	.48	.88
Gleb	.33	.88
Bloach	.56	.88
Pleem	.66	.88
Dwib	.28	.88
Smup	.35	.88
Sont	.71	.87
Doolt	.46	.88
Nast	.74	.87
Hink	.53	.88
Fesp	.55	.88
Bilk	.56	.88
Semp	.74	.87
Jast	.70	.87
Loask	.34	.89
Welp	.67	.87
Shand	.59	.88
Pelf	.70	.87

Note. The squared multiple correlations were not calculated as the determinant of the covariance matrix was zero or approximately zero.

For block one (CCVC items), reliability was adequate with coefficient α = .71 indicating that the items in this block were consistent and measuring the same underlying construct. Similar to the task as a whole, the item-total correlations for most of the items were above .30 indicating that the items were consistent with the total score and contributing to the reliability of the task (see Table 4.3). Five items (*Stek, Friss, Twish, Drak,* and *Gleb)* had item-total correlations below .30, indicating that the items were not consistent with the total score and were not be contributing to the reliability of the block. With the exception of *Gleb*, the items with low item-total correlations in this block were the same for the task as a whole. Consequently, and similarly to the task as a whole, the low correlations of the five items in this block may be due to restriction of range, as a consequence of the participants performing at ceiling on them. The restricted range of the items will also affect the reliability of the block and contribute to the lower reliability. The alpha-if-item-deleted figures for the items in the block indicated that the reliability could not be improved by deleting any items.

Table 4.3 Item-total Statistics for Block 1of the Phoneme Deletion Task (N = 36)

Phoneme Deletion Block 1 Items	Corrected item-total correlation	Cronbach's alpha if item deleted
Stek	.25	.70
Proosh	.63	.63
Friss	.01	.73
Twish	.09	.72
Drak	.09	.71
Spol	.51	.66
Treap	.51	.66
Gleb	.20	.71
Bloach	.51	.66
Pleem	.57	.67
Dwib	.44	.67
Smup	.33	.69

Note. The squared multiple correlations were not calculated as the determinant of the covariance matrix was zero or approximately zero.

For block two, the reliability was good and equal to that of the overall task coefficient $\alpha = .88$, indicating that the items in this block are consistent and measuring the same underlying construct. The item-total correlations for all the items were above .30, indicating that all the items were consistent with the total score and were contributing to the reliability of the task (see Table 4.4). The higher reliability and item-total correlations of this block provide some support for our explanation of the low item-total correlations of the five items in block one and hence its lower reliability. The performance of the participants in this block was more variable than in block one, with a lower percentage of participants (overall as well as each group) performing at ceiling. Similarly to the task as a whole and block one, for this block, the alpha-if-item-deleted figure for each item indicated that the reliability of the scale could not be improved by deleting any of the items. The two blocks of the task were also significantly correlated for accuracy $r(36) = .68$, $p < .001$ and speed $r(34) = .64$, $p < .001$ indicating stability across the blocks.

Table 4.4 Item-total Statistics for Block 2 of the Phoneme Deletion Task (N = 36)

Phoneme Deletion Block 2 Items	Corrected item-total correlation	Cronbach's alpha if item deleted
Sont	.72	.87
Doolt	.35	.88
Nast	.86	.87
Hink	.63	.88
Fesp	.59	.88
Bilk	.67	.88
Semp	.86	.87
Jast	.80	.87
Loask	.31	.89
Welp	.63	.87
Shand	.53	.88
Pelf	.71	.87

Note. The squared multiple correlations were not calculated as the determinant of the covariance matrix was zero or approximately zero.

To summarise, the Phoneme Deletion task as a whole had good reliability α = .88, indicating that the items in the task are consistent and may be measuring the same underlying construct. The reliability of block two was the same as that of the overall task; however the reliability of block one, although adequate α = .71, was less than ideal. The reliability of block one was adversely affected by five items with low item-total correlation and replacing these may increase its reliability as well as that of the scale as a whole.

4.4.2.2 Internal consistency reliability of the Nonword Repetition task.

The internal consistency reliability of the Nonword Repetition task was examined. The reliability of the task was coefficient α = .44, which is below an acceptable level, indicating that the items in the task were not consistent and may not be measuring the same underlying construct. With the exception of two items (*Tudryonippor* and *Stradiadtionmoy* – see IPA transcriptions of pronunciations in Table 4.5), the item-total correlations for all the other items were below .30, indicating that they were not consistent with the total score and were not contributing to the reliability of the task (see Table 4.5). Conversely, for *Tudryonippor* and *Stradiadtionmoy* their correlations indicated that they were consistent with the total score and were contributing to the reliability of the task. This was supported by the squared multiple correlations (R^2) figures for the task which indicated that *Tudryonippor* and *Stradiadtionmoy* contributed the most to the reliability of the task. Despite the low item-total correlations of these items, the alpha-if-item-deleted figures indicated that the reliability of the task would not be improved by deleting them. The low reliability of the task may be due to the low item-total correlations of most of the items and thus the task might be improved by replacing these with ones that are more consistent in order to make it suitable for inclusion in the BALI.

Table 4.5 Item-total Statistics for Nonword Repetition Task with Phonetic Transcriptions for Pronunciation in Square Brackets (N = 36)

Nonwords	Corrected item-total correlation	Squared multiple correlation	Cronbach's alpha if item deleted
Melpersai [‘mɛlpɚsaɪ]	.18	.12	.42
Nimutyco [‘njmjuːtɪko]	.02	.31	.47
Fibeshanoll [fɪbə’ʃænɔl]	.23	.41	.40
Nisiverupy [nɪsɪ’vɛɹʊpi]	-.04	.21	.45
Stamundingper [‘stæmʊndɪŋpɚ]	.21	.28	.40
Raksibledercon [ræksɪ’blɛdɚkɒn]	-.03	.26	.50
Gionityclafiden [‘ʤɒnɪtɪklæfɪdən]	-.05	.33	.50
Tudryonippor [tʊdɹɪ’ɒnɪpɚ]	.54	.54	.26
Clirestaibispion [klɪɹɛstaɪ’bɪspɪən]	.23	.18	.39
Stradiadtionmoy [stɹædɪ’ædʃənmɔɪ]	.53	.49	.25

4.4.2.3 Reliability of RAN and Inhibition tasks.

The reliability of the RAN and Inhibition (Opposite Direction trials) tasks was not assessed in terms of internal consistency; however the correlations of the different trials of the tasks suggest that they are likely to be reliable in terms of their stability. For the RAN tasks (Digits and Objects), two trials were administered, the second immediately after the completion of the first. The items in both trials were pseudorandomly presented and the order of presentation of the items in the two trials differed from the first. As reported in the data reduction and preliminary analyses section, both trials of the tasks were highly correlated RAN Digits $r(36) = .96$, $p < .001$and RAN Objects $r(36) = .88$, $p < .001$. Similarly, for the Inhibition tasks, two trials of the Opposite Direction portion of the task were administered, the second immediately after the completion of the first. The order of presentation of items in the second trial of the task was the reverse of that of the first. Both trials were highly correlated $r(36) = .90$, $p < .001$.

4.4.3 A Validation of the BALI for use with Adults

4.4.3.1 Comparisons of the performance of the dyslexic and control groups on the experimental measures.

If the BALI is to be an effective dyslexia screening test, it needs to be able to discriminate between the performances of adults with and without dyslexia. To this end, we compared the performances of the groups on the combined tasks of the BALI as well as on the tasks individually. The descriptive statistics of each group's performance on the BALI tasks along with Cohen's *d* are reported in Table 4.6. The results of the MANOVA analysis with the BALI tasks as the dependent variables and group (dyslexics and controls) as the independent variable indicated a significant effect of group on the tasks combined $F(8, 25) = 7.35$ $p < .001$; Pillai's Trace = 0.70. Follow up ANOVA (with Bonferroni adjusted alpha level of .006) revealed that, with the exception of the Inhibition task ($F(1, 32) = 4.13$ $p = .051$), the groups' performances were significantly different on all tasks. The dyslexic group performed less well than the control group, obtaining lower scores or taking more time to complete the tasks. Additionally, Cohen's *d* effect size indicated that the differences in the groups' performances on the tasks were also very large. The lack of a significant group difference on the Inhibition task suggests that the dyslexic group may not have an inhibition deficit. Alternatively, the Inhibition task may not be sufficiently sensitive to detect any deficit that may exist. Another possible though unlikely explanation (given the significant differences on the other tasks) is that of a lack of power. Nevertheless, given the moderate group difference as indicated by the effect size $d = 0.55$, further research with the task is required in order to eliminate the latter possibility.

Table 4.6 Mean Scores (Standard Deviation in Parenthesis) of Dyslexic and Control Groups on BALI Tasks and Effect Size (N = 36)

Experimental Tasks	Dyslexics (*n* = 18)	Controls (*n* = 18)	Cohen's *d*
Phoneme Deletion Accuracy[a]	19.56 (3.88)	22.56 (2.18)	.99
Phoneme Deletion Speed	131.00 (20.10)	100.88 (24.76)	1.34
Nonword Repetition Accuracy[b]	5.44 (1.69)	7.67 (1.09)	1.60
RAN Digits Time	19.11 (4.74)	13.44 (3.04)	1.46
RAN Objects Time	27.17 (6.21)	20.47 (2.86)	1.48
1 Minute Reading Fluency[c]	98.94 (18.05)	120.78 (19.90)	1.15
2 Minute Spelling Efficiency[d]	20.78 (5.58)	29.06 (3.62)	1.80
Inhibition	11.03 (7.61)	7.78 (4.22)	0.55

Note. [a]maximum score = 24. [b]maximum score = 10. [c]maximum score = 150. [d]maximum score = 40.

The results of the MANOVA indicated that the battery as a whole could effectively discriminate between dyslexic and control adults and it is therefore likely that all of the tasks are suitable for inclusion in a dyslexia screening test for adults. Additionally, with the exception of the Inhibition task, all tasks individually discriminated, with very large effect sizes, between dyslexic and control adults, providing empirical evidence to support their appropriateness for inclusion in the screener. To further examine the appropriateness of the tasks, we examined the correlations of the tasks with themselves and with the standardised measures.

4.4.3.2 Correlational analyses.

The correlation analyses were conducted with the total sample (*N* = 36) of dyslexic and control participants. The performance differences between the groups on the tasks were primarily quantitative, and the distributions of scores seemed to fall on a continuum with the dyslexic group at the lower end. Thus, the groups were combined to maximise the heterogeneity of the sample and to reduce restricted range effects. All the tasks included in the BALI were significantly correlated with at least two of the other tasks with moderate to high correlations ranging from r= .36, to r = .76 (see Table 4.7). In fact, most (five tasks) were significantly correlated with five or more tasks. The task with the best correlations was Phoneme Deletion Speed which was correlated with all other tasks, while Phoneme Deletion Accuracy and Inhibition had the lowest and correlated with only two other tasks. The correlation of the Inhibition tasks with only two tasks is understandable, given that the task was designed to measure a construct that is unrelated to literacy skills. For the Phoneme Deletion Accuracy task, its lack of correlations with most of the literacy measures is probably due to the restricted range of the scores on the task. Generally, the higher the inter-item correlation of items in a measure, the higher its internal consistency reliability (Streiner, 2003). Thus, the high correlations of the tasks suggest that a battery with them combined is likely to be reliable.

4.4.3.3 Concurrent validity.

In order to assess the concurrent validity of the experimental literacy tasks we examined their correlations with the standardised measures of literacy. Both tasks, 1Minute Reading and 2 Minute Spelling were significantly highly correlated with the standardised measure of the same construct, WRAT IV Word Reading $r(36) = .63$, $p < .001$, and WRAT IV Spelling $r(36) = .74$, $p < .001$ respectively. Additionally, the measures also correlated highly with standardised measures of similar (literacy) constructs. 1 Minute Reading correlated highly with WRAT IV spelling $r(36) = .66$, $p < .001$, and 2 Minute Spelling with the WRAT IV reading measure $r(36) = .74$, $p < .001$. The significant and high correlations of these two experimental tasks with the standardised measures of reading and spelling provide evidence of their concurrent and construct validity. That is, the tasks are likely measuring the constructs that they were designed to measure, reading and spelling abilities, and therefore may be considered valid measures of these abilities. The tasks that assess literacy related skills (namely: Phoneme Deletion Accuracy, Phoneme Deletion Speed, Nonword Repetition, and RAN Digits and Objects) had moderate to high correlations with the standardised measures of reading and spelling. This provides support for the construct validity of these tasks. For the Phoneme Deletion Accuracy task, although its correlations with the standard literacy measures were in line with the other tasks in the BALI, it is likely that these correlations were depressed due to the restricted range of the scores on the task. None of the literacy tasks (BALI or standardised measures) correlated with the Inhibition task, providing some evidence of their discriminant validity. In addition, none of the BALI tasks were correlated with the cognitive measure (nonverbal IQ) not associated with literacy skills.

The results indicate that the tasks included in the BALI (with the exception of the Inhibition tasks) are capable of discriminating, individually and collectively, between adults with and without dyslexia. The high correlations between the tasks suggest that the measures may have good internal consistency while the high correlations with the standardised measures support the concurrent and construct validity of the tasks.

Table 4.7. Correlations between Experimental Tasks and Standardized Measures (N = 36)

Measures	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Phoneme Deletion Accuracy	-	-.56**	.26	-.16	-.14	.22	.38*	.02	.31	.27	.51**	.46**	.11
2. Phoneme Deletion Speed		-	-.41*	.44**	.48**	-.49**	-.38*	.36*	-.11	-.16	-.43*	-.39*	-.23
3. Nonword Repetition Accuracy			-	-.33	-.38*	.50**	.74**	-.26	.19	.12	.53**	.52**	.00
4. RAN Digits Time				-	.76**	-.75**	-.50**	.47**	-.06	-.11	-.54**	-.46**	.11
5. RAN Objects Time					-	-.68**	-.58**	.27	-.28	-.36*	-.62**	-.60**	-.07
6. 1 Minute Reading Fluency						-	.65**	-.25	.24	.23	.63**	.66**	.10
7. 2 Minute Spelling Efficiency							-	-.22	.33	.36*	.74**	.74**	.04
8. Inhibition								-	.10	-.12	-.25	-.03	-.25
9. WRIT Matrices									-	.27	.38*	.41*	.20
10. WRIT Vocabulary										-	.49**	.42*	.49**
11. WRAT IV Word Reading											-	.82**	.25
12. WRAT IV Spelling												-	.05
13. WRAT IV Sentence Comprehension													-

* $p < .05$. ** $p < .01$

4.5 Discussion

The main purpose of this study was to investigate the capacity of the tasks to be included in the BALI, a new adult dyslexia screening test, to discriminate between adults with and without dyslexia, individually and collectively, and to evaluate their psychometric properties. We compared the performance of adults with and without dyslexia on the tasks as well as measures of cognitive and literacy abilities, and assessed the validity and reliability of the tasks with correlation and reliability analyses. The groups were comparable on the standardised measures of verbal and nonverbal IQ and comprehension but differed significantly on the measures of reading and spelling, with very large effect sizes, $d = 1.46$ and $d = 1.32$ respectively. Although, the dyslexics obtained significantly lower scores than the controls on the reading and spelling measures, their performances were in the average range, indicating that they may be considered compensated or high functioning dyslexics.

4.5.1 Reliability

The reliability analysis of the Phoneme Deletion task as a whole indicated that it had good internal consistency coefficient $\alpha = .88$, which is well within the recommended range .80 -.90 (Anastasi & Urbina, 1997; Field, 2009; Kline, 2000). This suggests that the items in the Phoneme Deletion task are consistent and are measuring the same underlying construct. The reliability of the task compares favourably to similar tasks in other adult dyslexia screening tests. The DAST manual reports test-retest reliability of $r = .90$ for its Phonemic Segmentation task (Nicolson & Fawcett, 1998), while the reported internal consistency reliability of the Spoonerism task of the YAA-R is coefficient $\alpha =.76$ (Warmington et al., 2013). In addition to assessing the reliability of the Phoneme Deletion task as a whole, we also assessed each block. The reliability of block two of the task was good $\alpha = .88$ and the same as the task as a whole. For block one, it was acceptable $\alpha =.71$ but below .80 what is considered ideal (Anastasi & Urbina, 1997). The analysis revealed that five items in block one had low item-total correlations which contributed to the lower reliability of this block. The low item-total correlations may be attributable to the restricted range of participants' scores on the items as most performed at ceiling on them. Administration of the task to a larger sample of individuals with more variable performance is desirable to test this hypothesis. The reliability of the block could be improved by replacing the five items with low item-total correlation with items that are more consistent with the other items in the block. However, such a manipulation would work against our objective to create a phoneme deletion block of items that poses minimal difficulty in terms of accuracy, and instead permits a valid estimate of phoneme deletion speed, a manipulation that was successfully accomplished in block one here. In addition, other methods of assessing reliability such as test-retest should be undertaken in order to validate the reliability of the task across time.

For the Nonword Repetition task internal consistency was coefficient $\alpha = .44$, which is below what is considered an acceptable level (Anastasi & Urbina, 1997; Field, 2009; Kline, 2000). The low reliability may be attributed to the low item-total correlations of eight of the 10 items in the task. The alpha-if -deleted figures for the items indicated that deleting items would not improve the reliability of the task. Consequently, the reliability of the task would need to be improved by the addition of new items with higher inter-item correlations, and the deletion of items with low correlations. We did not directly assess the reliability of the RAN and Inhibition tasks, however, and indication that the tasks are likely to be reliable was obtained from the correlations of the different trials of the tasks. In addition, all the tasks in the BALI were significantly correlated with at least two others, and most with five or more, with moderate to high correlations. However, the reliability of the 1 Minute Reading and the 2 Minute Spelling tasks also needs to be assessed in order to determine if they are consistent. Additionally, for the Phoneme Deletion and Nonword Repetition (after adjustments to improve reliability) tasks, test-retest reliability should be assessed to ensure that the tasks are consistent across time.

4.5.2 Validity

Evidence of the validity of the BALI was provided by the difference in the performance of the groups on its tasks. The performance of the groups differed on the battery as a whole, and with the exception of the Inhibition task, on each task individually, demonstrated its capacity to discriminate between them. Furthermore the effect sizes indicated that these differences were not only statistically significant, but also very large, ranging from .99 to 1.80. The dyslexic group’s performances on the tasks were less accurate and slower than the controls. The differences between the groups on the experimental tasks which measure literacy and literacy related skills were expected, given the nature of the deficits exhibited by adults with dyslexia. Additionally, as noted in the introduction, research indicates that the tasks can effectively discriminate between individuals with and without dyslexia. Also, our results are similar to those of several studies that have compared the performance of adults with and without dyslexia on similar measures (Miller-Shaul, 2005; Hatcher et al., 2002; Ramus et al., 2003b). However, the unique findings of this pilot study are that most of the tasks can be combined into a quick and effective dyslexia screening test for adults.

For the Inhibition task, the lack of a significant group difference suggests that the dyslexic groups may not have an inhibition deficit. Our results are contrary to most of the existing studies with children and adults with dyslexia that have reported the existence of an inhibition deficit (Baker & Ireland, 2007; Brosnan et al. 2002; Helland & Asbjørnsen, 2000). However, our results are similar to those of Lindgrén and Laine (2011) who found that the performance of university students with dyslexia did not differ from their counterparts without dyslexia, on the Stroop colour test. It is possible that if adults with dyslexia have an inhibition deficit, only a sub-group of individuals with the disorder are affected, although there was no evidence of this in our data. Alternatively, our failure to identify the deficit could be because our task is not sufficiently sensitive to detect any deficit that may exist. Also, although not significant, there was a moderate difference in the performance of the groups on the task $d = 0.55$; this suggests a slight though unlikely possibility of a lack of power. The lack of power is unlikely given the significant results on the other tasks with the same sample. Also, Brosnan et al. (2002), with an even smaller sample of 18 participants (9 with and 9 without dyslexia), found that the dyslexic participants were more inhibited on a group embedded figures test, where they were required to identify simple figures in a complex visual array. However, the difference in the results obtained by Brosnan et al. (2002) may be due to the type of inhibition task used which could be a more sensitive measure of inhibition. In any event, given the limited number of studies that have investigated an inhibition deficit in adults with dyslexia, and in order to eliminate the slight possibility that the lack of group difference was due to a lack of power, further research with a large sample is required.

Evidence of the concurrent validity of the experimental tasks in the BALI was provided by the correlation analysis. Most of the tasks selected for inclusion in the BALI were designed to assess literacy and related skills. In essence, if the experimental tasks are valid measures, they should correlate significantly with the standardised measures of literacy. This in fact was the case, as the correlation analysis revealed, the experimental tasks (with the exception of the Inhibition task) were significantly correlated with the standardised measures of reading and spelling with moderate to high correlation, thus providing evidence of the concurrent validity of the tasks. That is, they are actually measuring the construct that they were designed to measure.

4.6 Conclusion

The results of the study provide empirical evidence of the capacity of the tasks in the BALI to discriminate between adults with dyslexia, including compensated dyslexics, and typical readers. It also provided evidence of the validity and reliability of tasks. However, the results also indicated that the reliability of block one of the Phoneme Deletion is less than ideal while the reliability of the Nonword Repetition task is below an acceptable level. Improvements are therefore required in the reliability of these tasks. Additionally, the reliability of the other tasks (1 Minute Reading, 2 Minute Spelling, RAN, and Inhibition) should be assessed to ensure that they are at an acceptable level. For the Inhibition task, more research is also required in order to ascertain its capacity to distinguish between adults with and without dyslexia. The main study of the BALI which follows was designed to address these issues.

Chapter 5: Development of the Bangor Adult Literacy Index: Main Study (Study 2b)

5.1 Introduction

With a relatively small sample of participants, the results of the pilot study (Chapter 4 Study 2a) provided empirical evidence of the capacity of the tasks (with the exception of the Inhibition task) included in the BALI, to discriminate (individually and collectively) between adults with and without dyslexia, and therefore supported their inclusion in the screening test. Additionally, the pilot study provided evidence of the concurrent validity of the 1 Minute Reading and 2 Minute Spelling tasks, as well as the construct validity of these tasks and the others included in the BALI, with the possible exception of the Nonword Repetition task. Further, it provided evidence of the internal consistency of the Phoneme Deletion task. The reliability of the Nonword Repetition task was below an acceptable level. We conducted a follow up main study with the overall aim of further evaluating the tasks included in the BALI with a larger more representative sample of students. Specifically, we wanted to further evaluate the capacity of the tasks to effectively discriminate between adults with and without dyslexia individually and collectively. Additionally, we adjusted the Phoneme Deletion, Nonword Repetition, and 1 Minute Reading tasks (adjustments are described below along with rationales) and therefore wanted to re-assess these. Finally, we examined the test-retest reliability of the 1 Minute Reading and 2 Minute Spelling tasks, which was not done in the pilot study.

5.1.1 Error in the Reliability Analysis of Pilot Study

The Phoneme Deletion task was altered slightly for the present study with a view to improving its psychometric properties. A first pass of scoring in the Pilot Study suggested that although the task as a whole was reliable, block 2 (CVCC) had poor internal consistency. In fact, a second pass through the scoring (which provided the data for the analyses of Chapter 4), removing the earlier error of excluding participants who in fact scored zero (because they found the task too difficult), revealed that the Phoneme Deletion task has good reliability, as reported in Chapter 4. However, this second pass reanalysis took place after the data collection had begun for the present study with the adjusted measure. Consequently, we detail below the adjustments that were made, and report all results for the adjusted measure – albeit that the original measure had adequate psychometric properties.

5.1.2 Adjustments to BALI Tasks

Adjustments were made to the Phoneme Deletion and Nonword Repetition tasks to improve the internal consistency of the tasks. Although the results of the pilot did not indicate that the 1 Minute Reading task required any adjustments, it did raise some concerns about the length of the task, which are elaborated below.

5.1.2.1 Adjustments to 1 Minute Reading task.

The task was created to measure reading fluency and assessed the number of words read accurately in one minute. Although the results of the pilot study did not raise any concerns regarding its general structure, two of the control participants had completed the task in less than one minute. In order to ensure that the time required to complete the task was not less than a minute, we increased the number of words in the list. The original task consisted of 144, one to six syllables words with mean print familiarity *M* = 574.06, and mean Kucera-Francis written frequency *M* = 691.32. We increased the total number of words from 144 to 150, reduced the number of one-syllable words, and increased the number of mutli-syllable words. An example of the stimuli is shown in Appendix I. As in the pilot study, the words were selected using the MRC Psycholinguistic Database. The revised task consisted of: 20 one-syllable words, 60 two-syllable words, 53 three-syllable words, 13 four-syllable words, 3 five-syllable words, and 1 six-syllable word. The mean print familiarity and mean Kucera-Francis written frequency was very similar to the original task: *M* = 574.37 and *M* = 651.17 respectively. The words were presented in the same format as the pilot study, that is, in three columns on each side of an A-4 sheet and were printed in lower case letters, 16-point Arial bold font, ordered by the number of syllables with one syllable words first. The administration instruction for the task was the same as in the pilot study.

5.1.2.2 Adjustments to Phoneme Deletion task.

The task was adapted from Judge et al. (2006) and was created as a measure of phoneme awareness, measuring the ability to produce and manipulate speech sounds. The results of the analysis (based on first scoring on which the adjustments were made) revealed that the reliability of the task as a whole was adequate, coefficient α = .73. For block 1 of the task, reliability was also acceptable, coefficient α = .71. The reliability of the block was affected by the low correlations of the five items due to restriction of range, as participants performed at ceiling on the items. The items in block 2 of the task were more difficult and therefore some dyslexic participants could not complete the task resulting in a smaller sample size and skewed results. The reliability of the block was below an acceptable level, coefficient α = .49. The item-total correlations of five of the items were below .30, indicating that they were not consistent with the total score and may not be contributing to the reliability of the task. Additionally, two of the items in the task, *Nast* and *Semp*, were excluded from the analysis due to zero variance because of ceiling effects.

Based on these results, to improve the reliability of the task as a whole, we focused our adjustments on block 2 only, and attempted to increase the internal consistency of the items in this block. As a result, the two items (*Nast* and *Semp*) that had generated no variance due to ceiling level performance were exchanged, for *Galn* and *Heshp*. The word *Bilk* was also replaced; it had a very low negative item-total correlation indicating that it did not correlate well with the total score. Furthermore, inter-item correlations revealed that it was negatively correlated with all but one of the other items in the block; this item was replaced with item *Thoist*. An example of the stimuli is shown in Appendix J.

5.1.2.3 Adjustments to Nonword Repetition task.

The task was created to assess phonological verbal short term memory and articulation fluency. The results of the reliability analysis revealed that the reliability of the task was below an acceptable level, coefficient α = 44. The items in the task were not consistent and the item-total correlation indicated that only two of the 10 items had a correlation that was above the minimum acceptable. Additionally, the alpha-if-item-deleted figures indicated that the reliability of the task could not be improved by deleting any of the items. Therefore, major adjustments were made to the task in an effort to make the items more consistent and to improve reliability. Four words from the original version of the task were retained: *Stamundingper, Tudryonippor, Stradiadtionmoy,* and *Gionityclafiden*. *Tudryonippor* and *Stradiadtionmoy* were retained because they had good item-total correlations in the original task although we recognise that this is likely to change with a different combination of words. *Stamundingper* and *Gionityclafiden* were retained because their overall correlations with *Tudryonippor* and *Stradiadtionmoy* were better than most of the other items, and the performance of the groups differed on these tasks. Six new words were therefore included in the task. An example of the stimuli is shown in Appendix K. The words were created using the same method used in the pilot study by manipulating syllables and changing the sounds of real words. An effort was made to include stimuli with a greater number of consonant clusters than in the original set. The six words added to the task were *Thrasplabity, Vizationlisi, Senpretaretive, Breplaposenciary, Scriggaflumianal, and Rystupicarelapto*. The administration instruction for the task was the same as the pilot study.

5.2 Method

5.2.1 Participants

Participants were 80 native English speaking undergraduate and postgraduate students, 40 dyslexics, and 40 controls, recruited from Bangor University. All participants in the dyslexic group had a formal diagnosis of dyslexia and were registered with the Dyslexia Centre at the University. The group comprised 23 (57.5%) females and 17 (42.5%) males, mean age M = 21.73 (SD = 5.66), and range 18 – 50 years. The control participants reported no learning or literacy difficulties and all satisfied the minimum criteria of standard scores of 85 and above on cognitive and literacy measures. The group comprised 34 (85%) females and 6 (15%) males, mean age M = 19.28 (SD = 1.95), and range 18 – 30 years. The age difference between the groups was significant t (48.11) = -2.59, p = .013. The control participants were compensated with £4 per hour of printer credits for their participation. Dyslexic participants were compensated with £8 per hour to ensure the required numbers were recruited in the shortest possible time.

5.2.2 Measures

Most of the tasks administered to the participants were the same as those described in the pilot study (Chapter 4 Study 2a) except for the adjustments highlighted in the introduction (above) and are therefore not described here. The WRAT IV Sentence Comprehension subtest was not administered to two participants in the control group. Additionally, participants were also assessed on the Instines a computerised dyslexia screening test developed to identify characteristics associated with dyslexia in adolescents and adults. A description of the Instines is given in Chapter 6 which reports an evaluation of the test.

5.2.3 Procedure

The study was completed in two parts, with the second part scheduled 14 days after the first. The average number of days between the completion of the first (test) and the second (retest) parts of the study was M = 15 with a range of 14 – 21 days. For both parts of the study participants were assessed individually, with part one lasting approximately 80 – 90 minutes and part two approximately 20 – 30 minutes. Tasks were administered in a fixed order, in part one, as follows: Reading, Sentence Completion, and Spelling subtests of the WRAT IV, Matrices and Vocabulary subtests of the WRIT, and the experimental tasks of the BALI as follows: RAN Objects, Phoneme Deletion, 1 Minute Reading Fluency, RAN Digits, 2 Minute Spelling Efficiency, Inhibition, Nonword Repetition; the Instines was administered last. Part two of the study was used to retest participants on four of the tasks performed in part one of the study in order to examine the test-retest reliability of the tasks. The tasks administered, in the order of administration were: (1) Phoneme Deletion, (2) One Minute Reading Fluency, (3) Two Minute Spelling Efficiency, from the BALI, and (4) the Instines (which is reported in Chapter 6). The standardised measures were administered according to the published guidelines and the experimental task in accordance with established instructions. Ethical approval for the study was granted by the School of Psychology, Bangor University.

5.3 Data Analysis

Prior to analyses, the data set was checked for outliers. To identify outliers the mean for each group was calculated separately. Outliers above or below 2.5 standard deviations from the mean were adjusted to the score corresponding to 2.5 standard deviations from the mean. For the background measures, there were no outliers for the control group and only one score was adjusted for the WRIT Vocabulary subtest for the dyslexic group. For the experimental tasks, only one or two scores (5% or less of sample) were adjusted for each group on the following tasks: RAN Digits, RAN Objects, Phoneme Deletion Accuracy, Phoneme Deletion Speed, and Nonword Repetition Accuracy. As in the pilot study, we created composite scores for the two RAN tasks (Digits and Objects) by averaging the time taken to complete both trials of the tasks. The trials assessed the same construct and were highly correlated, RAN Digits $r(80) = .92$, $p < .001$ and RAN Objects $r(80) = .92$, $p < .001$. Similar to the pilot study, the groups did not differ in the number of errors made on the RAN Objects task; (controls $M = .14$, $SD = .30$, and dyslexics $M = .23$, $SD = .42$), $t(78)$ -1.07, $p =$ .289. However, they differed on the RAN Digits task; (controls $M = .00$, $SD = .00$, and dyslexics $M = .20$, $SD = .61$), $t(39)$ -2.08, $p = .044$, with the dyslexic group making more errors. Nevertheless, the majority of the dyslexics (85%) made no errors, suggesting this may not be important for distinguishing between the groups, and error rates were not included in the analyses. For the Inhibition task, scores were calculated by averaging the time taken to complete the two Opposite Direction trials, and subtracting the time taken to complete the Same Direction trial; we examined the correlation of the Opposite Direction trials and they were highly correlated $r(80) = .82$, $p < .001$. For the Same Direction trial, and trial 2 of the Opposite Direction, the groups did not differ in the number of errors made; (controls $M = .05$, $SD = .22$, and dyslexics $M = .15$, $SD = .53$), $t(51.97)$ -1.10, $p = .278$, and (controls $M = .32$, $SD = .97$, and dyslexics $M = .55$, $SD = 1.13$), $t(78)$ -.96, $p = .343$ respectively. For trial 1 of the Opposite Direction, the dyslexic group made significantly more errors; (controls $M = .20$, $SD = .46$, and dyslexics $M = 1.00$ $SD = 1.71$), $t(44.72)$ -2.86, $p = .006$. However, similar to the RAN tasks, the majority (77.5%) of the dyslexics made one or no error, suggesting this may not be important for distinguishing between the groups.

For the Nonword Repetition task, here again the results differed from that of the pilot study as the groups differed not only in their accuracy scores on the task; (controls *M* = 6.63, *SD* = 1.56, and dyslexics *M* = 5.48 *SD* = 1.84), $t(78)$ 3.01, $p = .003$, but also in the time taken to complete the task; (controls *M* = 35.14, *SD* = 3.69, and dyslexics *M* = 41.48 *SD* = 5.84), $t(65.86)$ -5.81, $p < .001$. However, as we were more interested in the difference in the groups' ability to perform the task (and not the time to complete), and for consistency with the pilot study, only the accuracy scores were included in the analysis. The 1 Minute Reading task was designed to be a measure of reading fluency and not accuracy. The results of the pilot study suggested that this was achieved, as the number of errors made by both groups was negligible. For this study, however, the groups differed both in the number of words read accurately $t(73.73)$ 4.98, $p < .001$, and errors made $t(48.68)$ -3.03, $p = .004$ with the dyslexics making more errors. Nevertheless, the number of errors made by the both groups was negligible, dyslexics *M* = 1.28 *SD* = 1.77 and controls *M* = .38, *SD* = .63.

The data set was also checked for appropriateness for multivariate analysis (MANOVA) and logistic regression. The assumptions of the logistic regression were satisfied. Two MANOVAs (followed up with univariate analyses (ANOVA) with Bonferroni adjusted alpha level) were conducted, one for the background measures and the other for the tasks included in the BALI. For the analysis of the background measures, the assumption of equality of error variances was violated by the WRAT Spelling subtest, therefore a more conservative alpha of .01 was used for significance, as recommended by Tabachnick and Fidell (2007). For the MANOVA of the BALI tasks, the assumptions of homogeneity of variance-covariance and equality of error variances were violated, and thus an alpha of .01 was also used and Pillai's Trace statistic reported as it is robust to the violation of assumptions (Tabachnick & Fidell, 2007). The Shapiro-Wilk statistic was used to test normality and variables not normally distributed were transformed. However, the results of the MANOVA with the untransformed variables are reported because analyses with the transformed variables produced the same pattern of results. We first compared the groups on the background measures and then assessed the reliability of four of the tasks in the BALI. Finally, we assessed the validity of the BALI by comparing the performance of the groups on the tasks (collectively and individually) and examined their correlations with the standardised measures of literacy (WRAT IV subtests).

5.4 Results

5.4.1 Comparisons of Performance of Dyslexic and Control groups on the Background Measures.

The descriptive statistics of the groups on the background measures along with Cohen's *d* are reported in Table 5.1. MANOVA analysis with group as the independent variable and the background measures as the dependent variables indicated a significant difference between the groups on the set of measures $F(5, 72) = 14.08$, $p < .001$; Pillai's Trace = 0.49. Follow up ANOVA (with Bonferroni adjusted alpha level of .01) revealed that the group's performances were significantly different on the measures of WRAT IV Word Reading, $F(1, 76) = 27.27$, $p < .001$ and WRAT IV Spelling, $F(1, 76) = 62.78$, $p < .001$, but not on the subtests of the WRIT and the WRAT IV Sentence Comprehension. On the WRAT IV Reading and Spelling subtests, the dyslexic group performed less well than the control group. Cohen's *d* effect size indicated that the differences in the group's performance on these measures were also very large, with WRAT IV Reading $d = 1.19$ and WRAT IV Spelling $d = 1.85$. Despite these differences, the scores obtained by the dyslexic group on both tasks were well within the average range. Thus, as was the case in the pilot study, the groups were similar in terms of their verbal and non-verbal IQ and only differed on their literacy skills, specifically reading and spelling. For both studies, the dyslexic groups could be described as compensated, with performances in the average range even on the literacy measures; the controls also performed in the average range but significantly better than the dyslexics on the reading and spelling measures.

Unlike the pilot study, there was a significant difference in the age of the groups, dyslexics $M = 21.73$ ($SD = 5.66$), controls $M = 19.28$ ($SD = 1.95$), $t\ (48.11) = -2.59$, $p = .013$. However, it is unlikely that this affected the results obtained here, as age is taken into account in the scoring procedure for the background measures. For the BALI tasks, we examined their correlations with age using the total sample. The results indicated that three tasks had low to moderate significant correlations with age as follows: Phoneme Deletion Accuracy $r(80) = -$

.28, $p < .05$; 1 Minute Reading $r(80) = -.24$, $p < .05$; and Inhibition $r(80) = -.30$, $p < .01$. However, because age was confounded with reading status (dyslexic vs. control), we further examined the correlations for each group and this revealed that for the control group, only RAN Objects had moderate significant correlation with age $r(40) = .35$, $p = .028$. For the dyslexic group there were no significant correlations with age. It is therefore, unlikely that the age differences between the groups will have affected the results obtained in the analyses conducted with these tasks.

Table 5.1 Mean Standard Scores (Standard Deviation in Parenthesis) of Dyslexic and Control Groups on Background Measures and Effect Size (N = 80)

Measures	Dyslexics (n =40)	Controls (n = 40)	Cohen's *d*
WRIT Matrices	106.50 (13.15)	104.98 (10.75)	0.13
WRIT Vocabulary	102.32 (11.15)	104.50 (11.43)	0.19
WRAT IV Word Reading	93.88 (7.88)	103.10 (7.62)	1.19
WRAT IV Spelling	92.32 (8.46)	111.98 (12.81)	1.85
WRAT IV Sentence Comprehension[a]	105.82 (12.54)	109.82 (9.93)	0.36

Note. [a]$n = 38$

5.4.2 An Examination of the Reliability of the Tasks in the BALI

5.4.2.1 Internal consistency and test-retest reliability of the Phoneme Deletion task.

We examined the internal consistency (Cronbach's alpha) and test-retest reliability of the Phoneme Deletion task as a whole, as well as for each block of the task. As indicated in the introduction, due to an error in the scoring of the task in the pilot study, the result of the earlier reliability analysis, based on first pass scoring, was incorrect and the subsequent adjustments made to the task were not necessary. Fortunately, these adjustments did not adversely affect the reliability of the task. The internal consistency of the task as a whole was excellent, coefficient $\alpha = .92$ indicating that the items in the task were consistent and are likely measuring the same construct. The item-total correlations for all the items were above .30, indicating that they were contributing to the reliability of the block (see Table 5.2). Additionally, the alpha-if-item-deleted figure for all the items indicated that the reliability of the scale could not be improved by deleting any items.

Table 5.2 Item-total Statistics for Phoneme Deletion Task with Percentage of Correct Responses for Dyslexic and Control Groups (N = 80)

Items	Corrected item-total correlation	Cronbach's alpha if item deleted	Dyslexics Percentage Correct	Controls Percentage Correct
Stek	.63	.92	80.0	97.5
Proosh	.65	.92	82.5	87.5
Friss	.43	.92	92.5	97.5
Twish	.64	.92	72.5	95.0
Drak	.60	.92	95.0	95.0
Spol	.46	.92	77.5	97.5
Treap	.48	.92	77.5	92.5
Gleb	.66	.92	75.0	97.5
Bloach	.59	.92	82.5	90.0
Pleem	.38	.92	87.5	95.0
Dwib	.63	.92	80.0	95.0
Smup	.62	.92	82.5	95.0
Sont	.61	.92	80.0	87.5
Doolt	.43	.92	52.5	67.5
Galn	.55	.92	60.0	82.5
Hink	.76	.92	77.5	92.5
Fesp	.65	.88	85.0	87.5
Thoist	.52	.88	67.5	85.0
Heshp	.31	.87	72.5	77.5
Jast	.74	.87	85.0	90.0
Loask	.40	.89	55.0	72.5
Welp	.61	.87	75.0	90.0
Shand	.66	.88	72.5	85.0
Pelf	.56	.87	65.0	87.5

Note. The squared multiple correlations were not calculated as the determinant of the covariance matrix was zero or approximately zero.

For block one (CCVC items), reliability was again excellent, coefficient α = .88, indicating that the items in this block were consistent and likely measuring the same construct. None of the items had item-total correlations below .30, indicating that all were contributing to the reliability of the block (see Table 5.3). The alpha-if-item-deleted figure for each item also indicated that the reliability of the scale could not be improved by deleting any of the items. The reliability is also higher than reported in the pilot study, α = .71. This increase is probably due to the larger sample size and greater variability in the performance of the participants; M = 11.00, SD = 1.62 for the pilot study with M = 10.60, SD = 2.50 for the present study.

Table 5.3 Item-total Statistics for Block 1of the Phoneme Deletion Task (N = 80)

Phoneme Deletion Block 1 Items	Corrected item-total correlation	Cronbach's alpha if item deleted
Stek	.62	.86
Proosh	.49	.87
Friss	.45	.87
Twish	.67	.86
Drak	.65	.86
Spol	.51	.87
Treap	.52	.87
Gleb	.67	.86
Bloach	.58	.86
Pleem	.45	.87
Dwib	.60	.86
Smup	.63	.86

Note. The squared multiple correlations were not calculated as the determinant of the covariance matrix was zero or approximately zero.

For block two, the reliability was also excellent, α = .86, reflecting inter-item consistency. The item-total correlations indicate that all the items were contributing to the reliability of the task (see Table 5.4). Similar to the task as a whole and block one, the alpha-if-item-deleted figure for each item in this block also indicated that the reliability of the scale could not be improved by deleting any of the items. The reliability of the block is only marginally lower than the reliability of α = .88 reported in the pilot study.

Table 5.4 Item-total Statistics for Block 2 of the Phoneme Deletion Task (N = 80)

Phoneme Deletion Block 2 Items	Corrected item-total correlation	Squared Multiple Correlation	Cronbach's alpha if item deleted
Sont	.56	.42	.85
Doolt	.44	.33	.86
Galn	.56	.47	.85
Hink	.69	.55	.84
Fesp	.68	.55	.84
Thoist	.54	.39	.85
Heshp	.30	.18	.87
Jast	.67	.61	.85
Loask	.45	.33	.86
Welp	.55	.50	.85
Shand	.64	.51	.85
Pelf	.60	.50	.85

5.4.2.2 Test-retest reliability of the Phoneme Deletion task.

The Phoneme Deletion Task was administered twice to the participants with a two weeks interval. Six participants from the control group did not participate in the second part of the study and therefore this analysis was conducted with a reduced sample of 74. Participants' scores were correlated using Pearson's *r* correlation and details of the correlations for each block as well as the total scores on the task are detailed in Table 5.5. The scores were highly correlated for each block as well as the task as a whole and were above the minimum (r = .80) required to support test-retest reliability (Kline, 2000). This suggests that scores on the task are consistent across time and indicates a 66% to 90% agreement between the participants' rank scores on the first and second administrations of the task.

Table 5.5 Correlations between Test and Re-test Scores on the Phoneme Deletion Task (N = 74)

Phoneme Deletion Task	Pearson's *r* Correlations
Phoneme Deletion Accuracy Block 1	.83
Phoneme Deletion Speed Block 1	.90
Phoneme Deletion Accuracy Block 2	.81
Phoneme Deletion Speed Block 2	.93
Phoneme Deletion Accuracy Blocks 1 & 2	.89
Phoneme Deletion Speed Blocks 1 & 2	.95

** $p < .001$

5.4.2.3 Internal consistency reliability of the Nonword Repetition task.

The internal consistency reliability of the Nonword Repetition task was also examined. The reliability of the task was $\alpha = .45$, which is below an acceptable level, indicating that the items in the task were not consistent and may not be measuring the same construct. With the exception of *Tudryonippor*, the item-total correlations for all the other items were below .30 indicating that they were not consistent with the total score and were not contributing to the reliability of the task (see Table 5.6). This was supported by the squared multiple correlation (R^2) figures for these items. Despite the low item-total correlations of these items, the alpha-if-item-deleted figures indicated that the reliability of the task would not be improved by deleting them. The results here are similar to those of the pilot study and indicate that the adjustments made to the task did not have the intended effect of improving the internal consistency of the task. These results along with the results of the MANOVA (see section on validation below), which indicates that the performance of the groups did not differ on the task, suggest that the Nonword Repetition task is not suitable for inclusion in the BALI.

Table 5.6 Item-total Statistics for Nonwords Repetition Task with Phonetic Transcriptions for Pronunciation in Square Brackets and Percentage of Correct Responses for Dyslexic and Control Groups (N = 80)

Nonwords	Corrected item-total correlation	Squared multiple correlation	Cronbach's alpha if item deleted	Dyslexics Percentage Correct	Controls Percentage Correct
Stamundingper [‘stæmʊndɪŋpɚ]	.24	.09	.40	65.0	65.0
Thrasplabity [‘θɹæsplabɪtɪ]	.14	.06	.43	75.0	87.5
Tudryonippor [tʊdɹɪ’ɒnɪpɚ]	.33	.17	.37	65.0	87.5
Stradiadtionmoy [stɹædɪ’ædʃənmɔɪ]	.12	.12	.44	60.0	67.5
Vizationlisi [vɪ’zeɪʃənlɪsɪ]	.20	.12	.41	60.0	85.0
Senpretaretive [‘sɛnpɹɛteɹətɪv]	.03	.10	.46	80.0	87.5
Gionityclafiden [‘ʤɒnɪtɪklæfɪdən]	.23	.14	.40	60.0	77.5
Breplaposenciary [‘bɹɛpləpɔnsɛnʃəɹi]	.06	.07	.47	35.0	52.5
Scriggaflumianal [‘skɹɪgəflumɪænəl]	.20	.08	.41	22.5	20.0
Rystupicarelapto [ɹaɪ’stʊpɪkæɹəlapto]	.24	.11	.40	25.0	30.0

5.4.2.4 Test-retest reliability of the 1 Minute Reading and 2 Minute Spelling tasks.

We assessed the test-retest reliability of the 1Minute Reading and 2 Minute Spelling tasks by correlating participants' scores across the two administrations of the tasks. The scores were highly correlated for both tasks, 1Minute Reading r (74) = .91, p <.001 and 2 Minute Spelling r (74) = .89, p <.001 and were above the minimum (r = .80) required to support test-retest reliability (Kline, 2000). This suggests that scores on the task are consistent across time and indicate an 83% and 79% (respectively) agreement between the participants' rank scores on the first and second administrations of the task. These results support the reliability (across time) of the tasks.

5.4.2.5 Reliability of RAN and Inhibition tasks.

The reliability of the RAN and Inhibition tasks was not directly assessed; however, the correlations of the different trials of the tasks suggest that they are likely to be reliable. For the RAN tasks (Digits and Objects), two trials were administered, the second immediately after the completion of the first. The items in both trials were randomly presented and the order of presentation of the items in the two trials differed from the first. As reported in the data analysis section, both trials of the tasks were highly correlated: RAN Digits $r(80) = .92$, $p < .001$ and RAN Objects $r(80) = .92$, $p < .001$. Similar to the RAN tasks, for the Inhibition tasks, two trials of the Opposite Direction portion of the task were administered, the second immediately after the completion of the first. The order of presentation of items in the second trial of the task was the reverse of that of the first. Both trials were highly correlated $r(80) = .82$, $p < .001$.

5.4.3 A Validation of the BALI for use with Adults

5.4.3.1 Comparisons of the performance of the dyslexic and control groups on the BALI tasks.

We again evaluated the capacity of the BALI to discriminate between dyslexics and controls (using a larger sample) by comparing the performances of the groups on the combination of tasks included in the BALI as well as on the tasks individually. The descriptive statistics of the groups' performance on the BALI tasks along with Cohen's d are reported in Table 5.7. The results of the MANOVA with the BALI tasks as the dependent variables and group (dyslexics and controls) as the independent variable indicated a significant difference between the groups on the tasks combined $F(8, 66) = 9.43$, $p < .001$; Pillai's Trace = 0.53. Follow up ANOVA (with Bonferroni adjusted alpha level of .006) revealed that the groups did not differ on the Nonword Repetition $F(1, 73) = 6.22$, $p = .015$ and the Inhibition $F(1, 73) = 3.11$, $p = .082$ tasks. However, the groups differed significantly on all the other tasks with the dyslexic group performing less well than the control group. Additionally, Cohen's *d* effect size indicated that the differences in the groups' performances ranged from moderate, $d = 0.74$, on Phoneme Deletion Accuracy, to very large, $d = 1.77$, on Phoneme Deletion Speed, followed by $d = 1.54$ on the 2 Minute Spelling tasks.

Table 5.7 Mean Standard Scores (Standard Deviation in Parenthesis) of Dyslexic and Control Groups on BALI Tasks and Effect Size (N = 80)

Experimental Tasks	Dyslexics (n = 40)	Controls (n = 40)	Cohen's *d*
Phoneme Deletion Accuracy[a]	18.47 (5.19)	21.65 (3.46)	.74
Phoneme Deletion Speed	136.73 (45.31)	78.71 (20.12)	1.77
Nonword Repetition Accuracy[b]	5.48 (1.84)	6.63 (1.56)	.68
RAN Digits Time	19.15 (6.65)	14.76 (2.95)	.92
RAN Objects Time	26.68 (6.29)	21.33 (3.61)	1.08
1 Minute Reading Fluency[c]	94.40 (24.69)	119.08 (19.31)	1.12
2 Minute Spelling Efficiency[d]	20.75 (5.49)	27.65 (3.50)	1.54
Inhibition	11.21 (7.50)	8.46 (4.29)	0.47

Note. [a]maximum score = 24. [b]maximum score = 10. [c]maximum score = 150. [d]maximum score = 40.

The results of the MANOVA were similar to those of the pilot study and indicated that the tasks combined could effectively discriminate between adults with and without dyslexia and were therefore likely to be suitable for inclusion in a dyslexia screening test for adults. While the tasks combined discriminated between the groups, the Inhibition and Nonword Repetition tasks did not. For the Inhibition task, the results here are similar to those of the pilot study and thus suggest that this task (and possibly executive function tasks in general) is not a sensitive indicator of dyslexia in adulthood; this replication of a null effect also argues against the explanation of a lack of statistical power in the pilot study. For the Nonword Repetition task, the results here are contrary to those of the pilot study. The performances of the dyslexic groups in both studies were similar; however, the controls in the pilot study obtained higher mean scores $M = 7.67$ on the task than those in this study $M = 6.63$ with less variability $SD = 1.09$ and $SD = 1.56$ respectively. This suggests that the task may not be a robust discriminator of adults with and without dyslexia.

5.4.3.2 Discriminant analyses.

To further assess the capacity of the BALI (as a whole) to discriminate between adults with and without dyslexia, we conducted a logistic regression. Based on the results of the MANOVA analysis, we excluded the Inhibition and Nonword Repetition tasks from the BALI as the tasks did not discriminate between the groups. Additionally, the reliability of the Nonword Repetition task was below an acceptable level (see results of the reliability analysis) and was therefore not suitable for inclusion in the BALI. Therefore, six tasks were included as predictors in the logistic regression: Phoneme Deletion Accuracy, Phoneme Deletion Fluency, 1 Minute Reading, 2 Minute Spelling, RAN Digits and RAN Objects, with group membership (dyslexics or controls) as the dependent variable. Five participants were excluded from the analysis because of missing data; therefore, the sample size was reduced to 75, with 37 dyslexics and 38 controls. The results indicated that the model with all the predictors was statistically significant χ^2 $(6, N = 75) = 63.90$, $p < .001$, indicating that the predictors combined were able to distinguish between the dyslexic and control participants. The predictors combined explained a large amount of the variance in the groups .57 (Cox and Snell R^2) and .76 (Nagelkerke R^2). The Hosmer-Lemeshow Goodness of Fit Test indicated that the model was a good fit χ^2 $(7, N = 78) = 7.07$, $p = .422$ for the data. This suggests that the model prediction is not significantly different from the observed values. As detailed in Table 5.8, the BALI correctly classified 68 of the 75 participants included in the analysis, yielding an overall classification rate of 90.7%. It correctly classified 34 of the 37 dyslexics for a sensitivity rate of 91.9%, and 34 of the 38 controls for a specificity rate of 89.5%, a false positive rate of 10.5% and false negative rate of 8.1%. The sensitivity rate is well above the minimum recommended 80% and specificity rate is in line with recommendations 90% (Friberg, 2010; Glascoe & Byrne, 1993).

Table 5.8. Classification Result of Logistic Regression Analysis for Dyslexics and Controls

Participants	Predicted Group Membership		
	Dyslexics	Controls	% Correct
Dyslexics	34	3	91.9
Controls	4	34	89.5
Overall %			90.7

As detailed in Table 5.9, the Wald statistic indicated that four of the six BALI tasks were significant predictors of group membership namely: Phoneme Deletion Fluency χ^2 (1, *N* = 75) = 7.20, *p* = .007, RAN Digits χ^2 (1, *N* = 75) = 4.02, *p* = .045, RAN Objects χ^2 (1, *N* = 75) = 6.16, *p* = .013, and 2 Minute Spelling χ^2 (1, *N* = 75) = 4.56, *p* = .033. The strongest predictors were Phoneme Deletion Fluency and RAN Objects. As the analysis indicated that Phoneme Deletion Accuracy and 1 Minute Reading were not significant predictors we excluded these tasks from the list of predictors and re-ran the analysis. The model was statistically significant χ^2 (4, *N* = 75) = 62.34, *p* < .001, and explained a similar proportion of the variance in the groups .56 (Cox and Snell R^2) and .75(Nagelkerke R^2) as the model with all six tasks. The Hosmer-Lemeshow Goodness of Fit Test indicated that the model was a good fit χ^2 (7, *N* = 78) = 5.41, *p* = .610 for the data. However, the difference between the -2 log-likelihood for the model with the six predictors 40.06 and the model with the four predictors 41.62 was 1.56, which is not statistically significant. This indicates that the model with four predictors was not significantly better (or worse) than the one with the six predictors. This was supported by the classification of the participants. The model with only four predictors correctly classified 66 of the 75 participants for an overall classification rate of 88%. It correctly classified 33 of the 37 dyslexics for a sensitivity rate of 89.2%, and 33 of the 38 controls for a specificity rate of 86.8%, a false positive rate of 13.2% and false negative rate of 10.8%. Although the results indicated that Phoneme Deletion Accuracy and 1 Minute Reading were not significant unique predictors, deleting these from the model resulted in a reduction in the accuracy of the classification of participants, and they were therefore kept in the model, as they contributed to the overall accuracy of the model.

Table 5.9 Logistic Regression Analysis Predicting Participants' Group Membership from BALI Tasks (N = 75)

Predictors	β	S. E.	Wald χ^2	Odds Ratio	95% C. I. For Odds Ratio		*P* Value
					Lower	Upper	
Phoneme Deletion Accuracy	0.09	0.21	0.17	1.09	.72	1.65	.682
Phoneme Deletion Speed	0.07	0.03	7.20	1.07	1.02	1.13	.007
RAN Digits Time	-0.41	0.20	4.02	0.67	0.45	0.99	.045
RAN Objects Time	0.29	0.12	6.16	1.34	1.06	1.69	.013
1 Minute Reading Fluency	-0.04	0.03	1.33	0.96	0.90	1.03	.249
2 Minute Spelling Efficiency	-0.26	0.12	4.56	0.77	0.61	0.98	.033
Constant	1.35	8.96	0.02	3.87			

5.4.3.3 Correlational analyses.

The correlation analyses were conducted with the total sample (*N* = 80) of dyslexic and control participants. Similar to the Pilot Study, the performance differences between the groups on the tasks were primarily quantitative, and the distributions of scores seem to fall on a continuum with the dyslexic group at the lower end. Thus, the groups were combined to maximise the heterogeneity of the sample and to reduce restricted range effects. We examined the intercorrelations of the tasks in the BALI (excluding the Inhibition and Nonword Repetition tasks). Every task was significantly correlated with at least three of the other tasks with moderate to high correlations (see Table 5.10). Three of the tasks, Phoneme Deletion Fluency, 1 Minute Reading and 2 Minute Spelling were significantly correlated with all the other tasks. The 1 Minute Reading task was very highly correlated with RAN Digits $r(80) = -.85$, $p < .001$. The high correlation suggests that the tasks may be measuring the same construct; however, a more probable explanation is the nature of the relationship between the tasks. Research indicates that RAN tasks are highly correlated with reading ability and in particular reading fluency across different orthographies (Katzir et al., 2006; Norton & Wolf, 2012). For the Phoneme Deletion Accuracy task, compared to the pilot study, the correlations between it and the other tasks in the BALI generally improved. This was particularly true for its correlation with the 1 Minute Reading task ($r(80) = .30$, $p < .001$) which was not significant in the pilot study. In general, the correlations between the items in the BALI in this study were higher than those in the pilot because of the increase in the variability of the scores obtained by the participants in this study. The high inter-item correlations of the items suggests that a battery with the tasks combined is likely to be reliable.

5.4.3.4 Concurrent validity.

The concurrent validity of the tasks in the BALI was assessed by examining their correlations with the standardised measures of literacy. The 2 Minute Spelling task was highly correlated with all the standardised measures of literacy WRAT IV Word Reading $r(80) = .70$, $p < .001$, WRAT IV Spelling $r(80) = .76$, $p < .001$, and WRAT IV Comprehension $r(78) = .28$, $p = .014$. Additionally, its highest correlation was with the standardised measure of the same construct (WRAT IV Spelling) and was above the minimum level (.75) required to support concurrent validity (Kline, 2000). The 1 Minute Reading task, was highly correlated with WRAT IV Spelling $r(80) = .60$, $p < .001$, moderately with WRAT IV Word Reading $r(80) = .35$, $p = .001$, but not with WRAT IV Comprehension $r(78) = .10$, $p = .377$. The moderate correlation of the 1 Minute Reading task and the Word Reading subtest, which is not timed and is designed to assess word reading accuracy, is lower than expected, particularly as the task was designed to measure a similar construct and in view of the high correlation of the tasks ($r = .63$) in the pilot study. Nevertheless, its correlations with both the reading and spelling standardized measures suggest that it is likely measuring the construct it was designed to measure, reading fluency. Additionally, its high correlation with the RAN Digits task provides further support that the task is more likely measuring reading *fluency*. The other tasks in the BALI were also significantly correlated with at least one of the standardized literacy measures with low to high correlations (see Table 5.10). All the other tasks were designed to measure literacy and related sub-skills and therefore the results of the correlation analyses provide support for the concurrent and construct validity of these tasks.

Table 5.10 Correlations between BALI Tasks and Standardised Measures (N = 80)

Measures	1	2	3	4	5	6	7	8	9	10	11
1. Phoneme Deletion Accuracy	-	-.65**	-.19	-.14	.30**	.46**	.01	.18	.60**	.46**	.30**
2. Phoneme Deletion Fluency		-	.44**	.37**	-.56**	-.68**	.23	-.06	-.60**	-.66**	-.18
3. RAN Digits Time			-	.79**	-.85**	-.52**	-.05	-.02	-.21	-.48**	-.02
4. RAN Objects Time				-	-.68**	-.43**	-.12	-.09	-.10	-.39**	-.04
5. 1 Minute Reading					-	.61**	-.07	.16	.35**	.60**	.10
6. 2 Minute Spelling						-	-.10	-.25*	.70**	.76**	.28*
7. WRIT Matrices							-	.38**	.05	-.02	.20
8. WRIT Vocabulary								-	.45*	.41*	.68**
9. WRAT IV Word Reading									-	.79**	.57**
10. WRAT IV Spelling										-	.37**
11. WRAT IV Sentence Comprehension											-

* *p* < .05. ** *p* < .01

5.5 Discussion and Conclusion

The main purpose of this study was to further investigate, with a larger more representative sample of participants, the capacity of the tasks in the BALI to discriminate between adults with and without dyslexia. Also, we assessed the psychometric properties (validity and reliability) of the tasks and the effects of the changes made to three of the tasks based on the results of the pilot study. The capacity of the tasks to discriminate between adults with and without dyslexia (individually and collectively) was examined by comparing group performances. Correlation and reliability analyses were used to examine the validity and reliability of the tasks. The groups were comparable on the standardised measures of verbal and non-verbal IQ and reading comprehension, but differed significantly on the measures of word reading and spelling accuracy, with very large effect sizes $d = 1.19$ and $d = 1.85$ respectively. Although, the dyslexics obtained significantly lower scores than the controls on the reading and spelling measures, their performances were in the average range, indicating that they may be considered compensated or high functioning dyslexics. These results are similar to those of the pilot study as well as the findings of other studies that compare the performance of university students with and without dyslexia on measures of literacy (Hatcher et al., 2002; Kemp et al., 2009; Lefly & Pennington 1991).

5.5.1 Reliability

The reliability analyses provided evidence of the reliability of three of tasks included in the BALI, namely Phoneme Deletion, 1 Minute Reading and 2 Minute Spelling. The Phoneme Deletion task (as a whole) had excellent internal consistency coefficient $\alpha = .92$, and test-retest reliability accuracy $r = .89$, and fluency $r = .95$. Additionally, the reliability of both blocks of the tasks was similar to that of the task as a whole, coefficient $\alpha = .88$ and $\alpha = .86$, and test-retest reliability ranging from $r = .81$ to $r = .93$. The reliability of the task compares favourably with, and often exceeds, that of similar tasks in other adult dyslexia screening tests. For the DAST, test-retest reliability of its Phonemic Segmentation task reported is $r = .90$ (Nicolson & Fawcett, 1998), while the internal consistency of the Spoonerism task of YAA-R reported is coefficient $\alpha = .76$ (Warmington et al., 2013). In addition, as indicated in the Pilot Study, the task was specifically designed to reliably measure the speed of phonological awareness. This is a unique feature which sets it apart from similar tasks and makes it particularly appropriate for individuals who perform at ceiling on accuracy, where speed is the only discriminator. Similarly, the test-retest reliability of the BALI's 1 Minute Reading *r* =.91 and 2 Minute Spelling *r* =.89 tasks are comparable to that of the DAST *r* =.90 and *r* =.93 respectively. However, for the DAST, the interval between the first and second administrations of the test was only one week which may have facilitated recall and influenced the same pattern of responses and consequently higher correlations (Anastasi & Urbina, 1997). The 1 Minute Reading task was specifically designed as a measure of fluency and not efficiency, and the error rates on the task even for the dyslexic group were very low. This distinguishes the tasks from similar measures included in other test batteries. The validity of the tasks as a fluency measure is supported by its high correlation with the RAN Digits and relatively weaker correlation with the Word Reading subtest of the WRAT IV. So, even for dyslexic participants, the BALI provides a fairly pure estimate of word recognition fluency (automaticity). The results of both the pilot and the main study indicated that the reliability of the Nonword Repetition task was below an acceptable level with low inter-correlations among items on the task, which we were not able to improve even though major adjustments were undertaken after the results of the pilot. We cannot fully explain this given that the method used to create the words was the same; however it is noteworthy that nonword repetition tasks seem to lack robustness across studies (Gray, 2003; Parrila, Georgiou, & Corkett, 2007; Ramus et al., 2003b; Snowling et al., 1997) The low reliability of the BALI version(s) of the task indicates that it is also not a valid measure for inclusion (Kline, 2000). Although the reliability of the BALI as a whole was not assessed, the moderate to high inter-correlations of the tasks suggest that the measure should have good internal reliability.

5.5.2 Validity

The performance of the dyslexic and control groups on the tasks in the BALI provided evidence of their capacity to discriminate between adults with and without dyslexia as the groups differed significantly on the combination of tasks included in the BALI. Additionally, with the exception of the Inhibition and Nonword Repetition tasks, the groups also differed significantly on each task individually. The dyslexic group performed significantly less well than controls on the tasks, obtaining lower accuracy or taking more time to complete the tasks than the controls. Furthermore the effect size calculations revealed that these differences were not only statistically significant, but also large, ranging from d = 0.74 to d = 1.77. The tasks included in the BALI assess literacy and the literacy related skills (phonological processing) which research has consistently shown are deficient in adults with dyslexia, even those described as compensated (Bruck, 1992; Kemp et al., 2009; Ramus et al., 2003b; Snowling et al., 1997; Swanson & Hsieh, 2009). Our results are in line with these studies and illustrate that adults with dyslexia who have cognitive and literacy abilities in the average range still exhibit relative deficits in their literacy and phonological processing skills. The present study provides empirical evidence that the tasks which assess the behavioural deficits that are known to persist in adults with dyslexia can be combined into a quick and effective screening test (the BALI) to identify individuals at risk of dyslexia.

The Inhibition and Nonword Repetition tasks created for inclusion in the BALI did not effectively discriminate between the groups. For the Inhibition task, in this study as well as the pilot, the dyslexics tended to be slower (were more inhibited) than the controls; however the difference between the groups was not statistically significant. What is more, the inhibition effect did not differ between the groups. Only a few studies have investigated executive function deficits among adults with dyslexia, and the types of deficit investigated varied and included planning, organization, set shifting, problem solving, sequencing, impulsiveness, and inhibition (Baker & Ireland, 2007; Brosnan et al., 2002; Lindgrén & Laine, 2011; Weyandt, Rice, Linterman, Mitzlaff, & Emert, 1998). Of these studies, two have compared the performance of adults with and without dyslexia on tasks assessing inhibition and these produced different results (Brosnan et al., 2002; Lindgrén & Laine, 2011). Brosnan et al. (2002) found evidence of an inhibition deficit in children and adults with dyslexia. They examined the performance of children and adults with and without dyslexia on a range of executive function tasks including sequencing, organization, planning and inhibition. They found that the controls performed better than the dyslexics on a group-embedded figures test (GEFT) assessing inhibition. In addition to this study others have found inhibition deficit in children with dyslexia (Helland & Asbjørnsen, 2000; Protopapas, Archonti, & Skaloumbakas, 2007; Reiter, Tucha, & Lange, 2005). One possible explanation for the difference in the results of our studies, and that of Brosnan et al. (2002), is that our tasks may lack the sensitivity required to detect any inhibition deficit that may exist in adults with dyslexia. Our task is based on the same paradigm as the *Same World Opposite World* subtest of the Test of Everyday Attention for Children (Manly et al., 2001). The task required the participants to name as quickly as possible the same or the opposite direction of arrows in a long quasi-random sequence of arrows. The GEFT task, which requires the identification of simple figures hidden within more complex figures, may be cognitively more demanding and therefore more likely to detect any inhibition deficit that may exist in adults with dyslexia. Contrary to the results of the aforementioned studies, and similar to the results of our study, Lindgrén and Laine (2011) found that the performance of multilingual (including Swedish and Finnish) university students with and without dyslexia did not differ on the Stroop colour test, an established inhibition measure. The Stroop colour test is very similar to the task used in our study and the results of this study provide some support for our suggestion that this type of task may not be sufficiently sensitive to detect any inhibition deficit that may exist in adults with dyslexia. However, given the limited number of studies conducted in this area, more research is required in order to confirm the existence (or non existence) of an inhibition deficit in adults with dyslexia.

For the Nonword Repetition task, the results of the pilot study and this study differed. However, for both studies, the dyslexic group performed less well than the controls (obtaining lower scores) on the task. In the pilot study, the groups' performance differed significantly and the task effectively discriminated between the dyslexics and controls, collectively with the other tasks included in the BALI as well as by itself. However, in this larger study, the difference between the groups on the task did not reach significance. The results of similar studies that compare the performance of adults with and without dyslexia on nonword repetition tasks vary, with some finding that the performance of the groups differs (Parrila et al., 2007; Pennington, Van Orden, Smith, Green, & Haith, 1990; Ramus et al., 2003b) and others finding no differences (Snowling et al., 1997; Paulesu, et al., 1996). The lack of difference in the performance of the groups in our study may be because the dyslexics have developed strategies that effectively help them to perform the task and not exhibit their underlying deficit. Alternatively, our task may not be sufficiently sensitive (although very similar tasks were used in other studies that found group differences) to detect any deficit that may exist in adults with dyslexia. Another possibility is that the scoring procedure (whole word accuracy) was not appropriate for the population. A more fine-grained scoring system, in which partial accuracy would be considered, could be used in order to see if it is the scoring system used here that was inappropriate. However, the reliability analysis indicated that the internal consistency of the task is below an acceptable level, which also questions its validity.

The validity of the BALI was further assessed by correlational analysis. The analysis indicated that the tasks were correlated to at least one of the standardised measures of literacy with low (r = .25) to high (r = .76). The concurrent validity of the 2 Minute Spelling Efficiency task was strongly supported by its high correlations with the standardised measure of spelling r = .76 and reading r = .70. Its correlation with the standardised measure of spelling is just above the minimum required to support concurrent validity (Kline, 2000), indicating that it is a valid measure of spelling. There was also good support for the validity of the 1 Minute Reading task as a measure of fluency as it correlated with both the standardised reading r = .35 and spelling r = .60 measures and also RAN Digits r = -.85. Most of the adult dyslexia screening tests currently in use in the UK do not report the correlations of their subtests with standardised measures of the same or similar constructs. However, our results compare favourably with those of the York Adult Assessment-Revised (YAA-R). For the YAA-R, the spelling (error rate) subtest correlation with the WRAT spelling is r = -.52, and reading accuracy and reading time subtests correlations with the WRAT reading are r = .55 and r = -.40 respectively (Warmington et al., 2013).

The results of the logistic regression provided the most compelling evidence of the validity of the BALI as a screening test for adults with dyslexia. The BALI (with the Inhibition and Nonword Repetition tasks excluded) correctly classified 91.9% (sensitivity rate) of the dyslexics and 89.5% (specificity rate) of the controls for an overall classification rate of 90.7%. Its sensitivity rate is well above the minimum (80%) considered acceptable and its specificity rate is within the minimum (90%) (Friberg, 2010; Glascoe & Byrne, 1993). More importantly these rates compare favourably with the rates reported for other adult dyslexia screening tests such as the DAST sensitivity 94% and specificity 100% (Nicolson & Fawcett, 1998), LADSPLUS sensitivity 90.6% and specificity 90% (Singleton, Horne, & Simmons, 2009) and YAA-R sensitivity 80% and specificity 97.2 % (Warmington et al., 2013).

The results of this study and that of the pilot indicated that we achieved our objective of creating a quick and effective screening test that is capable of identifying adults with dyslexia even those who have compensated for their difficulties. Given the favourable results for the BALI thus far, further research is required to establish appropriate norms for the tasks. Also, the tasks should be further evaluated with a non-student population to ensure that they have a broader application among adult communities. Further research into the development and evaluation of the Nonword Repetition and Inhibition tasks is needed to determine whether these skills have relevance as behavioural markers of adult dyslexia. In conclusion, the results of this study provided further evidence of the validity and reliability of the tasks in the BALI and its capacity to effectively identify adults at risk of dyslexia. It also provides evidence that dyslexia in adults may be effectively identified using behavioural measures of literacy and phonological processing only, combined into a quick screening test.

Chapter 6: Evaluation of the Instines, a Screening Test for Dyslexia in Adults (Study 3)

6.1 Introduction

The Instines is one of a few computerized dyslexia screening tests currently available in the United Kingdom. A description of the test along with information on its development were provided in the review of dyslexia screening tests in Chapter 2. As highlighted in the review, there are no published independent empirical studies of the Instines and information regarding the research undertaken for its development is extremely limited. Additionally, James (2004) was critical of the development of the test (without providing details) and suggested that it lacked scientific rigour. Furthermore the criteria used to determine an individual's risk of dyslexia for the test overall, and the subtests are not known. Nevertheless, a promotional flyer for the test states that it is used by approximately 800 schools in the United Kingdom as well as other organizations including: Remploy, Job Centre Plus, Connexions, and the National Probation Service. There is therefore a need for research to be undertaken on the Instines to determine its validity as a dyslexia screening test.

6.2 Method

6.2.1 Participants

Participants were the same as those described in the BALI main study 2b Chapter 5, and consisted of 80 native English speaking students at Bangor University. Two of the control participants were not assessed with the Instines and therefore their data was excluded from this study. The total number of participants was therefore 78: 40 dyslexics and 38 controls. The dyslexic group comprised 23 (57.5%) females and 17 (42.5%) males, mean age M = 21.73 (SD = 5.66), and range 18 – 50 years. The control group comprised 32 (84.2%) females and 6 (15.8%) males, mean age M = 19.29 (SD = 2.0), and range 18 – 30 years. The age difference between the groups was significant t (49.12) = -2.56, p = .014.

6.2.2 Measures

The measures administered to the participants included: Matrices and Vocabulary subtests of the WRIT to assess verbal and nonverbal ability; Word Reading, Spelling, and Sentence Comprehension subtests of the WRAT IV to assess literacy skills; the BALI tasks; and the Instines. With the exception of the Instines, all the other measures were described in the BALI pilot study 2a and main study 2b in Chapters 4 and 5, therefore only the Instines is described here.

Instines. The Instines consists of eight subtests: (1) Spelling/Homophone, (2) Spatial Recognition, (3) Verbal Reasoning, (4) Reading Speed, (5) Directional Awareness, (6) Forward Digit Span, (7) Reverse Digit Span, and (8) Comprehension. It is self-administered and participants follow the verbal and written instructions given during the administration of the test. A percentile score is calculated for each subtest of the test and there is no overall score. The published literature on the Instines is very limited and we were not able to find literature with the psychometric details of the subtests included in the Instines. Also, there is no published manual for the test and the details of the subtests given below are based on training undertaken for the administration of the test and observation of participants during administration.

Spelling/Homophone Subtest. This task assessed orthographic and spelling recognition abilities. Participants were required to select the correct spelling from among four options, of which three were misspelt words (non-word homophones of the word); it consisted of 30 words. Instructions for the task were given verbally, via computer, and the words for which the participants had to select the correct spelling were read by the software. The task is timed; however, the amount of time given to complete the task, and how it is scored, is not stated.

Spatial Recognition Subtest. This task assessed visual processing abilities. Participants were required to construct visually presented patterns. A pattern divided into nine square tiles was presented in the upper right section of the computer screen. Below the pattern were six square tiles made up of tiles from the pattern to be constructed, as well as tiles that were not a part of the pattern. Participants were required to select the correct tiles, and complete the pattern, exactly as presented in the square space provided in the upper left section of the screen. Sometimes it was necessary to rotate a tile (by clicking on a button) in order to complete the pattern correctly. The software demonstrated how the task was to be completed. After completing the pattern, participants were required (through verbal instructions before starting the task and visual instructions at the completion of each pattern) to click on a button in the centre of the screen. If the pattern was not completed correctly, the software verbally indicated that the pattern was incorrect, and participants were required to complete the pattern correctly before proceeding to the next pattern. Instructions for the task were given verbally. The task was timed and terminated after the allotted time; however, the amount of time specified for the task is not stated. The number of items that participants completed varied, depending on how quickly they completed each pattern.

Verbal Reasoning Subtest. This task assessed the conceptual understanding of words. Participants were required to select a word from five words presented visually on screen to complete sentences of verbal analogies. The task consisted of 15 items. An example of the type of analogies included is: "Cloud is to sky as wave is to" options for selection: (1) "*lake*", (2) "*farewell*", (3) "*ship*", (4) "*sea*", (5) "*land*". Once the participant selected a word to complete the sentence, the sentence was read by the software. Instructions for the task were presented verbally and visually.

Forward Digit Span Subtest. This task assessed short term memory. Participants were required to produce (by typing) a series of digits presented in the order of presentation. It consisted of two parts, in the first part, the instructions were verbally presented and the items were displayed visually and also read by the software. In the second part of the task, the instructions as well as the items were presented visually. For both parts of the task, a series of digits, presented on the computer screen, disappeared after a few seconds, and participants were required to type the digits in the order presented (after the digits disappeared) once a green light appeared on screen. The first item consisted of two digits and the number of digits presented increased by one digit as the task progressed. The total number of items completed depended on the participant's performance on the items presented, such that participants with better performances completed more items. However, there is no specific information available on the discontinuation criteria for the task.

Reverse Digit Span Subtest. This task assessed working memory. It is similar to the Forward Digit Span subtest; however, participants were required to produce by typing a series of digits presented in the reverse order of presentation. The administration and presentation of the items of the task were the same as the Forward Digit Span task, the only difference being that participants were required to type the digits presented in the reverse order of presentation.

Directional Awareness. This task assessed visual-spatial orientation. It consisted of two parts, and instructions were presented verbally and visually. In the first part, participants were required to select the right and the left hands of an image of a man on the screen. For example, "Please click on my left hand with your right mouse button". In the second part, participants were required to navigate as far "north" as possible in a three dimensional maze using the arrow keys of the computer for two minutes.

Reading Speed and Comprehension Subtests. These tasks assessed reading speed and reading comprehension respectively. The two subtests were administered as one task. Participants were required to read a short passage and answer questions based on it. Participants indicated when they finished reading the passage by selecting "OK" on the computer screen. As soon as participants finished reading the passage, they answered nine questions. The questions were presented visually with five options from which the participants selected the correct response. Instructions for the task were given verbally.

6.2.3 Procedure

The procedure for the testing protocol has already been described in Chapter 5, Study 2b. We repeat it here for ease of reference for the interested reader. The study was completed in two parts, with the second part scheduled 14 days after the first. The average number of days between the completion of the first (test) and the second (retest) parts of the study was M = 15 with a range of 14 – 21 days. For both parts of the study participants were assessed individually, with part one lasting approximately 80 – 90 minutes and part two approximately 20 – 30 minutes. Tasks were administered in a fixed order, in part one, as follows: Reading, Sentence Completion, and Spelling subtests of the WRAT IV, Matrices and Vocabulary subtests of the WRIT, and the experimental tasks of the BALI in the following order: RAN Objects, Phoneme Deletion, 1 Minute Reading Fluency, RAN Digits, 2 Minute Spelling Efficiency, Arrows Inhibition, Non-word Repetition, and the Instines was administered last. Part two of the study was used to retest participants on four of the tasks performed in part one of the study in order to examine the test-retest reliability of the tasks. The tasks administered in part two of the study, in the order of administration were: (1) Phoneme Deletion, (2) One Minute Reading Fluency, (3) Two Minute Spelling Efficiency, from the BALI, and (4) the Instines. The standardised measures were administered according to the published guidelines and the experimental task in accordance with established instructions. The Instines was administered on a Toshiba Tecra M 11 laptop with a mouse and a keyboard attached. Administration time was approximately 30 minutes and all participants were monitored by the researcher during the administration. Ethical approval for the study was granted by the School of Psychology, Bangor University.

6.3 Data Analysis

Prior to analyses, the data set was checked for outliers, and outliers above or below 2.5 standard deviations from the mean were adjusted to 2.5 standard deviations. To identify outliers the mean for each group was calculated separately. For the background measures, there were no outliers for the control group and only one score was adjusted on the WRIT Vocabulary subtest for the dyslexic group. We used multivariate analysis of variance (MANOVA), followed up with univariate analyses (ANOVA) with Bonferroni adjusted alpha level to compare the groups on the background measures. The data set was checked for appropriateness for multivariate analysis (MANOVA) and the assumption of equality of error variances was violated by the WRAT Spelling subtest therefore a more conservative alpha of .01 was used for significance as recommended by Tabachnick and Fidell (2007). We used Mann Whitney *U* to compare the performance of the groups on the subtests of the Instines (as the data was not normally distributed) and to assess its capacity to discriminate between adults with and without dyslexia. Effect sizes, using Cohen's *d* formula, were calculated to determine the magnitude of any group difference. Logistic regression analysis was also used to further assess the capacity of the Instines to discriminate between adults with and without dyslexia. Correlation analysis (Spearman's r_s) was used to assess the concurrent validity and test-retest reliability of the subtests of the Instines. In the results section, we first report the results of the group comparison on the cognitive and literacy measures. This is followed by the validity assessment of the Instines, (a) comparing the performances of the groups on its subtests, (b) examining their capacity to predict group membership, and (c) assessing the correlation of selected subtests with the subtests of the standardized measures. Finally, we report the results of the assessment of the test-retest reliability of the Instines.

6.4 Results

6.4.1 Comparisons of the Performance of the Dyslexic and Control Groups on the Background Measures

These analyses replicated those of the previous (Chapter 5 BALI main study 2b). However, as the number of participants in the control group was reduced from 40 to 38, the data were re-analysed; the results were unchanged and therefore are only summarised here. Details of the groups' performances on the measures are reported in Table 6.1. The groups did not differ on the cognitive measures (WRIT Matrices and Vocabulary) and on the Sentence Comprehension subtest of the WRAT IV. However, the groups differed significantly on the WRAT IV Word Reading and Spelling subtests with the dyslexics performing less well than the controls. Cohen's *d* effect size indicated that the differences in the groups; performance on these measures were also very large. Despite the reduction in the number of participants in the control group, these results are similar to those reported in the BALI main study and suggest that the groups were similar in terms of their verbal and nonverbal abilities and only differed on their literacy skills (reading and spelling).

Table 6.1 Mean Standard Scores (Standard Deviation in Parentheses) of Dyslexic and Control Groups on Background Measures and Effect Sizes (N =78)

Measures	Dyslexics (*n* = 40)	Controls (*n* = 38)	*p* Value	Cohen's *d* Effect Size
WRIT Matrices	106.50 (13.15)	104.71 (10.97)	.517	0.15
WRIT Vocabulary	102.32 (11.15)	104.39 (11.71)	.426	0.18
WRAT IV Word Reading	93.88 (7.88)	103.08 (7.67)	.001	1.18
WRAT IV Spelling	92.32 (8.46)	112.00 (13.09)	.001	1.83
WRAT IV Sentence Comprehension	105.82 (12.54)	109.82 (9.93)	.124	0.36

6.4.2 Evaluation of the Validity of the Instines

4.4.2.1. Comparisons of performance of dyslexic and control groups on the subtests of the Instines. The validity of the Instines was assessed by examining its capacity to discriminate between adults with and without dyslexia. This was done firstly by comparing the performances of the dyslexic and control groups on the subtests. If the Instines is a valid measure for the detection of dyslexic traits in adults, then we would expect to find differences in the performances of adults with and without dyslexia on its subtests. The performances of the groups were compared with Mann Whitney *U* test, as only percentile scores for each subtest are produced by the Instines. Cohen's *d* effect size was calculated to determine the magnitude of any differences. The results revealed that the average percentile scores for both groups were generally within or above the 50^{th} percentile, indicating that the participants performed in the average range or above on the subtests (see Table 6.2 for details). The performances of the groups differed on four of the eight subtests of the Instines namely, in order of effect size magnitude: Spelling/Homophone (d = 1.40), Forward Digit Span (d = 0.83), Reading Speed (d = 0.71), and Directional Awareness (d = 0.49). With the exception of the Directional Awareness subtest, the dyslexics obtained lower scores on the subtests.

For the other four subtests, Spatial Awareness and Recognition, Verbal Reasoning, Reversed Digit Span, and Comprehension, the performances of the groups did not differ. This suggests that these subtests may not be sufficiently sensitive to detect differences between individuals or groups with and without dyslexia on the abilities being assessed. However, for the Verbal Reasoning and Comprehension subtests, it is likely that the groups were in fact similar in their performances on these subtests, as this was also true of their performances on standardised measure of similar cognitive abilities (nonverbal and verbal) and comprehension.

Table 6.2 Results of Mann-Whitney U Test of Difference in Mean Ranks of Performance of Dyslexic and Control Groups on the Subtests of the Instines with Effect Sizes, Mean Percentile Scores and Standard Deviation in Parentheses (N = 78).

	Dyslexic Group (n = 40)		Control Group (n = 38)				
Instines Subtests	Mean Rank	Mean Percentile (SD)	Mean Rank	Mean Percentile (SD)	Mann-Whitney U	Z	Cohen's d Effect Size
Spelling/Homophone	26.28	49.60 (18.06)	53.42	73.08 (15.59)	231.00	-5.31***	1.40
Spatial Awareness and Recognition	38.91	79.80 (31.28)	40.12	80.66 (32.68)	736.50	-.28	0.03
Verbal Reasoning	38.95	66.62 (31.01)	40.08	73.92 (17.92)	738.00	-.22	0.30
Reading Speed	33.24	63.85 (21.73)	46.09	76.89 (15.19)	509.50	-2.51**	0.71
Directional Awareness	44.46	58.40 (38.96)	34.28	39.37 (38.32)	561.50	-2.01*	0.49
Forward Digit Span	31.46	61.40 (19.73)	47.96	76.37 (16.41)	438.50	-3.28**	0.83
Reverse Digit Span	35.04	43.23 (26.45)	44.20	56.08 (23.50)	581.50	-1.84	0.52
Comprehension	36.25	74.40 (32.09)	42.92	84.26 (25.27)	630.00	-1.51	0.34

* $p < .05$. ** $p < .01$. *** $p < .001$.

4.4.4.2 Discriminant analyses. To further assess the capacity of the Instines (as a whole) to discriminate between adults with and without dyslexia, we conducted a logistic regression. For the logistic regression, the eight subtests of the Instines were the predictors (independent variables) and group membership (dyslexics or controls) the dependent variable. The results indicated that the model with all the predictors entered was statistically significant χ^2 (8, N = 78) = 42.53, p < .001, indicating that the predictors combined were able to distinguish between the dyslexic and control participants. The predictors combined explained a large amount of the variance in the groups .42 (Cox and Snell R^2) and .56 (Nagelkerke R^2). However, the Hosmer-Lemeshow Goodness of Fit Test indicated that the model was not a good fit χ^2 (8, N = 78) = 16.88, p = .031 for the data. This suggests that the model prediction was significantly different from the observed values. As detailed in Table 6.3, the Instines correctly classified 65 of the 78 participants included in the analysis for an overall classification rate of 83.3%. It correctly classified 32 of the 40 dyslexics for a sensitivity rate of 80% and 33 of the 38 controls for a specificity rate of 86.8%, a false positive rate of 13.16% and false negative rate of 20%. The sensitivity rate was at the minimum required, however, the specificity rate was below the minimum considered acceptable, of 90% respectively (Friberg, 2010; Glascoe & Byrne, 1993).

Table 6.3 Classification Results of Logistic Regression Analysis for Dyslexics and Controls Groups

Participants	Predicted Group Membership		
	Dyslexics	Controls	% Correct
Dyslexics	32	8	80
Controls	5	33	86.8
Overall %			83.3

Table 6.4 presents the results of the logistic regression. The Wald statistic indicated that only the Spelling/Homophone subtest was a significant predictor of group membership χ^2 (1, N = 78) = 8.53, p = .003. This supports the result of the Hosmer-Lemeshow Goodness of Fit Test which indicated the model with all the predictors is not a good fit for the data. Based on this, we excluded the other subtests from the list of predictors and re-ran the analysis. The model was statistically significant χ^2 (1, N = 78) = 33.19, p < .001 indicating that the Spelling/ Homophone subtest distinguished between the groups. When compared to the model with all the subtests included, it explained a reduced proportion of the variance in the groups .35 (Cox and Snell R^2) and .46 (Nagelkerke R^2). The Hosmer-Lemeshow Goodness of Fit Test indicated that the model was a good fit χ^2 (8, N = 78) = 9.17, p = .328 for the data. The difference between the model with all the subtests and the model with only the Spelling/Homophone subtest was compared by calculating the difference in -2 log likelihood using the formula χ^2 = 2 [(-74.89)-(-65.55)] = -18.68 which is statistically significant at 7 degrees of freedom indicating that the model with only the Spelling/Homophone subtest is better than the model with all the subtests included. However, the accuracy of the classification of participants with the Spelling/Homophone only model was much less than that of the model with all the subtests. This model classified only 58 of the 78 participants for an overall classification rate of 74.4%. It correctly classified 27 of the 40 dyslexics for a sensitivity rate of 67.5%, and 31 of the 38 controls for a specificity rate of 81.6%, a false positive rate of 18.4% and false negative rate of 32.5%. These sensitivity and specificity rates are also below the minimum acceptable thresholds (Friberg, 2010; Glascoe & Byrne, 1993).

Table 6.4 Logistic Regression Predicting Participants' Group Membership from the Subtests of the Instines (N = 78)

Predictors	B	*S E*	Wald χ^2	Odds Ratio	95% C. I. For Odds Ratio		*P* Value
					Lower	Upper	
Spelling/Homophone	-0.08	0.03	8.53	0.92	0.87	0.97	.003
Spatial Recognition	0.00	0.01	0.01	1.00	0.98	1.02	.939
Verbal Reasoning	0.00	0.02	0.00	1.00	0.97	1.03	.981
Reading Speed	-0.03	0.02	2.99	0.97	0.93	1.00	.084
Directional Awareness	0.02	0.01	3.00	1.02	1.00	1.03	.084
Forward Digit Span	-0.02	0.02	0.96	0.98	0.94	1.02	.326
Reverse Digit Span	-0.01	0.01	0.37	0.99	0.97	1.02	.542
Comprehension	-0.02	0.01	1.72	0.98	0.96	1.01	.190
Constant	9.76	2.73	12.81				

In order to improve the model, we conducted another logistic regression with Spelling/Homophone and four of the other subtests of the Instines: Reading Speed, Directional Awareness, Forward Digit Span, and Reversed Digit Span. The results of the previous logistic regressions indicated that these subtests would provide the largest reduction in the -2 log likelihood of the models, and therefore should increase the amount of variance explained by the model. The model with these five subtests was statistically significant χ^2 (5, N = 78) = 40.60, p < .001indicating that the five subtests combined distinguished between the dyslexic and control participants. It explained a greater proportion of the variance in the groups .41 (Cox and Snell R^2) and .54 (Nagelkerke R^2) than the model with the Spelling/Homophone subtest only. The Hosmer-Lemeshow Goodness of Fit Test indicated that the model was a good fit χ^2 (8, N = 78) = 9.25, p = .322 for the data.

The difference between the model with the five subtests and the model with only the Spelling/Homophone subtest was compared using the formula χ^2 = 2 [(-67.48)-(-74.89)] =

14.82 which is statistically significant at 4 degrees of freedom indicating that the model with five subtests is better than the model with the Spelling/Homophone subtest. Additionally, for this new model (with Spelling/Homophone, Reading Speed, Directional Awareness, Forward Digit Span and Reversed Digit Span) the overall classification of participants and the sensitivity rate were improved. It classified 65 of the 78 participants for an overall classification rate of 83.3%. It correctly classified 33 of the 40 dyslexics for a sensitivity rate of 82.5%, and 32 of the 38 controls for a specificity rate of 84.2%, a false positive rate of 15.8% and false negative rate of 17.5%. The difference between the model with the five subtests and the model with all the subtests $\chi^2 = 2\ [(-67.48)-(-65.55)] = -3.86$ was not statistically significant at 3 degrees of freedom indicating that the model with all the subtests is no better than the model with the five subtests. However, the model with the five subtests was a good fit for the data while the model with all the subtests was not; therefore the model with the five subtests would be considered the best of the three models assessed. This suggests that three of the subtests (Spatial Awareness and Recognition, Verbal Reasoning, and Comprehension) were not contributing to the discriminatory capacity of the test. In the model with the five subtests; (Spelling/Homophone, Reading Speed, Directional Awareness, Forward Digit Span and Reversed Digit Span), the Spelling/Homophone subtest was again the only significant predictor χ^2 (1, N = 78) = 10.12, p = .001. This suggests that although Reading Speed, Directional Awareness, Forward Digit Span, and Reversed Digit Span were not unique significant predictors they were nevertheless contributing to the overall accuracy of the model.

The results indicated that the capacity of the Instines to discriminate between adults with and without dyslexia is lower than what is required for a good screening test. Its specificity rate would result in an unacceptably large number of individuals without dyslexia being classified as dyslexics (false positives). The results also provide further evidence that at least three of the subtests of the Instines may not be capable of distinguishing between adults with and without dyslexia and may not be contributing to the discriminatory capacity of the Instines as a whole.

We also used the classification of participants by the Instines to assess its capacity to discriminate between the groups, by comparing its classification with their actual group membership for both administrations of the test. For the first administration of the test, of the

40 dyslexic participants, Instines correctly classified 22 (sensitivity rate 55%), and of the 38 control participants, it correctly classified 22 (specificity rate 57.9%). For the second administration of the test, four control participants did not take part in the second part of the study, and the test was not administered to two participants (one control and one dyslexic) due to computer failure; therefore only 72 participants were assessed twice on the Instines. For this administration, Instines correctly classified 24 of the 39 dyslexic participants (sensitivity rate 61.5%), and 22 of the 33 control participants (specificity rate 66.7%). This indicates that classification of participants by the Instines would result in large and unacceptable percentages of false negatives and false positives. It also supports the results of the Hosmer-Lemeshow Goodness of Fit Test which indicated that the model was not a good fit to the data. These results suggest that the capacity of the Instines to discriminate between adults with and without dyslexia is below an acceptable level.

4.4.4.3 Evaluation of the concurrent validity of the Instines subtests. Finally, to further evaluate the validity of the Instines we assessed the concurrent validity of three of its subtests: (1) Spelling/Homophone, (2) Verbal Reasoning, and (3) Comprehension. The subtests of the Instines were designed to assess a number of different abilities; for example, the Spelling/Homophone subtest was designed to assess orthographic and spelling abilities. If the subtests are valid measures of the abilities they were designed to assess, then scores on them should correlate significantly with scores on other measures of the same and similar abilities. Participants' scores on the subtests were correlated with their scores on the subtests of the standardised cognitive and literacy measures that assess the same or similar abilities using Spearman's r_s.

As indicated in Table 6.5, there were significant positive correlations between participants' scores on the Instines subtests and their scores on the standardised measures. The Spelling/Homophone subtest was positively correlated with WRIT vocabulary $r_s = .26$, $p = .021$, WRAT Sentence Comprehension $r_s = .39$, $p < .001$, and most notably with WRAT Spelling $r_s = .75$, $p < .001$. The latter correlation is in line with the minimum required for the subtest to demonstrate concurrent validity (Kline, 2000). This suggests that the Spelling/Homophone subtest is a valid measure of spelling abilities. The subtest was also positively correlated with the standardised measures of similar constructs vocabulary and comprehension. The Verbal Reasoning subtest of the Instines was moderately positively correlated with the WRIT Matrices, r_s = .32, p = .005 and Vocabulary, r_s = .38, p = .001, and WRAT IV Sentence Comprehension, r_s = .46, p < .001. Its highest correlation was with the Sentence Comprehension subtest of the WRAT IV, which suggests that it may have more in common with this subtest than the other two subtests. The Comprehension subtest of the Instines was moderately positively correlated with the Sentence Comprehension subtest of the WRAT IV, r_s = .36, p = .001. This correlation is well below the minimum required to support the concurrent validity of the task and suggests that it may not be a valid measure of reading comprehension. Although the Verbal Reasoning and Comprehension subtests were correlated with the standardized measures assessing the same or similar abilities, these correlations were lower than the minimum required to demonstrate concurrent validity.

Table 6.5 Spearman's r_s Correlations between Participants' Scores on the Standardised Measures of Cognitive and Literacy Abilities and the Subtests of the Instines (N = 78)

Measures	1	2	3	4	5	6	7
1. Instines Spelling/Homophone		.13	.22	.02	.26*	.75**	.39**
2. Instines Verbal Reasoning			.31**	.32**	.38**	.21	.46**
3. Instines Comprehension				.09	.15	.13	.36**
4. WRIT Matrices					.42**	-.08	.16
5. WRIT Vocabulary						.27*	.69**
6. WRAT IV Spelling							.34**
7. WRAT IV Sentence Comprehension							

Note. * p < .05. ** p < .01.

To summarise, there was little support for the validity of the Instines, as a dyslexia screening test for adults. The performances of the groups did not differ significantly on half of its subtests. Additionally, the logistic regression analysis indicated that a model with all the subtests was not a good fit for the data. Moreover, although its sensitivity rate was at the minimum required for a good screening measure, its specificity rate was below the acceptable threshold. However, there was strong support for the validity of one subtest, the

Spelling/Homophone subtest; it produced significant between-group differences, had high positive correlation with the standardised measure of the same ability, and was a significant unique predictor of group membership.

6.4.3 Evaluation of the Reliability of the Instines

4.4.3.1 Test-retest reliability of the Instines subtests. We evaluated the test-retest reliability of the Instines by correlating participants' scores on its subtests from the first and second administrations of the test. Based on the design of the Instines, which provides only percentile scores on the subtests and no total score, reliability could only be assessed using the test-retest method. The more similar participants' rank ordering of the two sets of scores, the higher the correlation will be; a minimum correlation of .80 is considered an acceptable reliability coefficient (Kline, 2000). Six participants (1 dyslexic and 5 controls), did not undertake a second Instines screening, therefore this analysis was carried out with 72 participants (39 dyslexics and 33 controls). The scores were correlated using Spearman's r_s correlation and details of the correlations for each subtest are detailed in Table 6.6. The correlations ranged from a low of $r_s = .30$ on the Comprehension subtest to a high of $r_s = .88$ on the Spelling/Homophone subtest. The Spelling/Homophone subtest was the only subtest with a correlation above the minimum required to establish test-retest reliability. This suggests that scores on the subtest are consistent across time and indicates a 77% agreement between the participants' rank scores on the first and second administrations of the test. The Comprehension subtest had the lowest correlations, $r_s = .30$, $p < .001$; however this correlation should be treated with caution as reliability of this subtest may have been affected by the fact that the passages read and the related questions answered were different for the two administrations of the Instines. Thus, with the exception of the Spelling/Homophone subtest, the results indicate that the Instines is composed of subtests which have relatively weak stability across time.

Table 6.6 Spearman's r_s Correlations Coefficients between the First and Second Screening Outcomes and Scores on the Subtests of the Instines of Dyslexic and Control Participants (N = 72)

Instines Subtests	Spearman's r_s Correlation
Spelling/Homophone	.88
Spatial Recognition	.47
Verbal Reasoning	.67
Reading Speed	.69
Directional Awareness	.40
Forward Digit Span	.56
Reverse Digit Span	.50
Comprehension	.30

** $p < .001$

Finally, as a measure of the reliability (stability) of classification of participants by the Instines, we compared the screening outcomes of the participants on the first and second administrations of the test. Of the 72 participants, 21 (10 dyslexics and 11 controls) had different classifications on their second screening. For the 10 dyslexic participants whose classifications were changed on their second screening, four were classified as dyslexic on their first screening and as not-dyslexic on their second screening, while for the remaining six dyslexics the opposite was true. For the 11 control participants whose classifications were changed, 4 were classified as not dyslexic on their first screening and as dyslexics on their second screening, while for the remaining 7 the opposite was true. This suggests that the Instines screening outcomes are not consistent across time.

The results of the test-retest reliability analysis indicated that, with the exception of the Spelling/Homophone subtest, the reliability of the battery was below the acceptable level. Also, over time there were changes in its screening outcomes with the same participants receiving different classifications. Taken together, our results suggest that the overall reliability and validity of the Instines falls below an acceptable level.

6.5 Discussion and Conclusion

The main purpose of this study was to evaluate the Instines as a screening test for adults with dyslexia by examining its psychometric properties (validity and reliability) and its capacity to discriminate between adults with and without dyslexia. We obtained data from 78 university students, 40 with a formal diagnosis of dyslexia (dyslexic group), and 38 with no reported learning or literacy difficulties (control group). The groups' performances on the standardised measures of cognitive and literacy abilities were in the average to above range. Their performances were similar on cognitive measures (WRIT Matrices and Vocabulary) and on the Sentence Comprehension subtest of the WRAT IV. However, the groups differed in their performances on the Word Reading and Spelling subtests of the WRAT IV, with large effect sizes d = 1.18 and d = 1.83 respectively.

6.5.1 Validity of the Instines

The validity of the Instines was assessed by: (1) comparing the performances of the dyslexic and control groups on the subtests, (2) examining its capacity to discriminate between adults with and without dyslexia, and (3) examining the concurrent validity of three of its subtests: Spelling/Homophone, Verbal Reasoning, and Comprehension. The results of the comparison of the performances of the dyslexic and control groups supported the validity of four of the eight subtests of the Instines, such that group differences emerged for the Spelling/Homophone, Reading Speed, Directional Awareness, and Forward Digit Span subtests. The results suggest that the remaining four subtests lack the psychometric quality required for discriminating between individuals and/or groups with and without dyslexia. Alternatively, these tasks may be assessing constructs that are not relevant to dyslexia in adults. The Instines therefore seems a weak dyslexia screening tool. These results are contrary to those we reported earlier for the BDT and the BALI (reduced battery), as well as for the LADPLUS (Singleton et al., 2009), where the performance of adults with and without dyslexia differed significantly on all the subtests of the tests.

Further evidence that the Instines lacks the discriminatory capacity required for a good screening test was provided by the results of the logistic regression. The model that included all the subtests of the Instines, although explaining a large proportion of the variance in the groups (.42 - .56), and correctly classifying 83.3% of the participants, was not a good fit for the data. Additionally, although the sensitivity rate (80%) reached the minimum required for a good screening test, the specificity rate 86.8% fell below the minimum 90%. The sensitivity rate is much lower than the rates reported earlier for the BDT (96%) and the BALI (91.9%), as well as other screening tests DAST (94%) (Fawcett & Nicolson, 1998), and LADSPLUS (90.6%) (Singleton et. al., 2009). More importantly, the sensitivity rate is well below that reported in the software's frequently asked questions section which indicates a 94% - 98% agreement between the screening outcomes of the Instines and diagnosis by a psychologist; details of how these figures were obtained were not given, however. The specificity rate of the Instines was also lower than that observed for of the BALI (89.5%), and that reported for the DAST (100%) (Fawcett & Nicolson, 1998), and the LADPLUS (90%) (Singleton et. al., 2009). This may result in a larger than acceptable proportion of adults without dyslexia being incorrectly identified as being at risk of dyslexia (false positives). Of the eight subtests in the Instines, the Spelling/Homophone subtest was the only unique significant predictor of group membership, although four other subtests while not being significant predictors also contributed to the predictive power of the test. A model with these five subtests (Spelling/Homophone, Reading Speed, Directional Awareness, Forward Digit Span, and Reverse Digit Span) provided a good fit for the data and increased the overall classification and sensitivity rates of the Instines; nevertheless, its specificity rate of 84.2% failed to meet the minimum threshold. These results suggest that the overall capacity of the Instines to discriminate between adults with and without dyslexia may largely be due to the Spelling/Homophone subtest.

Information in the *frequently asked questions* section of the software for the test states that screening outcomes are obtained by comparing the performance of the person being assessed to that of other dyslexic and non-dyslexic individuals in its data base; however no further details of this process are given. Nevertheless, if half of subtests of the Instines do not discriminate between adults with and without dyslexia, or are not sufficiently sensitive to detect differences between these groups, this is likely to adversely affect the overall capacity of the Instines to identify adults with dyslexic traits.

In order for the Instines to be a valid screening test, its subtests should represent valid constructs. Yet, only the Spelling/Homophone subtest yielded evidence of construct validity. The concurrent validity of two other subtests, Verbal Reasoning and Comprehension, was not supported, as their correlations with the standardised measures were relatively weak. The results suggest that at least two of the subtests of the Instines Verbal Reasoning and Comprehension may not be valid measures for the abilities they purport to assess. This would adversely affect the validity of the test as a whole.

Further and even more compelling evidence that the Instines may not be a valid dyslexia screening test for adults is provided by the lack of stability in the classification of participants by the Instines. All the participants in the dyslexic group had a formal diagnosis of dyslexia from an appropriately qualified professional. Also, none of the participants in the control group reported learning or literacy difficulties, and their scores on the cognitive and literacy measures provide support for this. We are therefore, confident that most, if not all, individuals in the respective groups (particularly the dyslexic group) were correctly classified. However, the Instines only correctly classified 44 (56%) of the 78 participants (22 dyslexics and 22 controls), which is just above chance.

6.5.2 Reliability

The results of the assessment of the reliability of Instines were also unimpressive. Only the Spelling/Homophone subtest had an acceptable level of test-retest reliability r_s = .88, (above the minimum required), indicating that scores on the subtest are consistent across time. For all the other subtests, the test-retest reliability ranging from a low of r_s = .30 on the Comprehension subtest and a high of r_s = .69 on the Reading Speed subtest, were below the minimum (Kline, 2000).

To conclude, the study provided strong empirical support for the validity and reliability of the Spelling/Homophone subtest of the Instines but not for the test as a whole. The Spelling/Homophone subtest discriminated between the performance of the participants with and without dyslexia. It was highly correlated with the standardised measure assessing a similar ability and the scores of the participants on it were consistent (reliable) across time. For the test as a whole, its capacity to discriminate between adults with and without dyslexia is limited as only half of the subtests are capable of this. More importantly, the test-retest reliability of most of its subtests is below acceptable level.

Chapter 7: Cognitive Predictors of Literacy Skills in Adults with Dyslexia (Study 4)

The ongoing debate on the use of IQ discrepancy in the definition of dyslexia has resulted in a number of studies investigating the validity of the definition. These studies have mostly examined the cognitive profiles of children, with and without a diagnosis of reading disabilities, and with and without a discrepancy between their IQ and literacy skills (Badian, 1994; Fletcher et al., 1994; Francis et al., 2005; Stanovich & Siegel, 1994; Stuebing et al., 2002). The performances of these children were compared on a variety of measures and the majority of the studies concluded that IQ should not be used in the definition of reading disability as both groups of children exhibit similar cognitive deficits with similar predictors (Fletcher et al., 1994; Stanovich & Siegel, 1994), require the same intervention, and respond in a similar manner to intervention (Stage, Abbott, Jenkins, & Berninger, 2003). For more details, please see Literature Review, Chapter 1. Compared to the research undertaken with children, much less research on the validity of the IQ-discrepancy definition has been conducted with adults. However, Swanson and Ching-Ju (2009), in a meta-analysis of 52 studies comparing adults with and without reading disabilities on a range of cognitive and vocational measures, concluded that verbal IQ might be important for the identification of reading disabilities in adults. For the meta-analysis, studies including adults with reading disabilities, with and without an IQ discrepancy, were analysed. The researchers found larger effect sizes in studies that included participants with a discrepancy in their IQ and reading levels when compared to studies with participants whose IQ was consistent with their reading scores. Given the difference between Swanson and Ching-Ju's (2009) findings and those of the majority of studies with children, as well as the dearth of research with adults, there is a need for further research similar to that undertaken with children in order to elucidate the importance of IQ for the definition of dyslexia in adults.

7.1 Cognitive Predictors of Literacy Skills

Studies of literacy development and the predictors of literacy skills in children have identified the main cognitive predictors of skills such as reading, spelling, and reading comprehension. Although it is difficult making direct comparisons across studies that have used different combinations of predictor measures, and different measures of literacy, some similarities have been found across studies, and several skills emerge most systematically as predictors in alphabetic orthographies. For word-level reading and spelling, studies with children have consistently identified letter knowledge, phoneme awareness, and RAN, as significant early predictors in several alphabetic languages (Caravolas, Lervåg, Defior, Malková, & Hulme, 2013; Caravolas et al., 2012; Caravolas, Volín, & Hulme, 2005; Moll et al., 2014; Muter, Hulme, Snowling, & Stevenson, 2004; Parrila, Kirby, & McQuarrie, 2004). For example, research by Muter and colleagues, in a two years longitudinal study of the predictors of word reading accuracy and reading comprehension in English-speaking children at the start of schooling, found that word reading was predicted by letter knowledge and phoneme awareness, and reading comprehension by vocabulary knowledge, word reading and grammatical skills. In another study Caravolas et al. (2012) investigated the importance of phoneme awareness relative to other predictors in the development of reading and spelling in school beginners in four different language groups (English, Spanish, Czech and Slovak). These researchers found the same pattern of prediction across all languages, such that phoneme awareness, letter knowledge, and RAN, but not verbal short term memory or IQ, accounted for unique variance in both early word reading and spelling. Similarly, in a study of five European languages (English, French, German, Hungarian, and Finnish), with children (age range 111.4 to 129.7 months), phonological awareness and RAN were significant predictors of reading accuracy and spelling in all the languages (Moll et al., 2014). For reading comprehension, the important predictors that have been observed in studies with children include vocabulary, word reading accuracy, and word reading speed, as well as nonverbal skills/abilities, phoneme awareness, and grammatical skills (Caravolas, Volín, & Hulme, 2005; Clarke, Snowling, Truelove, & Hulme, 2010; Muter et al., 2004; Savage et al., 2005).

In addition to the predictors detailed above, the importance of oral language skills in reading comprehension has been highlighted by an established theory: The Simple View of

Reading. The theory which was originally proposed by Gough and Tunmer (1986) states that reading consists of two components, decoding and linguistic comprehension (Hoover & Gough, 1990). Decoding is defined as efficient word recognition, and linguistic comprehension as the ability to extract meaning from lexical information. Both components are equally important and necessary for success in reading which is not achievable with only one component. Further, the abilities required for reading comprehension are the same as those involved in linguistic comprehension. The theory makes a number of predictions. Firstly, decoding and linguistic comprehension skills will independently affect reading comprehension abilities. Secondly, poor reading comprehension skill can be caused by one of three conditions: (1) good decoding and poor linguistic comprehension skills; (2) good linguistic comprehension and poor decoding skills; (3) poor decoding and linguistic comprehension skills. Thirdly, reading comprehension is proportional to decoding and linguistic comprehension.

Surprisingly, we were not able to find studies that investigated the cognitive predictors of reading or spelling for adults with or without reading disabilities. However, research studies conducted with illiterate or low literacy adults have indentified predictors similar to those found in children (Durgunoğlu & Öney, 2002; Greenberg, Ehri, & Perin; 1997; Thompkins & Binder, 2003). Durgunoğlu and Öney (2002) in a study designed to determine the cognitive processes of literacy acquisition, examined the predictors of word recognition, spelling, and reading comprehension skills of 59 female adults in Turkey, who entered the study with no formal schooling, after they had undergone a 90-hour literacy program. Using regression analyses with pre-course listening comprehension, phonological awareness, and letter recognition as predictors, they found that phonological awareness predicted word recognition and spelling, and additionally letter knowledge was also a predictor of word recognition. For reading comprehension, phonological awareness and word recognition were significant predictors. In another interesting study, Greenberg, Ehri, and Perin (1997) compared the word reading processes in adult literacy students and children in grades three to five matched for reading level. Using multiple regression analyses, these researchers found that the unique predictors of word reading accuracy were the same for both groups, namely non-word decoding, sight word reading (irregular words), word likeness (selecting from a pair of nonwords the word that is more like a real word), rhyme word reading (selecting the pairs of words that rhymes) and spelling. Finally, Thompkins and

Binder (2003) used regression analysis to identify predictors of reading for 60 functionally illiterate adults enrolled in adult education programs. They found that phoneme awareness, spelling, and word-picture matching (matching pictures to words describing the pictures) were unique predictors of reading skills.

For predictors of reading comprehension, a few studies investigating this in adults with low literacy skills (Mellard, Fall, & Woods, 2010; Braze, Tabor, Shankweiler, & Mencl, 2007), and with dyslexia (Ransby & Swanson, 2003), have identified predictors similar to those for children. Mellard, Fall, and Woods (2010), investigated the predictors of reading comprehension (using path analysis) in 174 adults with low literacy skills who were enrolled in adult basic education and secondary education programmes. These researchers found that the model of reading comprehension, which best fitted the data, included: word reading and verbal IQ as direct and indirect predictors (mediator variables), reading fluency and language comprehension as direct predictors, and RAN Letters, non-word reading and auditory working memory as indirect predictors. Braze, Tabor, Shankweiler, and Mencl (2007) examined the predictors of reading comprehension of 44 young people aged 16-24 years who did not complete secondary education and were enrolled in adult education centres. In this study, the researchers used regression analysis and identified listening comprehension, non-word reading, and vocabulary as predictors of reading. Finally, Ransby and Swanson (2003) investigated the predictors of reading comprehension in young adults with dyslexia. These researchers compared the dyslexic adults to chronological age-matched adults and reading level-matched children on measures of phonological awareness, RAN, working memory, general knowledge (a measure of verbal IQ), vocabulary and listening comprehension. Hierarchical regression analyses were used to identify independent predictors of reading comprehension accuracy (Wide Range Achievement Test-Revised) and fluency (the Fast Reading subtest of the Stanford Diagnostic Reading Test). The analyses revealed that the predictors of reading comprehension accuracy were phonological awareness, RAN, vocabulary, general knowledge, and listening comprehension, and for reading comprehension fluency phonological awareness, working memory, and listening comprehension. In conclusion, although very limited, to date, the findings of existing research indicate that the cognitive predictors of literacy skills in adults are similar to those in children.

An important related issue is whether the cognitive predictors of literacy skills for adults with reading difficulties are the same for those who have a discrepancy between their IQ and their literacy skills, and for those who do not. Contrary to the results of the studies with children, the meta-analysis by Swanson and Hsieh (2009) suggests that the cognitive predictors of literacy in adults with reading difficulties, with and without an IQ-literacy level discrepancy, may differ, and verbal IQ, in particular, may play an important role. More research with adults is needed in order elucidate this issue. This is important because if there are no differences between the profiles of cognitive predictors of literacy in dyslexic adults with and without discrepant IQ levels, then this would support the elimination of the IQ discrepancy definition of dyslexia in both children and adults. Moreover, the diagnostic assessment process for dyslexia in adults should reflect this change. In this study, we examined whether two groups of dyslexic adults with relative differences in their levels of general ability (verbal and nonverbal IQ), who were diagnosed on the basis of a discrepancy definition, differ in their cognitive profile. We asked whether the groups, university students and non-students, would differ in their cognitive predictive profiles relative to each other and relative to a control group of typical adult (student) readers. The cognitive predictors examined included verbal and nonverbal IQ (WRIT Matrices and Vocabulary subtests), to investigate the effects of IQ on literacy skills. Also examined were reading, spelling, and reading comprehension (WRAT IV Word Reading, Spelling, and Sentence Comprehension subtests), and literacy related sub-skills (phoneme awareness and RAN), which research indicates are significant predictors of literacy skills in both children and adults. We were interested in the potential effect of IQ, over and above other key measures, on their literacy skills. Specifically we attempted to answer the following questions:

1. Do the profiles of cognitive predictors of literacy differ in adults with and without dyslexia?
2. Do the profiles of cognitive predictors of literacy differ as a function of general ability (IQ) in adults with dyslexia?
3. `Do verbal and nonverbal IQ themselves predict literacy skills of adults with dyslexia?
4. Do the potential relationships between verbal and nonverbal IQ and literacy skills differ in groups of adults with dyslexia with differences in IQ?

7.2 Method

7.2.1 Participants

Three groups of participants, all native English speakers, were included in the study. Two groups of student participants were recruited from Bangor University: a control group, and a student dyslexic sample. A third group of non-student dyslexics was recruited from Remploy. Remploy is a major provider of employment services to individuals with disabilities, in long term unemployment and also those in employment, in the United Kingdom. The student participants were the same as those described in Chapter 5, Study 2b; the characteristics of each group are detailed in Table 7.1 below. The student dyslexics were all formally assessed for dyslexia by qualified professionals and the control participants reported no learning or literacy difficulties and satisfied the minimum criteria of standard scores of 85 and above on the cognitive and literacy measures. The selection criteria for the non-student dyslexics were a standardised score of 70 or greater on the Matrices subtest of the WRIT, and a standardised score of 85 or less on either the Word Reading or Spelling subtests of the WRAT IV. Twenty seven of these participants were unemployed, while 13 were employed. The majority (38) were educated up to secondary level, and 2 up to undergraduate level. Four participants reported mental health difficulties; in addition to learning/reading disability, these included depression, anxiety, and obsessive compulsive disorder. The age difference between the groups was statistically significant $F(2, 117) = 90.35$, $p < .001$, with each group differing from the other two groups.

Table 7.1. Participants' Characteristics by Sub-samples

Characteristics	Sub-samples		
	Controls	Student Dyslexics	Non-student Dyslexics
n	40	40	40
Age			
M	19.27	21.72	40.35
SD	1.95	5.66	11.86
Gender			
Male	6	17	25
Female	34	23	15

7.2.2 Measures

Participants were all assessed on a range of measures including standardised tests of verbal and nonverbal IQ (WRIT Matrices and Vocabulary subtests), reading, spelling, reading comprehension (subtests from the WRAT IV), the BALI, and the Instines as described in Chapter 4 Study 2a and Chapter 5 Study 2b. For this study, the measures used were: WRIT Matrices and Vocabulary subtests, WRAT IV Reading, Spelling, and Sentence Comprehension subtests, and the Phoneme Deletion Accuracy and RAN Digits tasks from the BALI. The WRIT Matrices and Vocabulary subtests were used to assess the nonverbal and verbal IQ of the participants. Given our interest in the effects of IQ on the literacy skills of the participants, we recognised that it would have been more prudent to assess IQ on more comprehensive measures of nonverbal and verbal abilities; however, the number of cognitive predictors included in our analyses was restricted by the sample size. The Phoneme Deletion Accuracy and RAN Digits tasks were included to assess phoneme awareness and phonological processing speed as literacy sub-skills.

7.2.3 General Procedure

All participants were tested individually in a session lasting approximately 80 - 90 minutes. The standardised measures were all administered in accordance with administration instructions, and the BALI tasks in line with established instructions (see Chapter 4 Study 2a). Data from non-student dyslexics were collected at Remploy offices in England and Wales for the participants who were unemployed and, for the employed participants, at their places of employment. Data from students were collected at Bangor University.

7.2.4 Data Analysis

Prior to analyses, the data were checked for outliers and scores above or below 2.5 standard deviations from the mean were adjusted to the score corresponding to 2.5 standard deviations from the mean. To identify outliers the mean for each group was calculated separately. For the control and dyslexic groups the outliers adjusted were the same as that listed for the BALI main study in Chapter 5. For the non-student dyslexic group one score was adjusted on the WRAT IV Sentence Comprehension subtest and two scores on the RAN Digits task. The performances of the groups on the measures administered were compared using one way analysis of variance (ANOVA) with post-hoc comparison, and Bonferroni correction for multiple comparisons. For post-hoc comparison, Tukey HSD test was used for measures with equal variance and Games-Howell for measures with unequal variances. The data set were also checked for the assumptions of path analysis following Tabachnick and Fidell's (2007) recommendations and these were satisfied. Multigroup path analyses in Mplus (Version 7.11; Muthén & Muthén, 2013) were used to identify the predictors of literacy scores, to examine the similarities/differences in the groups' cognitive profiles, and the effect of IQ on literacy abilities.

7.2.5 Results

7.2.5.1 Comparison of the Performance of the Groups on the Measures Administered.

We first compared the groups' performances on the measures administered using ANOVA. Descriptive statistics for the groups on the measures are reported in Table 7.2 below. The results of the ANOVA revealed significant differences between the groups on all the measures as follows: WRIT Matrices $F(2, 117) = 18.37$, $p < .001$, $\eta^2 = 0.24$, WRIT Vocabulary $F(2, 117) = 60.99$, $p < .001$, $\eta^2 = 0.51$, WRAT IV Reading $F(2, 117) = 57.30$, $p < .001$, $\eta^2 = 0.49$, WRAT IV Spelling $F(2, 117) = 120.83$, $p < .001$, $\eta^2 = 0.67$, WRAT IV Sentence Comprehension $F(2, 117) = 53.68$, $p < .001$, $\eta^2 = 0.48$, Phoneme Deletion $F(2, 117) = 51.16$, $p < .001$, $\eta^2 = 0.47$, RAN Digits $F(2, 117) = 9.98$, $p < .001$, $\eta^2 = 0.15$. For the standardised measures, post-hoc comparisons indicated that the student participants, (controls and student dyslexics) (as highlighted in Chapter 5 Study 2b) were comparable on the measures of verbal and nonverbal IQ, and reading comprehension, and only differed on the reading and spelling measures. The groups also performed in the average to above average range on these measures. Conversely, the performances of the non-student dyslexics were significantly different from the control and student dyslexic groups on all the standardised measures. Additionally, on these measures (with the exception of the WRIT Matrices subtest), the non-student dyslexics performed in the low average range. For the phoneme deletion task, post-hoc comparisons indicated differences in the performance of all three groups, with the controls performing better (obtaining higher scores) than the dyslexic groups, and the student dyslexics performing better than the non-students. For the RAN Digits task, post-hoc comparisons indicated differences in the performance of the control and dyslexic groups, with the controls performing better (obtaining lower scores), but not between the two dyslexic groups. The magnitude of the differences in the groups' performances varied from small (RAN Digits $\eta^2 = 0.15$) to very large (WRAT Spelling $\eta^2 = 0.67$).

Table 7.2. Mean Standard Scores (Standard Deviation in Parenthesis) of Groups on Measures and Effect Sizes (N = 120)

Measures	Controls (*n* = 40)	Student Dyslexics (*n* = 40)	Non-student Dyslexics (*n* = 40)	η^2
WRIT Matrices	104.98 (10.75)	106.50 (13.15)	91.18 (13.30)	0.24
WRIT Vocabulary	104.50 (11.43)	102.32 (11.15)	76.28 (15.18)	0.51
WRAT IV Word Reading	103.10 (7.62)	93.88 (7.88)	83.45 (9.07)	0.49
WRAT IV Spelling	111.98 (12.81)	92.32 (8.46)	77.75 (7.56)	0.67
WRAT IV Sentence Comprehension	109.82 (9.93)	105.82 (12.54)	84.25 (12.63)	0.48
Phoneme Deletion	21.65 (3.46)	18.47 (5.19)	7.83 (9.17)	0.47
RAN Digits	14.93 (3.01)	19.30 (6.34)	19.18 (5.03)	0.15

As reported earlier, the age difference between the groups was statistically significant with each group differing from the other two groups. For the standardised measures, age is taken into account in the scoring procedure, and therefore it is unlikely that the age difference affected the results obtained here or in the other analyses conducted with these measures. For the other measures (Phoneme Deletion and RAN Digits), we examined their correlations with age for the total sample. Both tasks were significantly correlated with age; Phoneme Deletion $r(120) = -.52$, $p < .01$, and RAN Digits $r(120) = .26$, $p < .01$. However, a further examination of the correlations for each group separately revealed that for both tasks the correlations with age were not significant in any of the groups: non-student dyslexic group -- Phoneme Deletion $r(40) = .08$, $p = .615$ and RAN Digits $r(40) = .29$, $p = .074$, student dyslexics -- Phoneme Deletion $r(40) = -.28$, $p = .079$ and RAN Digits $r(40) = .06$, $p = .719$, and controls -- Phoneme Deletion $r(40) = .10$, $p = .531$ and RAN Digits $r(40) = -.12$, $p = .448$. This suggests that the significant correlations between Phoneme Deletion, RAN Digits and age (with the total sample) are likely due to the group differences on the tasks and not an association between age and the measures (see also Ramus, Pidgeon, & Frith, 2003). It is therefore, unlikely that the age differences between the groups affected the results obtained in the analyses conducted with these measures.

To summarize, the control group's performance differed significantly from that of the non-student dyslexics on all the measures, while only differing from the student dyslexics on reading and spelling. For the dyslexic groups (students and non-student) their performances differed significantly on all the measures except RAN Digits. This suggests that the difference in the performance of the controls and the student dyslexics on the reading and spelling measures could not be attributed to IQ. Therefore, with IQ differences in only the dyslexic groups, it was possible to address the question of the role of IQ on their literacy skills.

7.2.6 Cognitive Predictors of Literacy Skills.

We compared the profile of cognitive predictors of literacy skills (reading, spelling, and reading comprehension) among the groups, using multigroup path analyses. We were particularly interested in comparing the pattern of predictors for the two dyslexic groups and the specific roles of verbal and nonverbal IQ as predictors of literacy skills over and above literacy related skills. For each path model, we included known cognitive predictors of literacy skills along with verbal and nonverbal IQ and conducted a series of mutligroup path analyses in Mplus using Maximum Likelihood Estimator. We started with a fully saturated model in which all the predictors were allowed to vary freely across groups. Then we systematically constrained significant predictors across groups in order to determine if the effects of the predictors were the same. Predictors were constrained in order of significance and were retained only if they did not adversely affect model fit. We tested each model with chi-square significance test along with other model fit indices to determine the best fitting model for each literacy skill across the groups.

Word reading model. For the word reading model, we were interested in the effects of IQ on the literacy skills of the participants, and therefore included both measures of verbal and nonverbal IQ as predictors of scores on the WRAT IV Reading test, a measure of single word reading accuracy. Spelling (WRAT IV Spelling subtest) was also included as a predictor as the literature indicates that it is related to reading, with high correlation between the skills and shared cognitive processes (phonological awareness and alphabet knowledge) (Cataldo & Ellis, 1988; Santoro, Coyne, & Simmons, 2006). Also, previous studies have found that spelling is a significant predictor of reading in adults with low literacy (Greenberg et al., 1997; Thompkins & Binder, 2003). Finally, we also included Phoneme Deletion Accuracy and RAN Digits as predictors, as there is strong and consistent research evidence that they are two of a number of cognitive processes underlying word recognition (and other literacy skills) and are unique independent predictors (Caravolas et al., 2012; Georgiou, Parrila, & Papadopoulos, 2008; Kirby et al., 2010; Moll et al., 2014). The results of the BALI main study (see Chapter 5 Study 2b) indicated that Phoneme Deletion speed was better than Phoneme Deletion accuracy at discriminating between adults with and without dyslexia. However, Phoneme Deletion accuracy was used for the path analyses because data on Phoneme Deletion speed was missing for the majority of the non-student dyslexics. The non-student dyslexics failed the practice items on the tasks and therefore the full task was not administered and no scores for time were available. Therefore, the saturated model for word reading included: verbal and nonverbal IQ (WRIT Vocabulary and Matrices), Phoneme Deletion, RAN Digits, and WRAT IV Spelling, as direct predictors of reading. With the exception of RAN Digits, word reading was significantly positively correlated with all the other predictors in the model with moderate to high correlations (see Table 7.3).

Table 7.3. Correlations between Variables in Path Models (N = 120)

Measures	1	2	3	4	5	6	7
1. WRIT Matrices	-	.60**	.32**	.27**	.58**	-.15	.44**
2. WRIT Vocabulary		-	.63**	.51**	.78**	-.05	.59**
3. WRAT IV Word Reading			-	.80**	.74**	-.17	.70**
4. WRAT IV Spelling				-	.62**	-.45**	.66**
5. WRAT IV Sentence Comprehension[a]					-	-.19*	.71**
6. RAN Digits						-	-.28**
7. Phoneme Deletion Accuracy							-

Note. [a]$N = 118$

* $p < .05$. ** $p < .01$

In the saturated model, RAN Digits was a significant predictor for the dyslexic groups but not the controls; however, it was also excluded from the final model, as the Beta coefficient scores (for both dyslexic groups) were positive. This is unusual given the well established negative relation between RAN tasks and reading in children, such that high scores on RAN are correlated to low scores on reading (Georgiou, Parrila, & Kirby, 2006; Georgiou, Parrila, & Liao, 2008; Swanson, Trainin, Necoechea, & Hammill, 2003). What is more, RAN Digits and reading were not significantly correlated for the sample as a whole (see Table 7.3) or for any of the groups: controls $r(40) = -.29$, $p = .072$, student dyslexics $r(40) = .10$, $p = .549$, and non-student dyslexics $r(40) = .08$, $p = .605$, suggesting that this association was not important in the current data set.

The final, partially constrained, model included three significant cognitive predictors (verbal IQ, Phoneme Deletion, and WRAT IV Spelling). The model was an excellent fit to the data, $\chi^2(4, N = 120) = 2.05$, $p = .727$, comparative fit index (CFI) = 1.00, Tucker-Lewis index (TLI) = 1.05, root mean square error of approximation (RMSEA) = 0.000, 90% confidence interval = [0.000, 0.173], standardized root mean square residual (SRMR) = .024. The path model for the control group is shown in Figure 7.1 and for the dyslexic groups in Figure 7.2. The profile of significant predictors for the dyslexic groups (students and non-students) was the same, and similar to that of the control group. The only difference in the profile of cognitive predictors between the dyslexic and control groups was that verbal IQ was a predictor for the two dyslexic groups only. The path weights from phoneme deletion were constrained to be equal across all groups indicating that the relative effect of this predictor on reading was the same for all groups. Furthermore, verbal IQ and spelling were also constrained to be equal across the dyslexic groups; therefore not only were the predictors of reading the same for both dyslexic groups, but so too were the relative effects of these predictors.

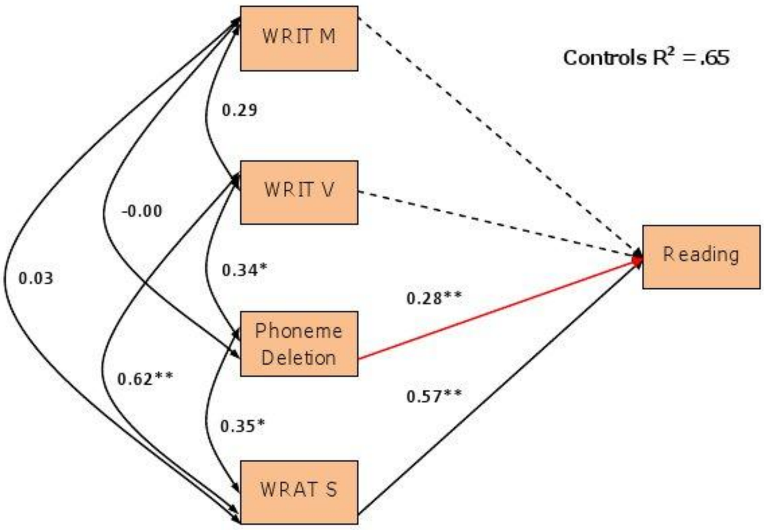


Figure 7.1. Path model of reading for control group with unstandardized path weights and correlations. Red line represents variable constrained across all groups. WRIT M = WRIT Matrices; WRIT V = WRIT Vocabulary; WRAT S = WRAT IV Spelling.

** p < .05, ** p < .01.*

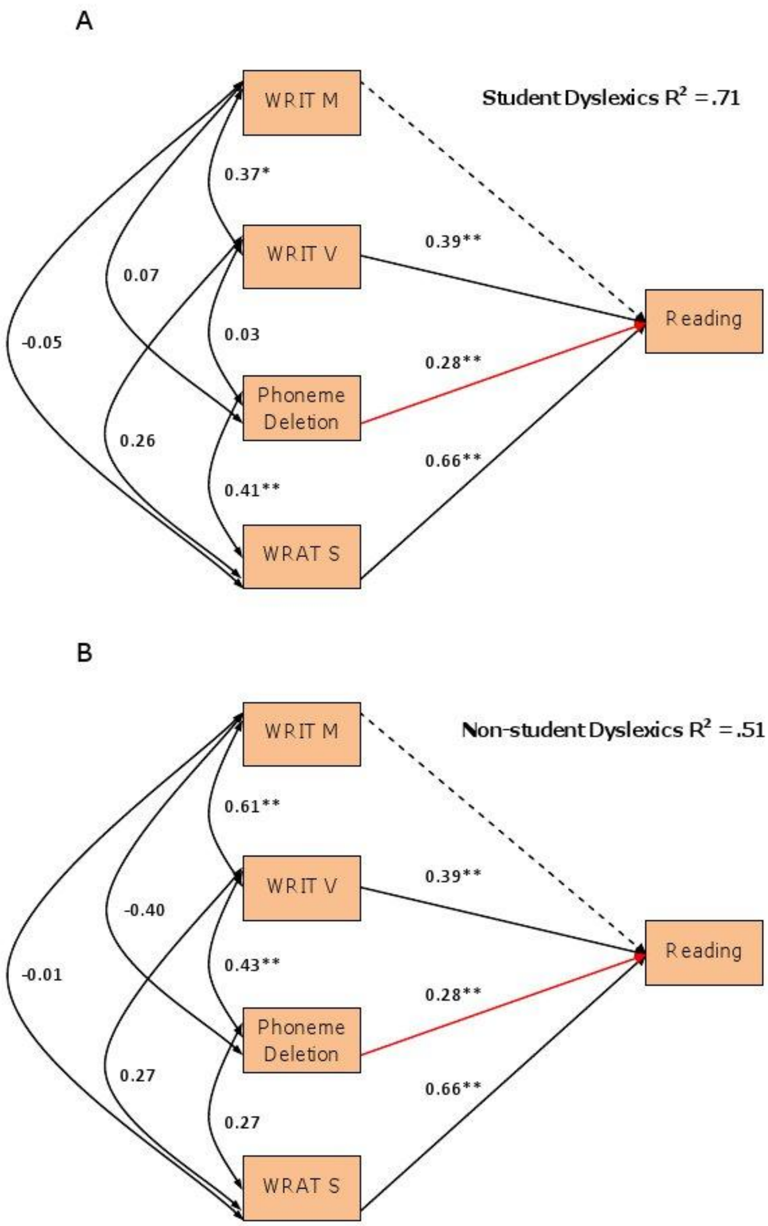


Figure 7.2. Path model of reading for dyslexic groups (A) student dyslexic and (B) non-student dyslexics with unstandardized path weights and correlations. Red line represents variable constrained across all groups. WRIT M = WRIT Matrices; WRIT V = WRIT Vocabulary; WRAT S = WRAT IV Spelling.

$* p < .05$, $** p < .01$.

The models accounted for high proportions of the variance in reading for all the groups, but more so for the student dyslexic group R^2 = .70, and the controls R^2 = .63, with a lower proportion explained for the non-student dyslexics R^2 = .51. For the non-student dyslexic group, the amount of variance explained by the model suggests there may be other cognitive predictors of reading that are important for this group that were not included in the model. The standardized beta coefficient scores indicated that the importance of the significant predictors was very similar for the dyslexic groups and this differed from that of the controls, however, spe1ling was the most important predictor for all groups, (controls β = .59, student dyslexics β = .60, and non-student dyslexic β = .43) and accounted for most of the variance in reading. For the dyslexic groups the importance of the other predictors (verbal IQ and phoneme deletion) was very similar, for the student dyslexics verbal IQ β = .31and phoneme deletion β = .28, and for the non-student dyslexics verbal IQ β = .38 and phoneme deletion β = .33.

Spelling model. For the spelling model, the predictors included: WRIT Matrices, WRIT Vocabulary, Phoneme Deletion, and RAN Digits; these were the same as the reading model and the rationale for their inclusion was similar. In addition, WRAT Reading was also included because, as already stated, it has a well established relationship with spelling. All the predictors were initially included in the model as direct predictors of spelling; however, based on information from the model modification indices during the process of identifying the best fitting model for the data, the role of reading was changed to a mediator. The saturated model therefore included verbal and nonverbal IQ (WRIT Vocabulary and Matrices subtests), Phoneme Deletion, and RAN Digits, as predictors of spelling; and WRAT Reading as a mediator variable between the foregoing measures and spelling. Spelling was significantly correlated to all the predictors in the model with low to high correlations (see Table 7.3).

The final model was an excellent fit to the data, $\chi^2(3, N = 120) = 0.472$, $p = .925$, comparative fit index (CFI) = 1.00, Tucker-Lewis index (TLI) = 1.134, root mean square error of approximation (RMSEA) = 0.000, 90% confidence interval = [0.000, 0.086], standardized root mean square residual (SRMR) = .026. The path model for the control group is shown in Figure 7.3, the student dyslexic and non-student dyslexic groups in Figures 7.4, and 7.5 respectively. For the dyslexic groups, the model yielded the same pattern of significant predictors. First, reading was the most powerful predictor of spelling, as would be expected. Second, RAN Digits made a direct contribution to spelling, but no indirect contribution via reading. In contrast, verbal IQ and phoneme deletion both accounted for variance in spelling indirectly through reading. Nonverbal ability did not account for unique variance in spelling for either the dyslexic group, or the control group. The significant predictors for the control group were similar to those for the dyslexic groups; however, verbal IQ predicted spelling both indirectly via reading and directly.

In fitting the final model, RAN Digits was constrained to be equal across all groups indicating that the relative effect of this predictor on spelling was the same for all groups. Beyond RAN, the weighting of the paths in the control group differed from those of the dyslexic groups. Reading made the strongest contribution, while the remaining three predictors accounted for roughly similar proportions of variance in spelling. In addition, verbal IQ was a significant direct predictor for the control group only, suggesting that the group may be more reliant on this ability for their spelling attainment than the dyslexic groups. For the dyslexic groups, verbal IQ had only an indirect effect through reading and was constrained to be equal, indicating that its effect was comparable.

The models accounted for a high proportion of the variance in spelling for the controls $R^2 = .70$ and student dyslexics $R^2 = .74$ groups, and a relatively lower proportion of the variance in the non-student dyslexics $R^2 = .47$. Similar to the results for the reading model, this suggests that for the non-student dyslexic group there may be important predictor/s that is/are not included in the model. The standardized beta coefficient scores indicated that the importance of the significant predictors was very similar for the dyslexic groups and this differed from that of the controls; however, reading was the most important predictor (direct and indirect) for all groups, (controls $\beta = .55$, student dyslexics $\beta = .85$, and non-student dyslexics $\beta = .69$) and accounted for most of the variance in spelling.

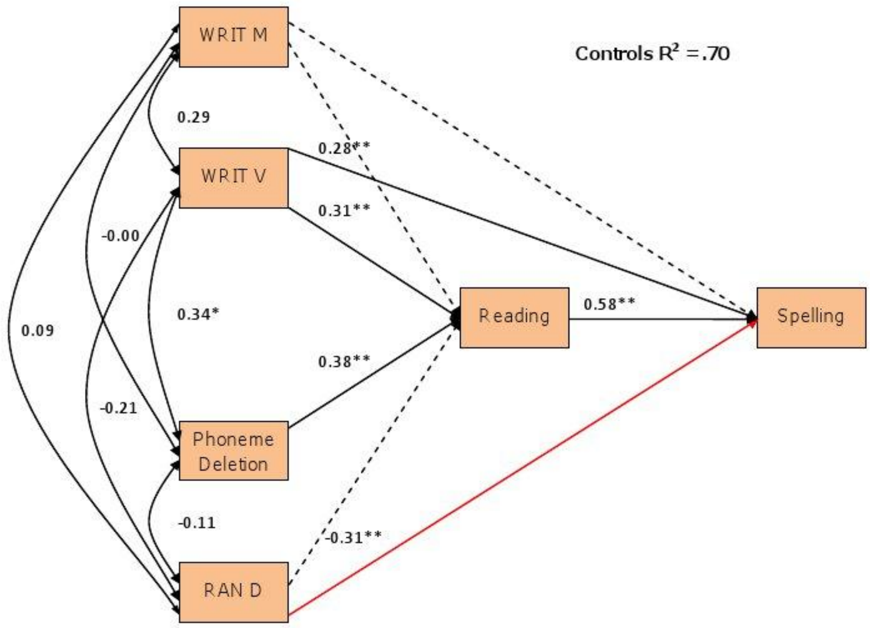


Figure 7.3. Path model of spelling for control group with unstandardized path weights and correlations. Red line represents variable constrained across all groups. WRIT M = WRIT Matrices; WRIT V = WRIT Vocabulary; RAN D = RAN Digits.

* p <.05, ** p < .01.

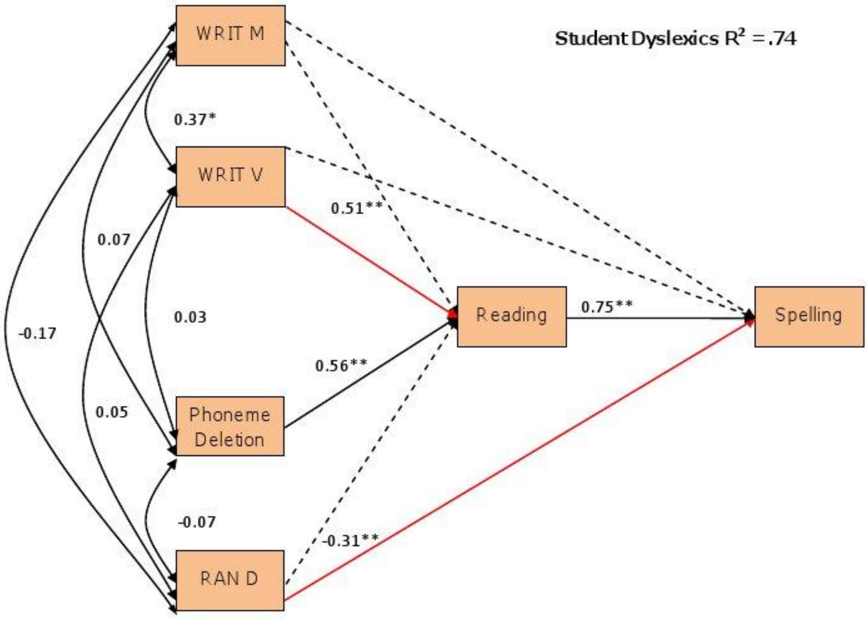


Figure 7.4. *Path model of reading for student dyslexic group with unstandardized path weights and correlations. Red lines represent variable constrained across all groups and dyslexic groups only. WRIT M = WRIT Matrices; WRIT V = WRIT Vocabulary; RAN D = RAN Digits.*

** p <.05, ** p < .01.*

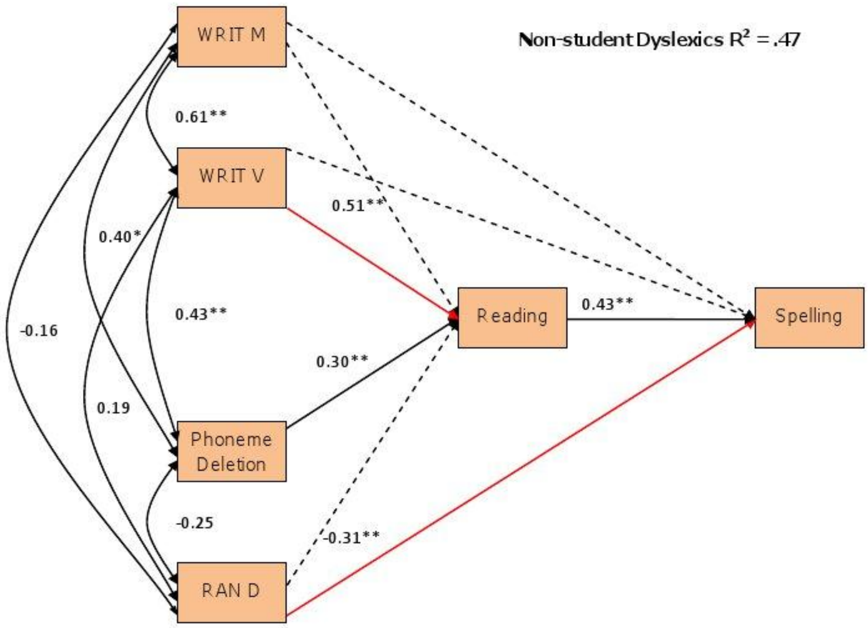


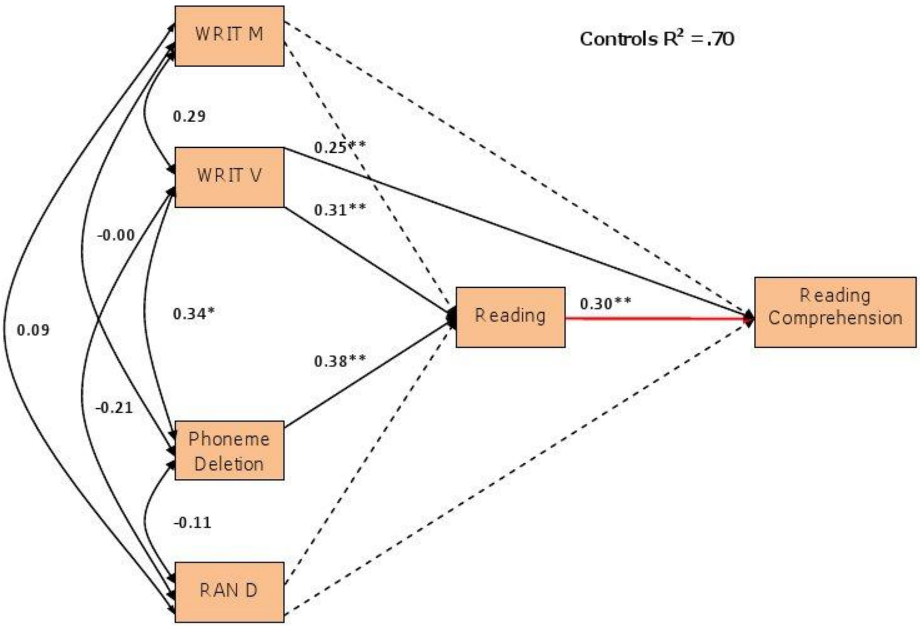
Figure 7.5. Path model of reading for non-student dyslexic group with unstandardized path weights and correlations. Red lines represent variable constrained across all groups and dyslexic groups only. WRIT M = WRIT Matrices; WRIT V = WRIT Vocabulary; RAN D = RAN Digits.

** p <.05, ** p < .01.*

Reading comprehension model. For the reading comprehension model, the predictors included were WRIT Matrices and WRIT Vocabulary (for reasons already stated). WRAT Reading and Phoneme Deletion were also included, based on the findings of previous studies which indicated that these skills were significant predictors of reading comprehension for children and adults (Caravolas et al., 2005; Durgunoğlu & Öney, 2002; Muter et al., 2004; Savage et al., 2005). Finally, RAN Digits was included as it is a reading related sub-skill and had a significant negative (although low) correlation with reading comprehension. WRAT Reading was included in the model as a direct predictor and a mediator, based on information from the model modification indices during the process of identifying the best fitting model for the data. Additionally, Mellard et al. (2010) in their study of reading comprehension with adults with low literacy also found that a model with reading as a direct predictor and mediator for reading comprehension was a good fit for their data. Therefore, for reading comprehension, the saturated model included WRAT Reading as a mediator and direct predictor, with verbal and nonverbal IQ, Phoneme Deletion, and RAN Digits as direct and indirect predictors (through WRAT Reading). Reading comprehension was significantly correlated with all the predictors in the model with low to high correlations (see Table 7.3).

The final model was an excellent fit to the data, $\chi^2(4, N = 120) = 2.42$, $p = .659$, comparative fit index (CFI) = 1.00, Tucker-Lewis index (TLI) = 1.06, root mean square error of approximation (RMSEA) = 0.000, 90% confidence interval = [0.000, 0.189], standardized root mean square residual (SRMR) = .065. The path model for the control group is shown in Figure 7.6, the student dyslexic, non-student dyslexic groups in Figures 7.7 and 7.8, respectively. The pattern of significant predictors for the dyslexic groups was very similar. For both groups, verbal and nonverbal IQ, and reading were direct predictors, and verbal IQ and phoneme deletion indirect predictors through word reading. Additionally, for the non-student dyslexic group, RAN Digits was also a significant direct predictor. For the control group, similarly to the dyslexic groups, verbal IQ and reading were significant direct predictors; in addition, verbal IQ, and phoneme deletion were indirect predictors. Reading was constrained to be equal across all groups, indicating that the effect of this predictor on reading comprehension was the same for all groups. Additionally, for the dyslexic groups, verbal IQ (as a direct and indirect predictor) was also constrained to be equal across the groups, indicating that its effect was the same for both groups.

Unlike the word reading and spelling models, nonverbal IQ was a significant predictor for the dyslexic groups but not the controls, suggesting that it may be an important determinant of reading comprehension attainment in dyslexic adults but not in adults without dyslexia. The model accounted for a high proportion of the variance in reading comprehension for all groups: controls $R^2 = .70$, student dyslexics $R^2 = .64$, and non-student dyslexics $R^2 = .65$. The standardized beta coefficients indicated that the importance of the significant predictors varied across the groups. For the control group, the coefficient scores indicated that both verbal IQ $\beta = .50$ and reading $\beta = .48$ were of similar importance as direct predictors, accounting for a similar amount of the variance in reading comprehension. Likewise, verbal IQ $\beta = .38$ and phoneme deletion $\beta = .38$ were of similar importance as indirect predictors. For the student dyslexic group, reading $\beta = .47$ was the most important direct predictor accounting for most of the variance in reading comprehension, followed by verbal IQ β = .37 and nonverbal IQ β = .25, while phoneme deletion β = .57 was the most important indirect predictor and then verbal IQ β = .41. For the non-student dyslexic group, reading β = .31, verbal IQ β = .29, and nonverbal IQ β = .30 were of similar importance as direct predictors, accounting for a similar amounts of the variance in reading comprehension, and RAN Digits carried the least importance β = .21, while verbal IQ was the most important indirect predictor β = .48 and then phoneme deletion β = .34.



.

Figure 7.6. Path model of reading comprehension for control group with unstandardized path weights and correlations. Red line represents variable constrained across all groups. WRIT M = WRIT Matrices; WRIT V = WRIT Vocabulary; RAN D = RAN Digits.

* p <.05, ** p < .01.

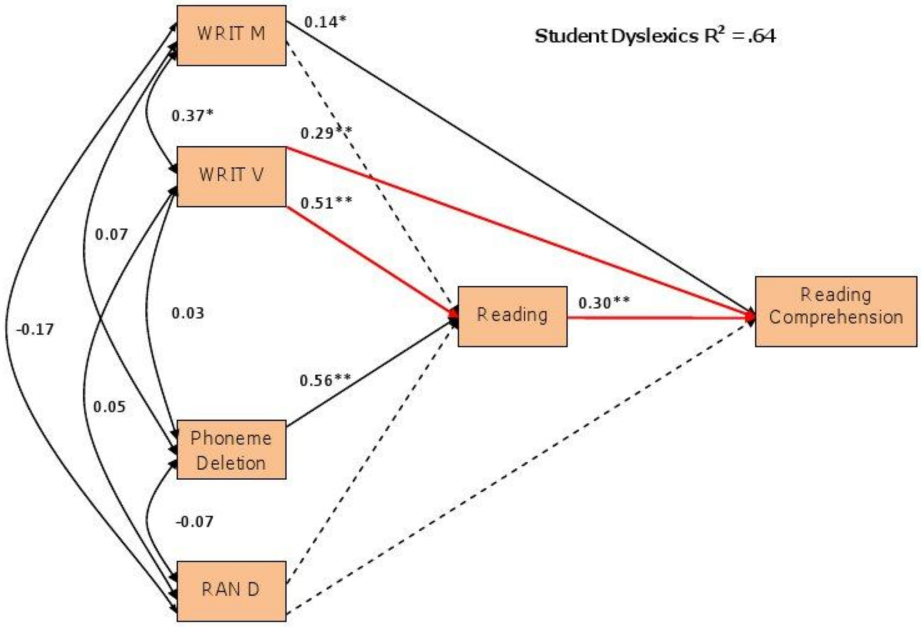


Figure 7.7. Path model of reading comprehension for student dyslexic with unstandardized path weights and correlations. Red lines represent variable constrained across all groups and dyslexic groups only. WRIT M = WRIT Matrices; WRIT V = WRIT Vocabulary; RAN D = RAN Digits.

** p <.05, ** p < .01.*

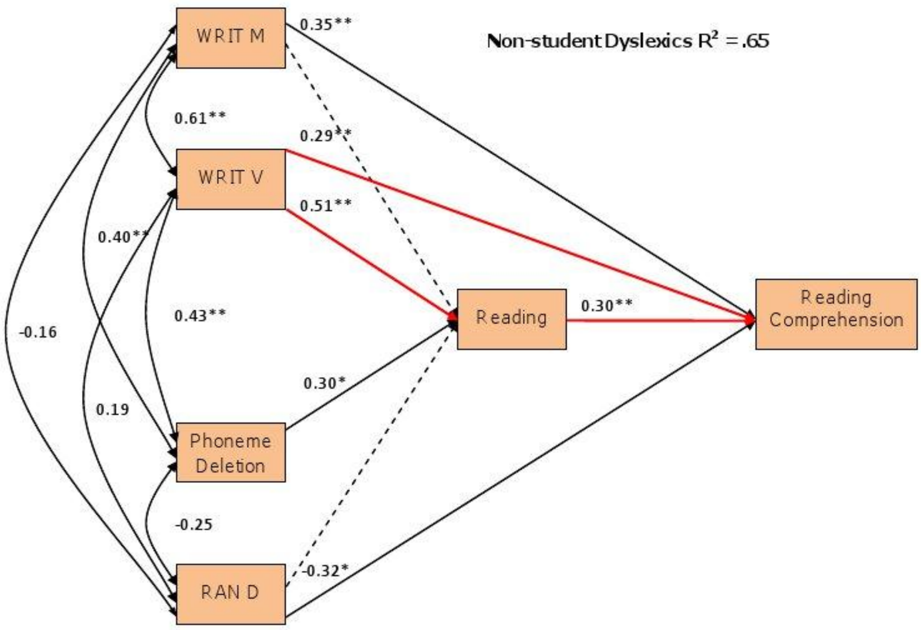


Figure 7.8. Path model of reading comprehension for non-student dyslexic with unstandardized path weights and correlations. Red lines represents variable constrained across all groups and dyslexic groups only. WRIT M = WRIT Matrices; WRIT V = WRIT Vocabulary; RAN D = RAN Digits.

** p <.05, ** p < .01.*

To summarise the results of the multi-group path analyses, the profiles of cognitive predictors of literacy skills were the same or very similar for the dyslexic groups and there were also similarities between the profiles of predictors for the dyslexic groups and the controls. For the dyslexic groups, their profile of cognitive predictors was the same for word reading and spelling, while reading comprehension differed slightly. For word reading, not only were the cognitive predictors the same, but also the relative effects of these predictors did not differ. For spelling, despite similar predictor profiles, there were differences in the relative effects of some of the predictors. Importantly, the effect of verbal IQ was the same. For reading comprehension, RAN Digits was a significant direct predictor for the non-student dyslexic group only. In addition, although nonverbal IQ was a significant direct predictor for both groups, its effect differed and was greater for the non-student dyslexics. Importantly, as with reading and spelling, the effect of verbal IQ for both groups was the same.

The main differences in the profile of cognitive predictors for the dyslexic and control groups were: (1) the number of cognitive predictors and (2) the significance and effect of verbal and nonverbal IQ. For word reading and reading comprehension, the dyslexic groups' profiles included more predictors than the control group, suggesting they required more cognitive resources in order to perform these tasks. Verbal IQ was a significant direct and/or indirect predictor of all the literacy skills for the dyslexic groups, while for the control group, this was true for spelling and reading comprehension, but not word reading. In general, verbal IQ had a greater effect (as indicated by the higher Beta scores) on the literacy skills of the dyslexic groups than the controls, although for spelling, verbal IQ was a significant direct and indirect predictor for the control group and only a direct predictor for the dyslexic group. Nonverbal IQ was not a significant predictor of any literacy skill for the control group, but was a direct predictor of reading comprehension for the dyslexic groups. Overall, therefore, verbal and nonverbal IQ were more important to the cognitive profile of the dyslexic groups than to that of the controls.

7.3 Discussion

The main purpose of this study was to investigate the profiles of cognitive predictors of literacy skills among adults (students and non-students) with and without dyslexia. We were particularly interested in the similarities and/or differences in the cognitive predictors of the two groups of adults with dyslexia and the effects of verbal and nonverbal IQ. The groups differed in age; however, it is unlikely that this affected the results, as age was not significantly correlated with the other variables for any of the groups. The student participants (controls and student dyslexics) differ from the non-students in their level of education: all the students were educated up to university level while most (except two) of the non-students up to secondary level. We compared the groups on standardised measures of verbal and nonverbal IQ and literacy skills (word reading, spelling, and reading comprehension) as well as on tasks that are known predictors of these skills, phoneme deletion and RAN Digits. We then used multigroup path analysis to identify the profiles of cognitive predictors of literacy skills, and the similarities/differences of these across the groups as well as the similarities/differences of their effects.

The comparison of the groups' performances on all the measures revealed that the student participants (controls and non-student dyslexics) were comparable on verbal and nonverbal IQ, and reading comprehension, but differed on the measures of word reading, spelling, phoneme deletion, and RAN. Additionally, the student dyslexics performed in the average range on the standardised measures of verbal and nonverbal IQ, and literacy, suggesting that they may be considered compensated or high functioning dyslexics. This is not unusual for this population (university students) and their cognitive profile is line with the results of other studies with similar populations (Hatcher et al., 2002; Kemp et al., 2009; Trainin & Swanson, 2003). Conversely, the performances of the controls differed significantly from those of the non-student dyslexics on all the measures, with controls performing better (more accurate and quicker). Similarly, with the exception of RAN Digits, the performances of the student dyslexics differed significantly from those of the non-student dyslexics on all the other measures with the students performing better (more accurate and quicker). To summarise, the controls performed better than both dyslexic groups on the literacy and literacy related measures, while the student dyslexics performed better than the non-student dyslexics. Also, the student groups (controls and student dyslexics) differed from the non-students on the verbal and nonverbal IQ measures. This difference between the two groups of dyslexics allowed us specifically to consider the effects of general ability (IQ) on their literacy skills.

7.3.1 Cognitive Predictors of Literacy Skills in Adults with and without Dyslexia

One of the objectives of this study was to identify the cognitive predictors of literacy skills in adults with and without dyslexia. Using multigroup path analyses, we developed models for the cognitive predictors for three groups of adults: student dyslexics, non-student dyslexics and student controls. The results indicated that all the models provided very good fits for the data with excellent model indices and explained large proportions of the variances in the literacy scores of the participants. The profiles of cognitive predictors of literacy skills were generally similar for the dyslexic and control groups. In addition, the cognitive predictors were similar to those reported by previous studies with children and adults. For word reading, the significant predictors across the groups were phoneme awareness and spelling which are similar to predictors of reading identified by previous studies with children (Caravolas et al., 2013; Caravolas et al., 2005; Moll et al., 2014; Muter et al., 2004; Oakhill & Cain, 2012; Parrila et al., 2004; Vellutino, Tunmer, Jaccard, & Chen, 2007) and adults with literacy difficulties (Durgunoğlu & Öney, 2002; Greenberg et al., 1997; Thompkins & Binder, 2003). In addition to phoneme awareness and spelling, verbal IQ was also a significant predictor of word reading for the dyslexic groups but not the controls. Our results are consistent with those of Caravolas et al. (2012), who found that verbal IQ was not a unique predictor of reading efficiency or spelling in normally developing children in English and several other languages. The results are also consistent with those of Swanson and Hsieh (2009) who found the verbal IQ moderated the magnitude of the deficit exhibited by dyslexic adults on a range of measures. The results suggest that for word reading, the dyslexic groups were using more cognitive resources that the controls to accomplish the task. A possible explanation for this might be that, for the controls, their abilities in word recognition were sufficient for them to accomplish the task of reading without relying on other cognitive resources. Conversely, the dyslexic groups, whose word reading skills are not as well developed, needed additional cognitive resources – especially their word knowledge -- to compensate for their deficit and to accomplish the task. Nonverbal IQ was not a significant predictor of word reading for any of the groups, and this suggests that it may not be an important determinant of single word reading attainment in adults.

For spelling, the predictors included in the models were verbal and nonverbal IQ, phoneme awareness, RAN, and reading (as direct predictor and mediator). The significant predictors across the groups were verbal IQ, phoneme awareness, RAN and reading. Our findings are consistent with those reported by other studies with children (Caravolas et al., 2005; Moll et al., 2014) and adults (Durgunoğlu & Öney, 2002). Although the sets of predictors were the same for all groups, for the control group verbal IQ was a direct and indirect (through reading) predictor, but for the dyslexic groups it was only an indirect predictor. This suggests that the control group was utilising vocabulary knowledge more than the dyslexics, which may contribute to their better performance on the task. Thus, the development of spellings skills, more so than reading skills, may depend on word knowledge, and broader verbal skills, even when a ‘normal’ level of spelling competence has been reached. Similar to the models for reading, nonverbal IQ was not a significant predictor for any of the groups, suggesting that it may not be an important determinant of spelling attainment in adults.

For reading comprehension, the predictors included in the models were verbal and nonverbal IQ, phoneme awareness, RAN, and reading (as direct predictor and mediator). The significant predictors across the groups were verbal IQ, phoneme awareness, RAN, and reading. Similar predictors of reading comprehension were identified in research with children (Caravolas et al., 2005; Oakhill & Cain, 2012) adults with low literacy (Mellard et al., 2010) and adults with dyslexia (Ransby & Swanson, 2003). In addition, nonverbal IQ was also a significant predictor of reading comprehension for the dyslexic groups but not the controls, and RAN was a significant predictor for the non-student dyslexics but not the other groups. As with the results for the reading model, these results suggest that the dyslexic group required more cognitive resources to accomplish the task than the controls. It is likely therefore, that for the controls (as with word reading), their abilities in reading comprehension (as well as the other predictors) were sufficient for them to accomplish the task without relying on non-verbal cognitive recourses. Conversely, the dyslexic groups with literacy deficits at the word level as well as in the other literacy-related predictors required additional cognitive resources to extract meaning from text. The student dyslexics were comparable to the controls on reading comprehension, and this gives us more evidence (in the path models) that they attain their performance by relying on a broader set of abilities, possibly because their word reading skills are not automated.

Despite the fact that the models for all the literacy measures fitted the data, and accounted for a large proportions of the variance in all the groups, the proportion of variance explained for the non-student dyslexic group, reading R^2 = .51 and spelling R^2 = .47 was relatively lower than that for the student controls R^2 = .65, R^2 = .70 and student dyslexics R^2 = .71, R^2 = .74 respectively. This suggests that there may be other variables important for this group that were not included in the model. Although it is difficult to say definitively what these variables may be, research indicates that socio-economic factors such as education and income may affect educational attainment (Hecht, Burgess, Torgesen, Wagner, & Rashotte 2000; Kieffer, 2012; Sastry & Pebley, 2010). The educational level of the non-students who were mostly educated up to the secondary level, differed from the student groups. Therefore, the inclusion of educational level as a predictor in the model might have increased variance explained for this group. However, this was not possible in our analyses, as the data collected for the participants on education were categorical and not continuous, and all but two non-students fell into the one category of ‘secondary education attainment’.

To summarize, the profile of cognitive predictors of literacy skills in adults with and without dyslexia was similar and consistent with predictors identified for both children and adults in existing research studies. In addition, there were more similarities between the profiles of cognitive predictors for the two dyslexic groups, than between these groups and the controls. Overall, the dyslexic groups tended to require more cognitive resources to accomplish the literacy tasks than the controls.

7.3.2 Comparison of the Profile of Cognitive Predictors of Literacy Skills for the Dyslexic Groups

Another question the study sought to address was whether the profiles of cognitive predictors of literacy differ in adults with dyslexia with differences in IQ (and educational attainment, as it turns out) level. The dyslexic groups differed in their verbal and nonverbal abilities. The results of the path analyses indicated that, despise this difference in their IQ level, the cognitive predictors for both groups were the same for word reading and spelling, and differed slightly for reading comprehension. For word reading, not only were the cognitive predictors for both groups the same, but also the relative weightings of each predictor, and could be constrained to be equal across the groups. For spelling, the results were similar to reading, with both groups having the same significant cognitive predictors two of which were constrained to be equal indicating that the effects of these variables were the same for both groups. The greatest difference in the cognitive profiles of the groups was for reading comprehension, nevertheless there were still more similarities than differences between them. For both groups, verbal and nonverbal IQ, phoneme awareness, and reading were significant predictors. Additionally, the effects of word reading and verbal IQ were the same. The main difference in the cognitive profile of the groups was that RAN was an additional significant predictor for the non-student dyslexics but not the students.

These results are consistent with the findings of the majority of similar studies of children with IQ discrepant and IQ consistent reading disabilities (Stanovich & Siegel, 1994; Stuebing et al., 2002; Vellutino, Fletcher, Snowling, & Scanlon, 2004), which found that the cognitive profiles of the children were very similar regardless of IQ level. Although the participants with dyslexia in this study were assessed based on the IQ discrepancy definition of dyslexia, there were significant differences in the IQ levels of the two groups. Given the difference in the groups' verbal and nonverbal IQ, the results suggest that the profile of cognitive (literacy and related) predictors of literacy skills for adults with dyslexia may be the same regardless of IQ level.

7.3.3 The Effects of Verbal and Nonverbal IQ on the Literacy Skills of Adults with Dyslexia with Difference in IQ

Finally, we also wanted to determine if the effects of IQ on the literacy skills differed for adults with dyslexia with difference in IQ levels. The results of the path analyses indicated that the effects of IQ were very similar for both groups. Nonverbal IQ was not a significant predictor of word reading and spelling for the groups. This is an important finding given the significant differences in the performance of the groups on the measures and suggests that nonverbal IQ may not be important for assessing attainment in reading and spelling for adults with dyslexia. Nonverbal IQ was a significant predictor of reading comprehension for both groups; however its effect varied across the groups, student dyslexic $\beta = 0.14$ and non-student dyslexic $\beta = 0.35$. The reliance on their nonverbal abilities (and not only language skills) for reading comprehension (and not for reading and spelling) may be a reflection of the more complex nature of the task (García & Cain, 2014). This also suggests that nonverbal abilities may be more relevant to reading comprehension than to word level literacy skills, at least for adults with reading disabilities. Also, if we applied the simple view of reading theory to the data, the reading comprehension ability of the dyslexic groups would be affected by their poorer decoding skills. For the student dyslexics, although their decoding skills were lower than the controls, they were adequate as indicated by their reading scores on the standardised reading measures. The fact that the dyslexic students performance on the reading comprehension task was comparable to that of the controls, may be attributed to the adequacy of their decoding skills along with the use of additional cognitive resources, that is, nonverbal IQ. For the non-student dyslexic group, decoding skills were below average and therefore although they also used additional cognitive resource to accomplish the reading comprehension task, they could not achieve the same results as the student dyslexics. Verbal IQ was a significant predictor of all the literacy skills for both dyslexic groups and importantly it was constrained to be equal across the groups, indicating that its effect was the same for both groups. Overall, therefore, the results indicated that despite the difference in their IQ level, the effect of IQ on their literacy skills was essentially the same for both dyslexic groups.

The findings of this study concur with that of Swanson and Hsieh (2009), which found that verbal IQ influenced the outcomes of performance of adults with reading disability on literacy measures. However, our findings contradict their support for the use of cut-off scores in IQ for determining reading disability. In our study, although the dyslexics in both the student and non-student groups were selected on the basis of the discrepancy definition of dyslexia, the IQ levels of the groups differed. Despite this fact however, the profiles of cognitive predictors of literacy skills for both groups were the same or very similar. More importantly, the effect of verbal IQ on the literacy skills for both groups was the same. This suggests that the cognitive profiles of adults with reading disabilities with and without a discrepancy between their IQ and literacy attainment may be the same. Our results therefore suggest that, similar to the findings of many studies with children, an IQ, literacy attainment discrepancy may not be required for identifying adults with dyslexia.

7.3.4 Conclusion

To our knowledge, this is the first study that has directly compared the cognitive predictors of literacy skills in adults with and without dyslexia. We obtained empirical evidence that the cognitive predictors of literacy skills in adults with and without dyslexia are similar to those identified for children. Additionally, our results suggest that the profile of cognitive predictors of literacy skills for adults with dyslexia may be the same regardless of IQ level. These results add to the growing research evidence that indicate that an IQ attainment discrepancy may not be required for identifying individuals with dyslexia, although general cognitive abilities probably play a more important compensatory role as literacy deficits become more severe.

Chapter 8: General Discussion

8.1 Summary of Results

This thesis focused on dyslexia in adulthood, its effective identification and its manifestations in student and non-student populations. Overall, compared to research with children with dyslexia, research with adults is limited and in some areas nonexistent. This restricts our understanding of the disorder in adults, and our ability to assist in the mitigation of the effects of the disorder, and to provide appropriate intervention and accommodation. This is likely to have negative effects on not only the adults affected by the disorder individually, but also society as a whole. This body of research included a critical review of current dyslexia screening tests currently available for use with adults, the evaluation of two existing dyslexia screening tests for adults, the development of a new test, and the examination of the profiles of cognitive predictors of literacy skills of adults with and without dyslexia. In the critical review of current dyslexia screening tests in Chapter 2, we assessed current research information available on six measures currently in use in the United Kingdom. This revealed that, generally, there is paucity of research information, and empirical evidence of the capacity of some of the tests was weak. Studies 1 and 3 evaluated the psychometric properties of the Bangor Dyslexia Test (BDT) and the Instines, respectively. Studies 2a and 2b reported the preliminary research undertaken in the development of the Bangor Adult Literacy Index. Finally, in Study 4 we examined the profiles of cognitive predictors of literacy skills for adults with dyslexia with differences in general ability levels.

In Study 1, we evaluated the psychometric properties of the BDT, a paper based dyslexia screening test suitable for use with individuals aged 7 years to adulthood. The results of the evaluation of the test highlighted strengths as well as a few weaknesses. There was good support for the validity of the test, as scores on its subtests, as well as on the task as a whole, discriminated between adults without and without dyslexia. Additionally, the accuracy with which it identified adults with dyslexia, as indicated by its sensitivity rate, was excellent, and above rates reported for all the other screening tests reviewed in this thesis. However, the accuracy of its identification of adults without dyslexia, that is, its specificity rate, was below the minimum acceptable. Consequently, a large number of adults (17.5% false positives) without dyslexia would incorrectly be identified as having the disorder. Further evidence of the construct validity of the test was provided by the correlations of scores on the test with standardised measures of literacy. Scores on the BDT were significantly negatively correlated with moderate to large correlations to scores on the standardised measures of literacy. As expected, high scores on the BDT were associated with low scores on the measures of literacy. The reliability of the test was acceptable, coefficient $\alpha = .69$; but below the level considered ideal. The study also provided empirical evidence that the Months Forward subtest of the BDT is not sufficiently sensitive for detecting dyslexia in adults. The dyslexic group performed at ceiling on the subtest and the vast majority (88.1%) obtained a negative score (no indication of dyslexia). Furthermore, the reliability analysis indicated that it was not contributing to the reliability of the BDT. Overall, however, the results indicate that the psychometric properties of the BDT are adequate, and it can effectively identify adults at risk of dyslexia.

In Study 3, we evaluated the psychometric properties of the Instines, a computerised dyslexia test designed for use with adolescents (aged 12 years and older) and adults. The results provided little support for the validity or the reliability of the measure, and suggest that the Instines is not appropriate for identifying adults with dyslexia. Only half of its eight subtests were capable of discriminating between adults with and without dyslexia. The impact of this on the measure as a whole could not be fully assessed, as the criteria used to determine an individual's risk of dyslexia are not available. Nevertheless, discriminant analysis indicated that it met the sensitivity threshold (80%) for accuracy in the identification of adults with dyslexia. Importantly, however, the analysis also indicated that three of the eight subtests were not contributing to the discriminatory capacity of the test. Also, although acceptable, this rate is lower than rates reported for most of the other screening tests reviewed. In addition, the specificity rate at which it accurately identified adults without dyslexia is below the minimum acceptable level, and would result in a large number of (13.2%) false positives. As with its sensitivity rate, this is also below that reported for other screening tests. Users of the Instines would therefore be disadvantaged compared to users of other screening tests. Further, the construct validity of two of its subtests (Verbal Reasoning and Comprehension) was not supported as they had low to moderate correlations with the standardised literacy measures, suggesting that they may not be valid measures. Most importantly, the results of the test-retest reliability analysis indicated that the Instines might not be a reliable test as most of its subtests (five) were well below the acceptable level of reliability. The inadequacy of the test was demonstrated by the comparison of the screening outcomes of the participants on the two administrations of the test. A very large percentage (43%) of the participants received different classifications on the two administrations of the test with extremely low sensitivity (55%) and specificity (57.9%) rates on the first administration of the test, which only improved slightly on the second. Additionally, information on the criteria used to identify individuals at risk at dyslexia on the subtests and on the test as whole is not available. Users of the test therefore have very little insight into how its screening outcomes are derived. Also, although this was not an issue for our sample because the test is self administered it does require that the person being screened is able to read or listen (most instructions are also given verbally) and follow the instructions correctly. The practical implications of using the Instines will be discussed in the section dedicated to this topic.

In studies 2a and 2b, we evaluated the psychometric properties of several tasks selected for inclusion in a cost effective, quick, and effective paper based screening test for dyslexia in adults named the Bangor Adult Literacy Index (BALI). The studies provided empirical evidence that tasks which assess the behavioural manifestations of dyslexia can be combined into a quick and effective dyslexia screening test, the BALI. The results of both studies supported the validity and reliability of most of the tasks developed for inclusion in the test. All the tasks created for the BALI (with the exception of the Nonword Repetition and Inhibition tasks, which will be excluded from the final battery) effectively discriminated between adults with and without dyslexia. The accuracy with which the test identified adults with and without dyslexia was good, with excellent sensitivity (91.9%) and adequate specificity rate (89.5%), and compares favourably to existing dyslexia screening tests for adults. Furthermore, the subtests correlated significantly with moderate to high correlations with the standardised measures of literacy providing evidence of their construct validity. Importantly the internal consistency and/or test-retest reliability of the subtests were excellent and were at or well above the minimum required. The psychometric properties of the BALI are on a par with or stronger than some others, for example the Instines, QuickScan, and aspects of the YAA-R. This is despite the fact two of the tasks created for inclusion (Nonword Repetition and Inhibition) were not psychometrically sound and the battery was therefore reduced from eight to six tasks. Like the YAA-R, however, the BALI awaits external evaluation.

In study 4, we examined the profile of cognitive predictors of literacy skills of adults with and without dyslexia, including two groups of dyslexics with differences in IQ level. We also examined the effects of verbal and nonverbal IQ on the literacy skills of the groups. The study was designed to increase our understanding of the cognitive predictors of literacy skills of adults with and without dyslexia. Additionally, it sought to obtain empirical evidence of the validity of the discrepancy definition of dyslexia for the identification of adults with dyslexia. The study provided empirical evidence that the cognitive predictors of literacy skills of adults with and without dyslexia are similar and are consistent with predictors of literacy for children (Caravolas et al., 2013; Caravolas et al., 2005; Moll et al., 2014; Muter et al., 2004; Parrila et al., 2004). Additionally, and more importantly, the profile of cognitive predictors of literacy skills of groups of adults with dyslexia who differed in terms of their general abilities (verbal and nonverbal IQ) was very similar, and there were greater similarities between the dyslexic groups than between them and typical readers. These results concur with a growing number of studies that have reported similar findings with children (Fletcher et al., 1992; Fletcher et al., 1994; Francis et al., 2005; Stage et al., 2003; Stuebing et al., 2009; Stuebing et al., 2002). However, the results of Study 4 were contrary to the findings of the meta-analysis conducted by Swanson and Hsieh (2009) which provided support for the use of an IQ discrepancy definition of dyslexia for adults. These researchers found that, compared to their counterparts without dyslexia, adults with dyslexia with a discrepancy between their IQ and literacy skills exhibited greater deficits (as indicated by larger effect sizes) than adults with dyslexia whose IQ was consistent with their literacy skills. The findings of our study that: (1) the profile of cognitive predictors of literacy skills of adults with dyslexia with differences in IQ level are the same or very similar, and (2) the effects of both verbal and nonverbal IQ on their literacy skills are the same or very similar, do not support the use of a discrepancy definition for the identification of dyslexia in adults. The study also provided evidence that nonverbal IQ is not a significant predictor of word reading and spelling abilities in adults with or without dyslexia. However, for adults with dyslexia, nonverbal IQ was a significant predictor of reading comprehension abilities.

8.2 Implications of Using Screening Tests with Inadequate Psychometric Properties

Currently, the identification of dyslexia involves a two stage process: firstly, screening to identity individuals most at risk, and secondly assessment for diagnosis. The screening process is considered a necessary and cost effective way of identifying individuals most at risk of the disorder. Once identified individuals at risk of the disorder are usually then referred for a full assessment by an educational psychologist or other qualified professional. Positive assessment of dyslexia usually results in access to appropriate resources to assist the individual to cope with the disability and to fulfill their potential. Screening is therefore an important first part of this process and its outcome can have important consequences for the individual screened. There is general agreement that early identification of individuals with dyslexia is important as this facilitates early invention and/or accommodations, which results in better outcomes (Reid, 2011; Rose Review, 2009). For children, the Rose Review (2009) recommended the monitoring of their progress by teachers as the best method for identifying those at risk of dyslexia. For adults, however, a screening process is probably required. This is because adults in academia, in the workplace, or in the general community, do not normally enjoy the benefit of long-term oversight by a trained professional teacher; their tutors or line managers may have neither the expertise not the time for such oversight. Thus, the screening process with adults usually includes a number of different procedures such as a semi-structured interview, assessment of literacy and other cognitive skills, and a dyslexia screening test. There is currently no gold standard for dyslexia screening tests and therefore the psychometric properties of these tests are an important and objective way of accessing their suitability. Independent research studies that evaluate the psychometric properties of these screening tests are essential to facilitate this evaluation and further test development.

No screening test can achieve 100% accuracy in its classifications, as there is usually a tradeoff between sensitivity and specificity, such that an increase in one will result in a decrease in the other (Grimes & Schulz, 2002). However, it is important that the sensitivity and specificity rates of the tests are at least at the minimum recommended to ensure effective identification and avoid the adverse consequences associated with incorrect classifications (false positive and false negative). These consequences for people with dyslexia may include: (1) no access to resources that might assist in the mitigation of their difficulties; (2) the inability to fulfill their potential, particularly in educational attainment, employment, and income (McLaughlin, Speirs, & Shenassa, 2014; Vogel & Holt, 2003); and (3) increased risk of emotional and mental health disorders such as anxiety and depression (Carroll & Iles, 2006; Undheim & Sund, 2008; Wilson, Armstrong, Furrie, & Walcot, 2009). For individuals without dyslexia who are incorrectly identified as dyslexic, although there may not be any adverse personal consequences, this may result in the inappropriate allocation of scarce resources and may even limit or deprive those in need of these resources. Furthermore, information about the psychometric properties of available screening tests allows users to compare the tests and select the most effective for their context and needs, and thus minimise incorrect classification and associated adverse consequences.

The evaluation of the BDT and the Instines produced contrasting results and highlighted the need for high quality independent empirical research of the psychometric properties of screening tests to ensure that they are fit for purpose. The studies also highlight the need for research undertaken to be comprehensive and to examine both validity and reliability as both are essential characteristics of any good screening tool and are required to make appropriate evaluation of the suitability of the test. As illustrated by the results of the study of the Instines, a test may have some capacity to discriminate between individuals with and without dyslexia, but be incapable of doing this reliably, which ultimately affects its validity.

8.3 Screening for Dyslexia in Adults

If tests developed to screen for dyslexia in adults are to be effective it is important that the abilities assessed and the tasks included are capable of distinguishing between adults with and without the disorder. The two studies conducted with the BALI illustrated the validity of tasks included in it and provided support for the use of tasks assessing literacy and related sub-skills for the identification of adults with dyslexia. In developing the BALI to ensure its effectiveness, most of the tasks selected for inclusion had strong and consistent research evidence of their relevance for discriminating between adults with and without dyslexia. Given its theoretical framework and the existing research evidence for the discriminant powers of the tasks included, the effectiveness of the measure is not surprising. It does however highlight the need for good theoretical framework and strong research evidence in the development of screening tests.

The tasks included in the BALI also highlight the importance of assessing speed, not only accuracy, on the abilities being assessed by the test. Previous research studies have also produced similar findings (Re et al., 2011). The importance of speed is illustrated quite clearly by the performance of the participants on the Phoneme Deletion tasks for which participants were scored on both accuracy and speed. Although, the groups' performances differed significantly on both measures, the magnitude of the difference for speed (d = 1.77) was much greater than for accuracy (d = 0.74). It seems, therefore, that having a timed element to the task is more likely to elicit the existence of any deficit, as the dyslexics may require more time to complete the task in order to mitigate/compensate for their difficulties (Re et al., 2011). This might be especially true for compensated dyslexic adults like those in our studies with university students, as these individuals may be able, whether through remediation or compensation, to perform in the average range even on tasks which they find difficult, once there are no time restrictions. Our results support the use of extra time as an accommodation for adults with dyslexia in further education and in the work place. Also, the results of the logistic regression indicated that Phoneme Deletion Speed was the most discriminating tasks. This result also suggest that tasks assessing phonological awareness may be more sensitive than other measures of phonological processing (such as phonological short term memory and naming speed) for detecting dyslexic difficulties in adults, especially those who are well compensated. Our results are consistent with those of other studies that have also found that phoneme awareness tasks are the most effective tasks for detecting dyslexic difficulties in adults (Snowling et al., 1997).

Given the limited number of dyslexia screening tests for adults that are currently available, and the inadequacies highlighted by our research, there is a need for psychometrically sound and effective screening tests. The research undertaken to date with the BALI suggests that it is has the characteristics of a good screening test and would therefore be a useful addition to the current tests available.

8.4 Cognitive Profiles of Adults with Dyslexia

Two issues that need to be resolved in the current debate on the IQ discrepancy definition of dyslexia are whether the cognitive profiles of individuals with and without an IQ discrepancy are different, and, what exactly is the role of IQ in a dyslexia diagnosis? The results of our study suggest that the profiles of cognitive predictors among dyslexic adults with relatively higher or lower IQ, as estimated by the two proxy measures of WRIT Matrices and WRIT Vocabulary, are very similar. Thus, relative levels of general cognitive ability seem not to be associated with differences in the cognitive skills/building blocks that underlie literacy abilities. Likewise, the effect of IQ itself on the literacy attainments of both groups is similar, and nonverbal IQ was not found to be a significant predictor of word-level reading or spelling abilities in dyslexic and non-dyslexic adults. Our results are consistent with similar research with children (Fletcher et al., 1994; Stuebing et al., 2002) and do not support the IQ discrepancy definition of dyslexia. The similarities between the groups suggest that general cognitive abilities are simply not important causes of individual or group differences in word reading and spelling development. Our results therefore suggest that an evaluation of an individual's IQ should not be a determinant of a classification of dyslexia in adults, although of course it may be an important and informative measure in a full educational psychologist assessment. The elimination of the IQ discrepancy from the definition of dyslexia should result in more individuals with reading difficulties, regardless of their IQ, being able to access resources that may assist in mitigating their difficulties and reducing its adverse effects. This should have a positive impact not only on the life of the individual and their family but also on society as a whole.

In all the studies conducted in the thesis, the comparison of the performances of the dyslexic and control groups on the various measures administered revealed that the differences between the groups were quantitative and not qualitative. Thus, the pattern of the groups' performances on the measures was similar; however the controls performed better (more accurate and/or quicker). Our results support the view that dyslexia is dimensional disorder that exists at the lower end of a normal continuum of reading abilities and is not a qualitatively distinct disorder. Similar results have been reported for children with dyslexia (Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992).

8.5 Future Research

In general, more research that evaluates dyslexia screening tests by examining their psychometric properties is required to provide independent assessments of these tests. Also, research studies that compare the effectiveness of more than one screening test with the same population would advance the field and facilitate evaluation that is more objective. In addition, as most studies are conducted with students, research with non-student adults is also required to determine the effectiveness of the screening tests with the broader population.

For the BDT, the battery that can be said to comprise the most 'distal' set of diagnostic indicators of dyslexia of the existing screening tools, additional means of evaluation are called for. This is due to the heterogeneity and alleged subjectivity of the scoring of its subtests, which may have negatively affected its internal consistency. Therefore, methods such as test re-test analyses to assess its reliability would be useful. An assessment of inter-rater reliability is also desirable. Unlike the other tests, administrators of the BDT are required to make some clinical judgments in determining the score, on most of its subtests, of the person being assessed and this may affect the overall screening outcome. Eight of the ten subtest scores require this clinical judgment as they are determined not only by the accuracy of the response but also the difficulties encountered or explicit strategies used by the person being assessed to achieving it. This may result in variations in scoring and screening outcomes. An assessment of inter-rater reliability would determine if the measure is consistent across administrators. It would also indirectly assess the quality of the administrative instructions, particularly the interpretation of the scores, which was described as inadequate by Sutherland and Smith (1991). Finally, the BDT manual does not specify a cut off score for determining an individuals' at risk status, and our study suggests that this could improve its specificity rate. More research is required to determine the most appropriate cut off score. For the Instines, as ours first empirical study of its psychometric properties, it may be useful for further research to verify our results. However, until the makers of the Instines test make more information available about the construction of its component measures, the method of scoring performance, and the computation of the norms, further research may be a futile undertaking. Additionally, our study could only assess the construct validity of three of its eight subtests; therefore, an assessment of the others would be required.

For the BALI, further research on our part is required to establish norms for each of the subtests and to ascertain the best method of determining an individuals' at risk status on the test as a whole. Research with a non-student adult population is also required to ensure that the battery is suitable for this population. Although the measure effectively discriminates between adults with and without dyslexia, research to identify other tasks (particularly ones that are not directly literacy related) that may be included to further improve its sensitivity and specificity would be useful. Further research is also needed to verify the results of our study which examined the profile of cognitive predictors of literacy skills in adults. In our study, although the dyslexic groups differed in IQ level both consisted of individuals with discrepancies between their IQ and literacy skills. Further research is therefore required with groups of dyslexic adults with IQ discrepant and IQ consistent profiles to verify our findings. Also, intervention studies with dyslexic adults with IQ discrepant and IQ consistent profiles which investigate their outcomes are needed, to further elucidate the importance of IQ in the definition and assessment of dyslexia. Finally, as our study included only IQ and literacy related predictors, more research with cognitive and other predictors (such as age, education, socio-economic status) is required, to identify other variables that may be important in the cognitive profile of adults with dyslexia.

8.6 Conclusion

The studies reported in this thesis evaluated dyslexia screening tests for adults and examined the profiles of cognitive predictors of literacy skills. They provided empirical evidence of the relative capacity of the BDT and Instines to identify adults at risk of dyslexia and the appropriateness or otherwise of their psychometric properties. The development of the BALI should facilitate screening for dyslexia in adults by providing a psychometrically sound, quick, easy to administer, and well researched, screening test. The examination of the cognitive profiles of adults with dyslexia enhances our understanding of dyslexia in adults and contributes to improvements in screening, assessment, and remediation of the disorder.

References

Anastasi, A., & Urbina, S. (1997). *Psychological Testing*. (7th ed.). Upper Saddle River, N.J.: Prentice Hall.

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review, 105*, 678-723. doi: 10.1037/0033-295X.105.4.678-723

Badian, N. A. (1994). Do dyslexic and other poor readers differ in reading-related cognitive skills? *Reading and Writing: An Interdisciplinary Journal, 6,* 45–63.

Baker, S. F., & Ireland, J. L. (2007). The link between dyslexic traits, executive functioning, impulsivity and social self-esteem among an offender and non-offender sample. *International Journal of Law and Psychiatry, 30,* 492-503. doi:10.1016/j.ijlp.2007.09.010

Banich, M. T. (2009). Executive function: The search for an integrated account. *Current Directions in Psychological Science, 18,* 89-94. doi:10.1111/j.1467-8721.2009.01615.x

Barbosa, T., Miranda, M. C., Santos, R. F., & Bueno, O. F. (2009). Phonological working memory, phonological awareness and language in literacy difficulties in Brazilian children. *Reading and Writing: An Interdisciplinary Journal, 22*, 201-218. doi:10.1007/s11145-007-9109-3

Barth, A. E., Denton, C. A., Stuebing, K. K., Fletcher, J. M. Cirino, P. T., Francis, D. J., & Vaughn, S. (2010). A test of the cerebellar hypothesis of dyslexia in adequate and inadequate responders to reading intervention. *Journal of the International Neuropsychological Society, 16,* 526-536. doi: 10.1017/S1355617710000135

Bartlett, D., & Moody, S. (2000). *Dyslexia in the workplace*. London: Whurr.

Beaton, A., McDougall, S., & Singleton, C. (1997). Humpty Dumpty grows up? Diagnosing dyslexia in adulthood. *Journal of Research in Reading, 20,* 1-6.

Bishop, D., Bishop, S. J., Bright, P., James, C., Delaney, T., & Tallal, P. (1999). Different origin of auditory and phonological processing problems in children with language impairment: Evidence from a twin study. *Journal of Speech, Language & Hearing Research, 42*, 155-168.

Bosse, M., & Valdois, S. (2009). Influence of the visual attention span on child reading performance: A cross-sectional study. *Journal of Research in Reading, 32*, 230-253. doi:10.1111/j.1467-9817.2008.01387.x

Bosse, M., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition, 104*, 198-230. doi:http://dx.doi.org/10.1016/j.cognition.2006.05.009

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities, 40*, 226-243. doi: 10.1177/00222194070400030401

Brosnan, M., Demetre, J., Hamill, S., Robson, K., Shepherd, H., & Cody, G. (2002). Executive functioning in adults and children with developmental dyslexia. *Neuropsychologia, 40*, 2144-2155. doi:10.1016/S0028-3932(02)00046-5

Bruck, M. (1990). Word-recognition skills of adults with childhood diagnoses of dyslexia. *Developmental Psychology, 26*, 439-454. doi:10.1037/0012-1649.26.3.439

Bruck, M. (1992). Persistence of dyslexic’s phonological awareness deficits. *Developmental Psychology, 28*, 874- 886.

Bruck, M. (1993). Word recognition and component phonological processing skills of adults with childhood diagnosis of dyslexia. *Developmental Review, 13*, 258-268. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1006/drev.1993.1012

Bruck, M., & Waters, G. (1990). Effects of reading skill on component spelling skills. *Applied Psycholinguistics, 11*, 425-437. doi:10.1017/S0142716400009668

Byrne, B., Coventry, W. L., Olson, R. K., Samuelsson, S., Corley, R., Willcutt, E. G., . . . DeFries, J. C. (2009). Genetic and environmental influences on aspects of literacy and language in early childhood: Continuity and change from preschool to Grade 2. *Journal of Neurolinguistics, 22*, 219-236.

Callens, M., Tops, W., & Brysbaert, M. (2012). Cognitive profile of students who enter higher education with an indication of dyslexia. *Plos One, 7*, 1-14. doi:10.1371/journal.pone.0038081

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language, 45*, 751-774. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1006/jmla.2000.2785

Caravolas, M., Lervåg, A., Defior, S., Malková, G., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science, 24,* 1398-1407. doi: 10.1177/0956797612473122

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., . . . Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies *Psychological Science, 23*, 678-686. doi:10.1177/0956797611434536

Caravolas, M., Volín, J., & Hulme, C. (2005). Phoneme awareness is a key component of alphabetic literacy skills in consistent and inconsistent orthographies: Evidence from Czech and English children. *Journal of Experimental Child Psychology, 92*, 107-139.

Caron, C., & Rutter, M. (1991). Comorbidity in child psychopathology: Concepts, issues and research strategies. *Journal of Child Psychology and Psychiatry, 32,* 1063-1080. doi:10.1111/j.1469-7610.1991.tb00350.x

Carroll, J. M., & Iles, J. E. (2006). An assessment of anxiety levels in dyslexic students in higher education. *British Journal of Educational Psychology, 76*, 651–662.

Cataldo, S., & Ellis, N. (1988). Interactions in the development of spelling, reading and phonological skills. *Journal of Research in Reading, 11*, 86-109. doi:10.1111/j.1467-9817.1988.tb00153.x

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research, 48*, 1378-1396. doi:10.1044/1092-4388(2005/096)

Catts, H. W., Gillispie, M., Leonard, L. B., Kail, R. V., & Miller, C. A. (2002). The role of speed of processing, rapid naming, and phonological awareness in reading achievement. *Journal of Learning Disabilities, 35*, 509-524.

Chester, M. D. (2005). Making valid and consistent inferences about school effectiveness from multiple measures. *Educational Measurement: Issues and Practice, 24*, 40-52. doi:10.1111/j.1745-3992.2005.00022.x

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating Children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science, 21*, 1106-1116. doi:10.1177/0956797610375449

Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological testing and assessment: An introduction to tests and measurement* (6th Ed.). New York: McGraw-Hill.

Démonet, J., Taylor, M. J., & Chaix, Y. (2004) Developmental dyslexia. *Lancet, 363*, 1451-1460.

Denckla, M. B., & Cutting, L. E. (1999). History and significance of rapid automatized naming. *Annals of Dyslexia, 49*, 29-42.

Denckla, M. B., & Rudel, R. (1976). Rapid 'automatized' naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14,* 471-479. doi:10.1371/journal.pone.0038081

Durgunoğlu, A. Y., & Öney, B. (2002). Phonological awareness in literacy acquisition: It's not only for children. *Scientific Studies of Reading, 6,* 245-266. doi: 10.1207/S1532799XSSR0603_3

Dyslexia Foundation retrieved from: http://www.dyslexia-check.com/instines01.htm

Elbro, C., Nielsen, I., & Petersen, D. K. (1994). Dyslexia in adults: Evidence for deficits in nonword reading and in the phonological representation of lexical items. *Annals of Dyslexia, 44*, 203-226.

Evans, M. I., Galen, R. S., & Britt, D. W. (2005). Principles of Screening. *Seminars in Perinatology, 29*, 364-366. doi:10.1053/j.semperi.2005.12.002

Everatt, J. (1997). The abilities and disabilities associated with adult developmental dyslexia. *Journal of Research in Reading, 20*, 13-21. doi:10.1111/1467-9817.00016

Farmer, M. E., & Klein, R. M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic Bulletin & Review, 2*, 460-493.

Fawcett, A. J., & Nicolson, R. I. (1999). Performance of dyslexic children on cerebellar and cognitive tests. *Journal of Motor Behavior, 31*, 68-78.

Fawcett, A.J., & Nicolson, R. I. (1998). *The dyslexia adult screening test*. London: The Psychological Corporation

Fawcett, A. J., Nicolson, R. L., & Dean, P. (1996). Impaired performance of children with dyslexia on a range of cerebellar tasks. *Annals of Dyslexia, 46*, 259-283.

Fawcett, A. J., Singleton, C. H., & Peer, L. (1998). Advances in Early Years Screening for Dyslexia In the United Kingdom. *Annals of Dyslexia, 48,* 57-88.

Felton, R. H., Naylor, C. E., & Wood, F. B. (1990). Neuropsychological profile of adult dyslexics. *Brain and Language, 39*, 485-497. doi:10.1016/0093-934X(90)90157-C

Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.

Finch, A. J., Nicolson, R. I., & Fawcett, A. J. (2002). Evidence for a neuroanatomical difference within the olivo-cerebellar pathway of adults with dyslexia. *Cortex, 38*(4), 529-539. doi:10.1016/S0010-9452(08)70021-2

Fisher, S. E., Francks, C., Marlow, A. J., MacPhie, I. L., Newbury, D. F., Cardon, L. R., . . . Monaco, A. P. (2002). Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nature Genetics*, *30*(1), 86-91. doi: 10.1038/ng792

Fletcher, J. M., Francis, D. J., Rourke, B. P., Shaywitz, S. E., & Shaywitz, B. A. (1992). The validity of discrepancy-based definition of reading disabilities. *Journal of Learning Disabilities, 25*, 555–561.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Fowler, A., Francis, D. J., Stuebing, K. K., & Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86,* 6–23.

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2005). Psychometric approaches to the identification of LD: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities, 38,* 98-108.

Friberg, J. C. (2010). Consideration for Test Selection: How do Validity and Reliability Impact Diagnostic Decisions. *Child Language Teaching and Therapy, 22*, 77-92.

Frith, U. (1999). Paradoxes in the definition of dyslexia. *Dyslexia: An International Journal of Research and Practice, 5*, 192-214. doi:10.1002/(SICI)1099-0909(199912)

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, *84*, 77-111. doi:10.3102/0034654313499616

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology, 93*, 265-281. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/j.jecp.2005.08.003

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*, 336-360.

Georgiou, G. K., Parrila, R., & Kirby, J. (2006). Rapid naming speed components and early reading acquisition. *Scientific Studies of Reading, 10*, 199-220. doi:10.1207/s1532799xssr1002_4

Georgiou, G. K., Parrila, R., & Liao, C. (2008). Rapid naming speed and reading across languages that vary in orthographic consistency. *Reading and Writing: An Interdisciplinary Journal, 21*, 885-903. doi:10.1007/s11145-007-9096-4

Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology, 100*, 566-580. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1037/0022-0663.100.3.566

Glascoe, F. P. & Byrne, K. E. (1993) ‘The accuracy of three developmental screening tests.’ *Journal of Early Intervention*, 17, 368–79.

Glutting, J., Adams, W., & Shelow, D. (2000). *Wide Range Intelligence Test*. Lutz, FL: Psychological Assessment Resources.

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 10911-10916.

Gottardo, A., Siegel, L. S., & Stanovich, K. E. (1997). The assessment of adults with reading disabilities: What can we learn from experimental tasks? *Journal of Research in Reading, 20*, 42-54. doi:10.1111/1467-9817.00019

Gough, P., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6-10.

Grady, C. L. (2008). Cognitive neuroscience of aging. *Annals of the New York Academy of Sciences, 1124*, 127-144. doi:10.1196/annals.1440.009

Gray, S. (2003). Diagnostic accuracy and test–retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders, 36*, 129-151. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/S0021-9924(03)00003-0

Greenberg, D., Ehri, L. C., & Perin, D. (1997). Are word-reading processes the same or different in adult literacy students and third-fifth graders matched for reading level? *Journal of Educational Psychology, 89,* 262-275.

Griffiths, Y. M., & Snowling, M. J. (2002). Predictors of exception word and nonword reading in dyslexic children: The severity hypothesis. *Journal of Educational Psychology, 94*, 34-43. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1037/0022-0663.94.1.34

Grimes, D. A., & Schulz, K. F. (2002). Uses and abuses of screening tests. The Lancet, 359, 881-884.

Gross-Glenn, K., Jallad, B., Novoa, L., Helgren-Lempesis, V., & Lubs, H. A. (1990). Nonsense passage reading as a diagnostic aid in the study of adult familial dyslexia. *Reading and Writing: An Interdisciplinary Journal, 2*, 161-173.

Hämäläinen, J. A., Salminen, H. K., & Leppänen, P. H. T. (2013). Basic auditory processing deficits in dyslexia: Systematic review of the behavioral and event-related potential/ field evidence. *Journal of Learning Disabilities, 46*, 413-427. doi:10.1177/0022219411436213

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,* 423-434.

Haslum, M. N., & Kiziewicz, M. (2001). Computerised assessment of students for dyslexia: a preliminary report of an evaluation of the StudyScan screening and assessment suite. In M. Kiziewicz & I. Biggs (Eds.), *CASCADE – Creativity across science, art, dyslexia, education*. Bath: University of Bath. Retrieved from http://www.bath.ac.uk/cascade/

Hatcher, J., Snowling, M. J., & Griffiths, Y. M. (2002). Cognitive assessment of dyslexic students in higher education. *British Journal of Educational Psychology, 72*, 119-133. doi:10.1348/000709902158801

Hawelka, S., & Wimmer, H. (2005). Impaired visual processing of multi-element arrays is associated with increased number of eye movements in dyslexic reading. *Vision Research, 45,* 855-863. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/j.visres.2004.10.007

Hawelka, S., & Wimmer, H. (2008). Visual target detection is not impaired in dyslexic readers. *Vision Research, 48,* 850-852. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/j.visres.2007.11.003

Hawelka, S., Huber, C., & Wimmer, H. (2006). Impaired visual processing of letter and digit strings in adult dyslexic readers. *Vision Research, 46*, 718-723. doi:http://dx.doi.org/10.1016/j.visres.2005.09.01

Hecht, S. A., Burgess, S. R., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2000). Explaining social class differences in growth of reading skills from beginning kindergarten through fourth-grade: The role of phonological awareness, rate of access, and print knowledge. *Reading and Writing: An Interdisciplinary Journal, 12*, 99-127.

Heiervang, E., Stevenson, J., Hugdahl, K., (2002). Auditory processing in children with dyslexia. *Journal of Child Psychology and Psychiatry, 43*, 931-938.

Helland, T., & Asbjørnsen, A. (2000). Executive functions in dyslexia. *Child Neuropsychology, 6*, 37-48. doi:10.1076/0929-7049(200003)6:1;1-B;FT037

Higher Education Statistics Agency (2008). Retrieved from http://www.hesa.ac.uk

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*, 127–160. doi:10.1007/BF00401799

Hulme, C., & Snowling, M. J. (2009). *Developmental disorders of language learning and cognition*. Chichester, U.K.; Malden, MA: Wiley-Blackwell.

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read: Combining intervention studies with mediation analyses. *Psychological Science, 23*, 572-577. doi: 10.1177/0956797611435921

Hulme, C., Hatcher, P., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002). Phoneme awareness is a better predictor of early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology, 82,* 2–28.

Instines. Retrieved from http://www.dyslexia-check.com/index-1.htm

Irannejad, S., & Savage, R. (2012). Is a cerebellar deficit the underlying cause of reading disabilities? *Annals of Dyslexia, 62*, 22-52. doi:10.1007/s11881-011-0060-2

James, A. (2004, March). *An investigation and comparison of computer-based screenings for dyslexia for use in FE and HE institutions.* Paper presented at the 6th International Conference of the British Dyslexia Association, Coventry: England.

Jeffries, S., & Everatt, J. (2004). Working memory: Its role in dyslexia and other specific learning difficulties. *Dyslexia, 10*, 196-214. doi:10.1002/dys.278

Judge, J., Caravolas, M., & Knox, P. C. (2006). Smooth pursuit eye movements and phonological processing in adults with dyslexia. *Cognitive Neuropsychology, 23*, 1174-1189. doi: 10.1080/02643290600785931

Kail, R., & Hall, L. K. (1994). Processing speed, naming speed, and reading. *Developmental Psychology, 30*, 949-954. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1037/0012-1649.30.6.949

Kasler, J., & Fawcett, A. (2009). Screening for learning disabilities in young adult career counseling. *Work, 32*, 201-210.

Katzir, T., Kim, Y., Wolf, M., O'Brien, B., Kennedy, B., Lovett, M., & Morris, R. (2006). Reading fluency: the whole is more than the parts. *Annals of Dyslexia 56,* 51–82.

Kemp, N., Parrila, R. K., & Kirby, J. R. (2009). Phonological and orthographic spelling in high-functioning adult dyslexics. *Dyslexia, 15*, 105-128. doi: 10.1002/dys.364

Kieffer, M. J. (2012). Before and after third grade: Longitudinal evidence for the shifting role of socioeconomic status in reading growth. *Reading and Writing: An Interdisciplinary Journal, 25*, 1725-1746. doi:10.1007/s11145-011-9339-2

Kirby, J. R., Georgiou, G. K., Martinussen, R., Parrila, R., Bowers, P., & Landerl, K. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly, 45*, 341-362.

Kline, P. (1986). *A Handbook of Test Construction Introduction to Psychometric design*. London: Methuen & Company Limited.

Kline, P. (2000). *Handbook of psychological testing* (2nd ed.). London: Routledge.

Landerl, K., & Moll, K. (2010). Comorbidity of learning disorders: Prevalence and familial transmission. *Journal of Child Psychology and Psychiatry, 51*, 287-294. doi:10.1111/j.1469-7610.2009.02164.x

Lefly, D. L., & Pennington, B. F. (1991). Spelling errors and reading fluency in compensated adult dyslexics. *Annals of Dyslexia, 41*, 143-162. doi:10.1007/BF02648083

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*, 59-80. doi:10.1348/026151003321164627

Lennox, C., & Siegel, L. S. (1993). Visual and phonological spelling errors in subtypes of children with learning disabilities. *Applied Psycholinguistics, 14*, 473-488. doi:10.1017/S0142716400010705

Lesaux, N. K., Pearson, M. R., & Siegel, L. S. (2006). The effects of timed and untimed testing conditions on the reading comprehension performance of adults with reading disabilities. *Reading and Writing, 19*, 21-48. doi:10.1007/s11145-005-4714-5

Lindgrén, S., & Laine, M. (2011). Cognitive-linguistic performances of multilingual university students suspected of dyslexia. *Dyslexia, 17*, 184-200. doi:10.1002/dys.422

Lindgrén, S., & Laine, M. (2007). The adaptation of an adult group screening test for dyslexia into Finland-Swedish: Normative data for university students and the effects of language background on test performance. *Scandinavian Journal of Psychology, 48*, 419-432. doi:10.1111/j.1467-9450.2007.00592.x

Lobier, M., Zoubrinetzky, R., & Valdois, S. (2012). The visual attention span deficit in dyslexia is visual and not verbal. *Cortex, 48,* 768-773. doi:http://dx.doi.org/10.1016/j.cortex.2011.09.003

Lucid Ability Administrator's Manual, (2011). Retrieved from http://www.lucid-research.com/t/manuals#

Lucid Adult Dyslexia Screening Administrator's Manual, (2004). Retrieved from http://www.lucid-research.com/t/manuals#

Lucid Adult Dyslexia Screening Plus Administrator's Manual (2010). Retrieved from http://www.lucid-research.com/t/manuals#

Lucid Research Limited (2010). Lucid Adult Dyslexia Screening Plus Administrator's Manual. Retrieved from http://www.lucid-research.com

Lyon, G. R. (1995). Toward a definition of dyslexia. *Annals of Dyslexia, 45*, 1-27.

Manis, F. R., Custodio, R., & Szeszulski, P. A. (1993). Development of phonological and orthographic skill: A 2-year longitudinal study of dyslexic children. *Journal of Experimental Child Psychology, 56*, 64-86. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1006/jecp.1993.1026

Manly, T., Anderson, V., Nimmo-Smith, I., Turner, A., Watson, P., & Robertson, I. H. (2001). The differential assessment of children's attention: The test of everyday attention for children (TEA-ch), normative sample and ADHD performance. *Journal of Child Psychology and Psychiatry, 42*, 1065-1081. doi:10.1111/1469-7610.00806

Marshall, C. M., Snowling, M. J., & Bailey, P. J. (2001). Rapid auditory processing and phonological ability in normal readers and readers with dyslexia. *Journal of Speech, Language & Hearing Research, 44*, 925-940.

Martin, I. West Kent College StudyScan Review. Retrieved from http://www.quickscn.com/studyscn/reviews.htm

Maughan, B., Messer, J., Collishaw, S., Pickles, A., Snowling, M., Yule, W., & Rutter, M. (2009). Persistence of literacy problems: Spelling in adolescence and at mid-life.

Journal of Child Psychology and Psychiatry, 50, 893-901. doi:10.1111/j.1469-7610.2009.02079.x

McGrath, L. M., Pennington, B. F., Shanahan, M. A., Santerre-Lemmon, L. E., Barnard, H. D., Willcutt, E. G., . . . Olson, R. K. (2011). A multiple deficit model of reading disability and attention-deficit/hyperactivity disorder: Searching for shared cognitive deficits. *Journal of Child Psychology and Psychiatry, 52*, 547-557. doi:10.1111/j.1469-7610.2010.02346.x

McLaughlin, M. J., Speirs, K. E., & Shenassa, E. D. (2014). Reading disability and adult attained education and income: Evidence from a 30-year longitudinal study of a population-based sample. *Journal of Learning Disabilities*, *47*, 374-386. doi:10.1177/0022219412458323

Melby-Lervåg, M., & Lervåg, A. (2012). Oral language skills moderate nonword repetition skills in children with dyslexia: A meta-analysis of the role of nonword repetition skills in dyslexia. *Scientific Studies of Reading*, *16*, 1-34. doi:10.1080/10888438.2010.537715

Mellard, D. F., Fall, E., & Woods, K. L. (2010). A path analysis of reading comprehension for adults with low literacy. *Journal of Learning Disabilities, 43,* 154-165. doi: 10.1177/0022219409359345

Messbauer, V. C. S., & de Jong, P. F. (2003). Word, nonword, and visual paired associate learning in dutch dyslexic children. *Journal of Experimental Child Psychology, 84*, 77-96. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/S0022-0965(02)00179-0

Miles, T. (1983). *Dyslexia: The Pattern of Difficulties*. London: Granada Publishing Limited.

Miles, T. (1993). *Dyslexia: The Pattern of Difficulties*. 2nd Ed. London: Whurr Publishers.

Miles, T. (1997). The Bangor Dyslexia Test. Cambridge: Learning Development Aids.

Miles, T. R., Wheeler, T. J., & Haslum, M. N. (1998). Gender ratio in dyslexia. *Annals of Dyslexia*, *48*, 27-55.

Miller, B., McCardle, P., & Hernandez, R., (2010). Advances and remaining challenges in adult literacy research. *Journal of Learning Disabilities, 43,* 101-107. doi: 10.1177/0022219409359341

Miller-Shaul, S. (2005). The characteristics of young and adult dyslexics readers on reading and reading related cognitive tasks as compared to normal readers. *Dyslexia, 11*, 132-151. doi:10.1002/dys.290

Mody, M., Studdert-Kennedy, M., & Brady, S. (1997). Speech perception deficits in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology*, *64*(2), 199-231. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1006/jecp.1996.2343

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., ... Landerl, K., (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction, 29,* 65-77. doi: 10.1016/j.learninstruc.2013.09.003

Moody, S. (Ed.). (2009). *Dyslexia and Employment*. Chichester, UK: Wiley-Blackwell.

Muneaux, M., Ziegler, J. C., Truc, C., Thomson, J., & Goswami, U. (2004). Deficits in beat perception and dyslexia: Evidence from French. *Neuroreport, 15*, 1255-1259. doi:10.1097/01.wnr.0000127459.31232.c4

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Developmental Psychology*, *40*, 663–681. doi: 10.1037/0012-1649.40.5.665

National Working Party on Dyslexia (1999) *Dyslexia in Higher Education: Policy, Provision and Practice,* Hull: The Psychology Department, University of Hull.

National Working Party on Dyslexia in Higher Education (1999). *Dyslexia in higher education: Policy, provision and practice*. Hull, UK: University of Hull.

Needle, J. L., Fawcett, A. J., & Nicolson, R. I. (2006). Balance and dyslexia: An investigation of adults' abilities. *European Journal of Cognitive Psychology, 18,* 909-936. doi:10.1080/09541440500412304

Nichols, S. A., Mcleod, J. S., Holder, R. L., & McLeod, H. S. T. (2009). Screening for Dyslexia, Dyspraxia and Meares-Irlen Syndrome in Higher Education. *Dyslexia, 15,* 42-60. doi:10.1002/dys.382

Nicolson, R. I., & Fawcett, A. J. (2011). Dyslexia, dysgraphia, procedural learning and the cerebellum. *Cortex, 47*, 117-127. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/j.cortex.2009.08.016

Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: Reuniting the developmental disorders? *Trends in Neurosciences, 30*, 135-141. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/j.tins.2007.02.003

Nicolson, R. I., Fawcett, A. J., Brookes, R. L., & Needle, J. (2010). Procedural learning and dyslexia. *Dyslexia, 16*, 194-212. doi:10.1002/dys.408

Nicolson, R. I., & Fawcett, A. J. (1999). Developmental dyslexia: The role of the cerebellum. *Dyslexia, 5*, 155-177.

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,* 77-83.

Nicolson, R. I., & Fawcett, A. J. (1990). Automaticity: A new framework for dyslexia research? *Cognition*, *35*, 159-182. doi:10.1016/0010-0277(90)90013-A

Nicolson, R. I., Fawcett, A. J., & Dean, P. (2001). Developmental dyslexia: The cerebellar deficit hypothesis. *Trends in Neurosciences*, *24*, 508-511. doi:10.1016/S0166-2236(00)01896-8

Nicholson, R. I., Fawcett, A. J., & Miles, T. R. (1993). *Feasibility study for the development of a computerised screening test for dyslexia in adults*. Report OL176, Department of the Employment.

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, *63*, 427-452. doi:10.1146/annurev-psych-120710-100431

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16*, 91-121. doi:10.1080/10888438.2010.529219

Ofiesh, N., Mather, N., & Russell, A. (2005). Using speeded cognitive, reading, and academic measures to determine the need for extended test time among university students with learning disabilities *Journal of Psychoeducational Assessment, 23*, 35-52. doi:10.1177/073428290502300103

Pammer, K., Lavis, R., Hansen, P., & Cornelissen, P. L. (2004). Symbol-string sensitivity and children's reading. *Brain and Language, 89*, 601-610. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/j.bandl.2004.01.009

Parrila, R., Georgiou, G., & Corkett, J. (2007). University students with a significant history of reading difficulties: What is and is not compensated? *Exceptionality Education Canada, 17*, 195-220.

Parrila, R., Kirby, J. R., & McQuarrie, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: Longitudinal predictors of early reading development? *Scientific Studies of Reading, 8*, 3-26. doi:10.1207/s1532799xssr0801_2

Pasquini, E. S., Corriveau, K. H., & Goswami, U. (2007). Auditory processing of amplitude envelope rise time in adults diagnosed with developmental dyslexia. *Scientific Studies of Reading, 11*, 259-286. doi:10.1080/10888430701344280

Paulesu, E., Frith, U., Snowling, M., Gallagher, A., Morton, J., Frackowiak, R. S. J., & Frith, C. D. (1996). Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning *Brain, 119*, 143-157. doi:10.1093/brain/119.1.143

Payne, G. H., Miles, T. R., & Wheeler, T. J. (2007). Does the DDAT programme provide a cure for dyslexia? Results from the Bangor Dyslexia Test. *Dyslexia Review, 19,* 7-12.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition, 101*, 385-413. doi:10.1016/j.cognition.2006.04.008

Pennington, B. F., & Bishop, D. V. M. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology, 60*, 283-306. doi:10.1146/annurev.psych.60.110707.163548

Pennington, B. F., Cardoso-Martins, C., Green, P. A., & Lefly, D. L. (2001). Comparing the phonological and double deficit hypotheses for developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal, 14*, 707-755

Pennington, B. F., McGrath, L. M., Rosenberg, J., Barnard, H., Smith, S. D., Willcutt, E. G., . . . Olson, R. K. (2009). Gene × environment interactions in reading disability and attention-deficit/hyperactivity disorder. *Developmental Psychology, 45*, 77-89. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1037/a0014549

Pennington, B. F., Van Orden, G. C., Smith, S. D., Green, P. A., & Haith, M. M. (1990). Phonological processing skills and deficits in adult dyslexics. *Child Development, 61*, 1753. doi:10.1111/1467-8624.ep9103040638

Peterson, R. L. & Pennington, B. (2012). Developmental dyslexia. *The Lancet*, *379*, 1997-2007.

Pico Educational Systems Limited. Retrieved from http://www.quickscn.com/website/pages/research.html

Protopapas, A., Archonti, A., & Skaloumbakas, C. (2007). Reading ability is negatively related to stroop interference. *Cognitive Psychology, 54*, 251-282. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/j.cogpsych.2006.07.003

Quercia, P., Demougeot, L., Santos, M. D., & Bonnetblanc, F. (2011). Integration of proprioceptive signals and attentional capacity during postural control are impaired but subject to improvement in dyslexic children. *Experimental Brain Research, 209*, 599-608. doi:10.1007/s00221-011-2593-3

Rack, J. (1997). Issues in the assessment of developmental dyslexia in adults: theoretical and applied perspectives. *Journal of Research in Reading, 20,* 66-76.

Rae, C., Lee, M. A., Dixon, R. M., Blamire, A. M., Thompson, C. H., Styles, P., . . . Stein, J. F. (1998). Metabolic abnormalities in developmental dyslexia detected by 1H magnetic resonance spectroscopy. *The Lancet, 351*, 1849-1852. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/S0140-6736(97)99001-2

Ramus, F. (2003). Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology, 13,* 212-228.

Ramus, F., Pidgeon, E., & Frith, U. (2003a). The relationship between motor control and phonology in dyslexic children. *Journal of Child Psychology and Psychiatry, 44*, 712-722. doi:10.1111/1469-7610.00157

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003b). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain, 126,* 841-865. doi:10.1093/brain/awg076

Ransby, M. J., & Swanson, H. L. (2003). Reading comprehension skills of young adults with childhood diagnoses of dyslexia. *Journal of Learning Disabilities, 36*, 538-555.

Re, A. M., Tressoldi, P. E., Cornoldi, C., & Lucangeli, D. (2011). Which tasks best discriminate between dyslexic university students and controls in a transparent language? *Dyslexia, 17*, 227-241. doi:10.1002/dys.431

Reid, G. (2011). *Dyslexia* (3rd ed.). London: Continuum International Publishing.

Reid, G., & Kirk, J. (2001). *Dyslexia in adults: Education and employment*. Chichester; New York: John Wiley.

Reid, A. A., Szczerbinski, M., Iskierka-Kasperek, E., & Hansen, P. (2007). Cognitive profiles of adult developmental dyslexics: Theoretical implications. *Dyslexia, 13*, 1-24. doi:10.1002/dys.321

Reiter, A., Tucha, O., & Lange, K. W. (2005). Executive functions in children with dyslexia. *Dyslexia, 11*, 116-131. doi:10.1002/dys.289

Rice, M., & Brooks, G. (2004). *Developmental dyslexia in adults: a research review*. National Research and Development Centre for Adult Literacy and Numeracy.

Richardson, J. T. E., & Wydell, T. N. (2003). The representation and attainment of students with dyslexia in UK higher. *Reading and Writing: An Interdisciplinary Journal, 16,* 475-503.

Richardson, U., Thomson, J. M., Scott, S. K., & Goswami, U. (2004). Auditory processing skills and phonological representation in dyslexic children. *Dyslexia, 10*, 215-233. doi:10.1002/dys.276

Richlan, R., Kronbichler, M., & Wimmer, H. (2009). Functional abnormalities in the dyslexic brain: A quantitative meta-analysis of neuroimaging studies. *Human Brain Mapping, 30,* 3299-3308. doi:10.1002/hbm.20752

Rochelle, K. S. H., & Talcott, J. B. (2006). Impaired balance in developmental dyslexia? A meta-analysis of the contending evidence. *Journal of Child Psychology and Psychiatry, 47,* 1159-1166. doi:10.1111/j.1469-7610.2006.01641.x

Romani, C., Olson, A., Di Betta, A. M. (2005). Spelling Disorders. In: M. J. Snowling & C. Hulme, (Eds.), *The science of reading: A handbook* (pp. 431-447). Malden, MA: Blackwell publishing Ltd.

Rose Review. (2009). *Identifying and teaching children and young people with dyslexia and literacy difficulties.* Retrieved from http://publications.dcsf.gov.uk/eOrderingDownload/00659-2009DOM-EN.pdf

Rosen, S., & Manganari, E. (2001). Is there a relationship between speech and nonspeech auditory processing in children with dyslexia? *Journal of Speech, Language & Hearing Research, 44*, 720-736.

Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry, 16*, 181-197. doi:10.1111/j.1469-7610.1975.tb01269.x

Sanderson, A. (2000). Reflections on StudyScan. *Dyslexia, 6,* 284-290.

Santoro, L. E., Coyne, M. D., & Simmons, D. C. (2006). The reading-spelling connection: Developing and evaluating a beginning spelling intervention for children at risk of reading disability. *Learning Disabilities Research & Practice, 21,* 122-133. doi:10.1111/j.1540-5826.2006.00212.x

Sastry, N., & Pebley, A. R. (2010). Family and neighborhood sources of socioeconomic inequality in children's achievement. *Demography, 47*, 777-800.

Savage, R. S., Frederickson, N., Goodwin, R., Patni, U., Smith, N., & Tuersley, L. (2005). Relationships among rapid digit naming, phonological processing, motor automaticity, and speech perception in poor, average, and good readers and spellers. *Journal of Learning Disabilities, 38*, 12-28.

Savage, R., & Carless, S. (2005). Phoneme manipulation not onset-rime manipulation ability is a unique predictor of early reading. Journal of Child Psychology and Psychiatry, 46, 1297-1308. doi: 10.1111/j.1469-7610.2005.01436.x

Savage, R., Pillay, V., & Melidona, S. (2007). Deconstructing rapid automatized naming: Component processes and the prediction of reading difficulties. *Learning and Individual Differences, 17*, 129-146. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/j.lindif.2007.04.001

Shaywitz, S., Fletcher, J., Holahan, J., Shneider, A., Marchione, K., Stuebing, K. K., ... Shaywitz, B. A. (1999). Persistence of dyslexia: The Connecticut longitudinal study at adolescence. *Pediatrics, 104,* 1351–1359.

Shaywitz, S. E., Shaywitz, B. A., Pugh, K. R., Fulbright, R. K., Constable, R. T., W. Einar Mencl, W. E., ... Gore, J. C. (1998). Functional disruption in the organization of the brain for reading in dyslexia. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 2636-2641.

Shaywitz, B.A., Shaywitz, S. E., Pugh, K. R. Mencl, W. E., Fulbright, R. K., Skudlarski, P., ... Gore, J. C. (2002). Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biological Psychiatry, 52*, 101-110.

Shovman, M. M., & Ahissar, M. (2006). Isolating the impact of visual perception on dyslexics' reading ability. *Vision Research, 46*, 3514-3525. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/j.visres.2006.05.011

Siegel, L. S. (1999). Issues in the definition and diagnosis of learning disabilities: A perspective on Guckenberger v. Boston university. *Journal of Learning Disabilities, 32*, 304-319.

Singleton, C. (2004). Using computer-based assessment to identify learning problems. In L. Florian & J. Hegarty (Eds), *ICT and Special Educational Needs*, pp. 46-63. Milton Keynes, Uk: Open University Press.

Singeton, C., Horne, J., & Simmons, F. (2009). Computerised screening for dyslexia in adults. *Journal of Research in Reading, 32*, 137-152.

Singleton, C. H., Horne, J. K., Thomas, K. V., & Leedale, R.C. (2004). Lucid Adult Dyslexia Screening. East Yorkshire: Lucid Research Limited.

Singleton, C., & No To Failure (2009). *Intervention for Dyslexia A review of published evidence on the impact of specialist dyslexia teaching*. Retrieved from http:// thedyslexia-spldtrust.org.uk

Singleton, C., Horne, J., & Simmons, F. (2009). Computerised screening for dyslexia in adults. *Journal of Research in Reading, 32*, 137-152. doi:10.1111/j.1467-9817.2008.01386.x

Singleton, C., Thomas, K., & Horne, J. (2000). Computer-based cognitive assessment and the development of reading. *Journal of Research in Reading, 23*, 158-180. doi:10.1111/1467-9817.00112

Skottun, B. C. (2000). The magnocellular deficit theory of dyslexia: The evidence from contrast sensitivity. *Vision Research, 40*, 111-127. doi:10.1016/S0042-6989(99)00170-4

Snowling, M. J. (1995). Phonological processing and developmental dyslexia. (1995). *Journal of Research in Reading, 18*, 132-138.

Snowling, M., (1998). Dyslexia as a phonological deficit: Evidence and implication. *Child Psychology & Psychiatry Review, 3*, 4-11.

Snowling, M. J. (2001). From language to reading and dyslexia. *Dyslexia, 7*, 37-46. doi:10.1002/dys.185

Snowling, M. J. (2012). Changing concepts of dyslexia: Nature, treatment and comorbidity. *Journal of Child Psychology and Psychiatry, 53*, 1-3. doi:10.1111/j.1469-7610.2009.02197.x

Snowling, M., Dawes, P., Nash, H., & Hulme, C. (2012). Validity of a protocol for adult self-report of dyslexia and related difficulties. *Dyslexia, 18*, 1-15. doi:10.1002/dys.1432

Snowling, M. J., Goulandris, N., & Defty, N. (1996). A longitudinal study of reading development in dyslexic children. *Journal of Educational Psychology, 88*, 653-669. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1037/0022-0663.88.4.653

Snowling, M., Muter, V., & Carroll, J. (2007). Children at family risk of dyslexia: a follow-up in early adolescence. *Journal of Child Psychology and Psychiatry, 48*, 609-618.

Snowling, M., 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*, 31-41. doi:10.1111/1467-9817.00018

Stage, S. A., Abbott, R. D., Jenkins, J. R., & Berninger, V. W. (2003). Predicting response to early reading intervention from verbal IQ, reading-related language abilities, attention ratings, and verbal IQ-word reading discrepancy: Failure to validate discrepancy method. *Journal of Learning Disabilities, 36*, 24-33.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*, 24-53. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1037/0022-0663.86.1.24

Stein, J. (1993). Dyslexia-impaired temporal information processing? *Annals of the New York Academy of Sciences, 682*, 83-86. doi:10.1111/j.1749-6632.1993.tb22961.x

Stein, J. (2001). The magnocellular theory of developmental dyslexia. *Dyslexia, 7,* 12-36. doi:10.1002/dys.186

Stein, J., & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. *Trends in Neurosciences, 20*, 147-152. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/S0166-2236(96)01005-3

Stevenson, J., Langley, K., Pay, H., Payton, A., Worthington, J., Ollier, W., & Thapar, A. (2005). Attention deficit hyperactivity disorder with reading disabilities: Preliminary genetic findings on the involvement of the ADRA2A gene. *Journal of Child Psychology and Psychiatry, 46*, 1081-1088. doi:10.1111/j.1469-7610.2005.01533.x

Stoodley, C. J., Fawcett, A. J., Nicolson, R. I., & Stein, J. F. (2006). Balancing and pointing tasks in dyslexic and control adults. *Dyslexia, 12*, 276-288. doi:10.1002/dys.326

Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment, 80*, 217-222.

StudyScan Suite. Retrieved from http://www.quickscan.com

Stuebing, K. K., Barth, A. E., Molfese, P. J., Weiss, B., & Fletcher, J. M. (2009). IQ is not strongly related to response to reading instruction: A meta-analytic interpretation. *Exceptional Children, 76,* 31-51.

Stuebing, K. K., Fletcher, J. M., LeDoux, J. M., Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2002). Validity of IQ discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal, 39,* 469–518. doi: 10.3102/00028312039002469

Sucena, A., Castro, S. L., & Seymour, P. (2009). Developmental dyslexia in an orthography of intermediate depth: The case of European Portuguese. *Reading and Writing: An Interdisciplinary Journal, 22*, 791-810. doi:10.1007/s11145-008-9156-4

Sun, Y., Lee, J., & Kirby, R., (2010). Brain imaging findings in dyslexia. *Pediatrics & Noenatology, 51*, 89-96.

Sutherland, M. J., & Smith, C. D. (1991). Assessing Literacy Problems in Mainstream Schooling: a critique of three literacy screening tests. *Educational Review, 43,* 39-48. doi.org/10.1080/0013191910430104

Svensson, I., & Jacobson, C. (2006). How persistent are phonological difficulties? A longitudinal study of reading retarded children. *Dyslexia*, *12*, 3-20. doi:10.1002/dys.296

Swanson, H. L. & Hsieh, C. (2009). Reading disabilities in adults: A selective meta-analysis of the literature. *Review of Educational Research, 79*, 1362-1390. doi:10.3102/0034654309350931

Swanson, H. L., Trainin, G., Necoechea, D. M., & Hammill, D. D. (2003). Rapid naming, phonological awareness, and reading: A meta-analysis of the correlation evidence. *Review of Educational Research, 73*, 407-440.

Szenkovits, G., & Ramus, F. (2005). Exploring dyslexics' phonological deficit I: Lexical vs sub-lexical and input vs output processes. *Dyslexia, 11*, 253-268. doi:10.1002/dys.308

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson/Allyn & Bacon.

Talcott, J. B., Hansen, P. C., Assoku, E. L., & Stein, J. F. (2000). Visual motion sensitivity in dyslexia: Evidence for temporal and energy integration deficits. *Neuropsychologia, 38*, 935-943. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/S0028-3932(00)00020-8

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language, 9*, 182-198.

Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiological basis of speech: A case for the preeminence of temporal processing. *Annals of the New York Academy of Sciences, 682*, 27-47. doi:10.1111/j.1749-6632.1993.tb22957.x

Tallal, P., & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia*, *13*, 69-74. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/0028-3932(75)90049-4

Tallal, P., & Piercy, M. (1973). Developmental aphasia: Impaired rate of non-verbal processing as a function of sensory modality. *Neuropsychologia, 11*, 389-398. doi:http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/0028-3932(73)90025-0

Taylor, H. G., Lean, D., & Schwartz, S. (1989). Pseudoword repetition ability in learning-disabled children. *Applied Psycholinguistics, 10*, 203-219. doi:10.1017/S0142716400008523

Taymans, J. M., Swanson, H. L., Schwarz, R. L., Gregg, N., Hock, M., Gerber, P. J. (2009). *Learning to Achieve: A Review of the Research Literature on Serving Adults with Learning Disabilities*. National Institute for Literacy, Learning to Achieve: A Review of the Research Literature on Serving Adults with Learning Disabilities, Washington, DC 20008

Teare, P. (2003, April 30). Re: Screening Children [Online forum comment]. Retrieved from https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=dyslexia;a26b7ce5.03

Temple, E. (2002). Brain mechanisms in normal and dyslexic readers. *Current Opinion in Neurobiology, 12,* 178-183.

Thompkins, A. C., & Binder, K. S. (2003). A comparison of the factors affecting reading performance of functionally illiterate adults and children matched by reading level. *Reading Research Quarterly, 38,* 236-258.

Tønnessen, F. E. (1997). How can we best define ‘dyslexia’? *Dyslexia*, *3*, 78-92. doi:10.1002/(SICI)1099-0909(199706)3:2<78::AID-DYS71>3.0.CO;2-2

Trainin, G.,H.Lee. (2005). Cognition, metacognition, and achievement of college students with learning disabilities. *Learning Disability Quarterly, 28*, 261-272.

Tunick, R. A., & Pennington, B. F. (2002). The etiological relationship between reading disability and phonological disorder. *Annals of Dyslexia, 52*, 75-97.

Turner, M. (2008). *Psychological assessment of dyslexia*. Hoboken: John Wiley & Sons, Ltd.

Undheim, A. M., & Sund, A. M. (2008). Psychosocial factors and reading difficulties: Students with reading difficulties drawn from a representative population sample. *Scandinavian Journal of Psychology, 49*, 377-384.

Valdois, S., Bosse, M., & Tainturier, M. (2004). The cognitive deficits responsible for developmental dyslexia: Review of evidence for a selective visual attentional disorder. *Dyslexia, 10*, 339-363. doi:10.1002/dys.284

Valdois, S., Bosse, M., Ans, B., Carbonnel1, S., Zorman, M., David, D., & Pellat, J. (2003). Phonological and visual processing deficits can dissociate in developmental dyslexia: Evidence from two case studies. *Reading and Writing: An Interdisciplinary Journal, 16*, 541-572.

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

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Eduactional Psychology, 88*, 601-638.

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading, 11*, 3-32. doi:10.1080/10888430709336632

Vogel, S. A., & Holt, J. K. (2003). A comparative study of adults with and without self-reported learning disabilities in six English-speaking populations: What have we learned? *Dyslexia, 9*, 193-228. doi:10.1002/dys.244

Viholainen, H., Aro, M., Ahonen, T., Crawford, S., Cantell, M., & Kooistra, L. (2011). Are balance problems connected to reading speed or the familial risk of dyslexia? *Developmental Medicine & Child Neurology, 53*, 350-353. doi:10.1111/j.1469-8749.2010.03856.x

Vinegrad, M. (1994). Adult Dyslexia Checklist. Reading, United Kingdom: British Dyslexia Association.

Vukovic, R. K., Wilson, A. M., & Nash, K. K. (2004). Naming speed deficits in adults with reading disabilities: A test of the double-deficit hypothesis. *Journal of Learning Disabilities, 37*, 440-450.

Waber, D. P., Weiler, M. D., Wolff, P. H., Bellinger, D., Marcus, D. J., Ariel, R., . . . Wypij, D. (2001). Processing of rapid auditory stimuli in school-age children referred for evaluation of learning disorders. *Child Development, 72,* 37-49. doi:10.1111/1467-8624.00264

Wagner, R. K. (2005). Understanding genetic and environmental influences on the development of reading: Reaching for higher fruit. *Scientific Studies of Reading, 9,* 317-326. doi:10.1207/s1532799xssr0903_7

Wagner, R. K., Torgesen, J. K., Rashotte, C. A. (1999a). *Comprehensive Test of Phonological Processing* (*CTOPP*). Austin, TX: Pro-Ed

Wagner, R. K., Torgesen, J. K., Rashotte, C. A. (1999b). *Test of Word Reading Efficiency (TOWRE).* Austin, TX: Pro-Ed

Warmington, M., Stothard, S. E., & Snowling, M. (2013). Assessing dyslexia in higher education: The York Adult Assessment Battery-Revised. *Journal of Research in Special Educational Needs, 13*(1), 48-56. doi:10.1111/j.1471-3802.2012.01264.x

Warmington, M., Stothard, S. E., & Snowling, M. J. York Adult Assessment Battery-Revised (YAA-R) Retrieved from http://www.york.ac.uk/media/psychology/teachingandmarking/York%20Adult%20Assessment%20Battery%20Revised%20Form%20and%20Test.pdf

Wechsler, D. (1993). *Wechsler Objective Reading Dimensions (WORD)*. London: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-Third Edition (WAIS-III)*. London: The Psychological Corporation.

Wechsler, D. (2005). *Wechsler Individual Achievement Test-Second Edition (WIAT-II).* London: The Psychological Corporation.

Weyandt, L. L., Rice, J. A., Linterman, I., Mitzlaff, L., & Emert, E. (1998). Neuropsychological performance of a sample of adults with ADHD, developmental reading disorder, and controls. *Developmental Neuropsychology, 14*(4), 643-656. doi:10.1080/87565649809540734

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*, 237-255. doi:10.1111/j.1467-7687.2006.00483.x

Wilkinson, G. S., & Robertson, G. J. (2004). *Wide Range Achievement Test* (4th Ed.). Lutz. FL: Psychological Assessment Resources.

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*, 1345-1361. doi:http://dx.doi.org/10.1016/j.cortex.2010.06.009

Wilson, A. M., Armstrong, C. D., Furrie, A., & Walcot, E. (2009). The mental health of Canadians with self-reported learning disabilities. *Journal of Learning Disabilities, 42,* 24-40. doi:10.1177/0022219408326216

Wilson, A. M., & Lesaux, N. K. (2001). Persistence of phonological processing deficits in college students with dyslexia who have age-appropriate reading skills. *Journal of Learning Disabilities, 34,* 394-400. doi:10.1177/002221940103400501

Wimmer, H., Mayringer, H., & Landerl, K. (1998). Poor reading: A deficit in skill-automatization or a phonological deficit? *Scientific Studies of Reading, 2*, 321-340. doi:10.1207/s1532799xssr0204_2

Wimmer, H., Mayringer, H., & Raberger, T. (1999). Reading and dual-task balancing: Evidence against the automatization deficit explanation of developmental dyslexia. *Journal of Learning Disabilities, 32*, 473-478. doi:10.1177/002221949903200513

Wolff, P. (2002). Timing precision and rhythm in developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal, 15,* 179-206.

Wolf, M., & Denckla, M. B. (2005). *Rapid Automatized Naming and Rapid Alternating Stimulus Tests.* Austin, TX: Pro-Ed

Wolff, U., & Lundberg, I. (2003). A technique for group screening of dyslexia among adults. *Annals of Dyslexia, 53,* 324-339.

Zdzienski, D. (1998). Dyslexia in higher education: An exploratory study of learning support, screening and diagnostic assessment (Doctoral thesis, University of Leicester, Leicester, United Kingdom). Retrieved from http://ethos.bl.uk

Ziegler, J. C., Pech-Georgel, C., Dufau, S., & Grainger, J. (2010). Rapid processing of letters, digits and symbols: What purely visual-attentional deficit in developmental dyslexia? *Developmental Science, 13*, F8-F14. doi:10.1111/j.1467-7687.2010.00983.x

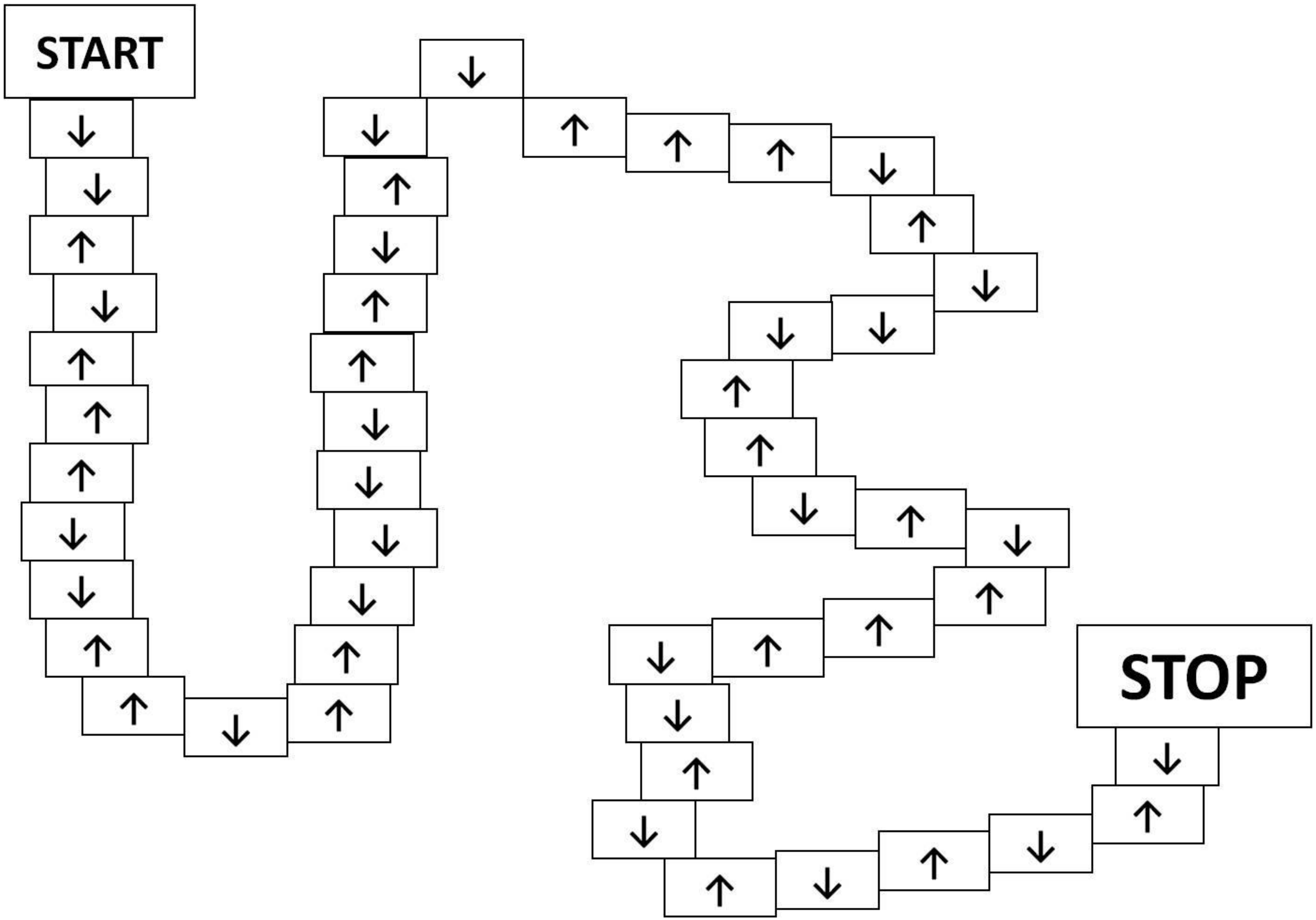
Appendix A: 1 Minute Reading Task for BALI Pilot Study

from	should	centre	Children	limited	quality
said	place	design	Number	reaction	continued
that	through	value	Party	following	industry
were	school	doctor	Question	committee	president
they	city	future	Woman	department	national
some	body	human	Entire	general	performance
have	even	letter	Service	provided	temperature
with	only	really	Without	however	property
time	upon	other	Different	material	society
more	very	money	Trouble	together	available
once	idea	effect	Working	beginning	literature
this	many	hotel	Century	attitude	development
help	into	level	Radio	hospital	original
kind	about	result	Family	experience	particular
less	father	series	Similar	religion	individual
must	under	figure	Agreement	important	population
know	always	little	Secretary	consider	understanding
when	table	higher	Officer	business	professional
great	better	matter	Anyone	government	situation
right	degree	never	Company	involved	information
since	power	system	Position	equipment	administration
there	after	nature	History	character	opportunity
which	union	office	Average	decision	university
point	island	second	Direction	production	responsibility

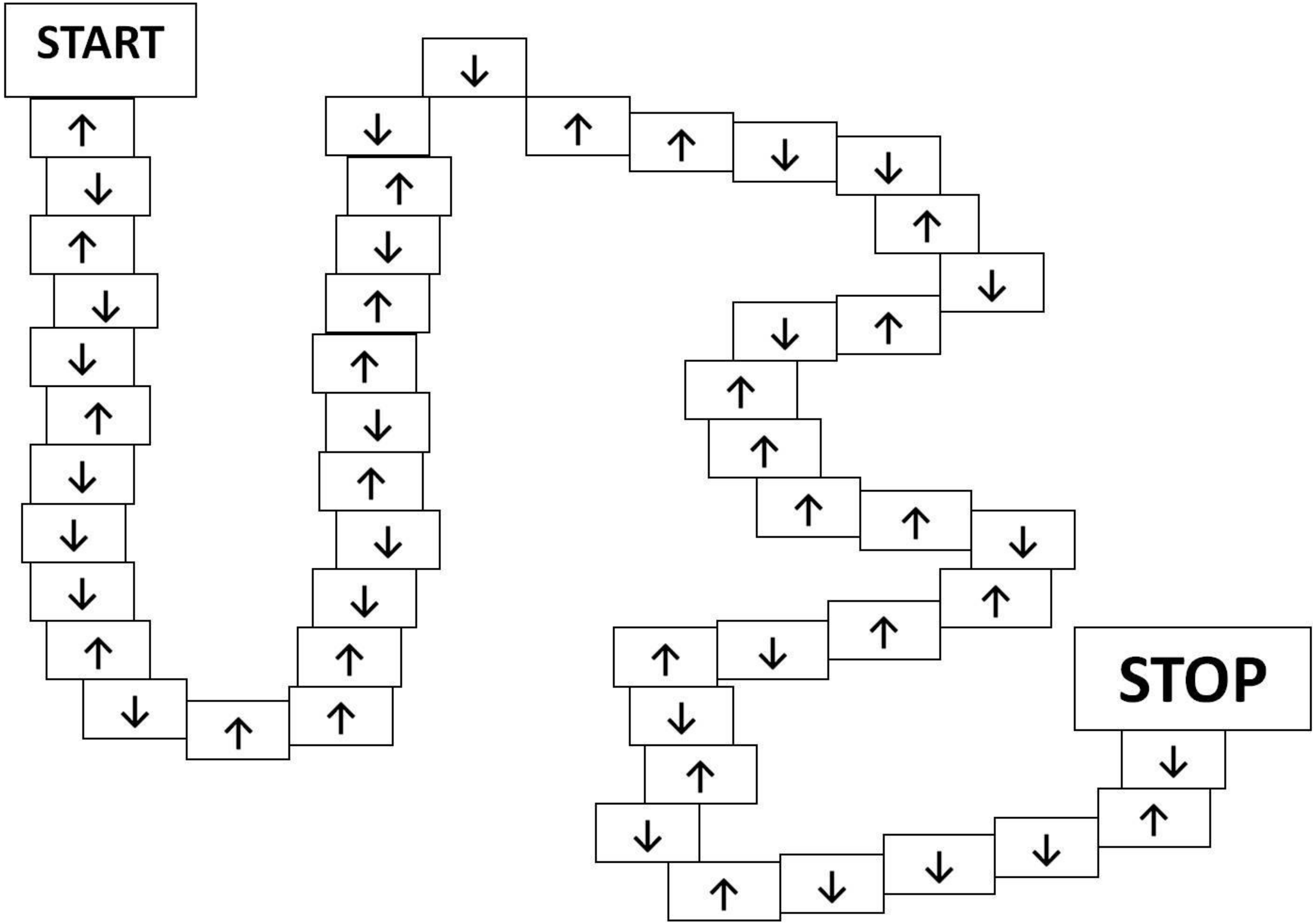
Appendix B: 2 Minute Spelling Task

Book	Succeed
Plan	Thought
Black	Daughter
Brilliant	Recognition
Yacht	Wreckage
Portrait	Neighbour
Control	Evidence
Deceit	Married
Attack	Situation
Spring	Foreigner
Traitor	Vociferous
Weird	Beautiful
Despair	Efficiency
Addition	Uncertainties
Snail	Ceiling
Reported	Wicker
Enough	Accommodation
Vaccine	Provision
People	Conceited
Hockey	Although
Quickness	Painted
Knowledge	Brittle

Appendix C: Inhibition Task Same Direction



Appendix D: Inhibition Task Opposite Direction



Appendix E: Phoneme Deletion Task BALI Pilot Study

Block 1 CCVC – Words	**Correct Response**	**Block 2 CVCC – Words**	**Correct Response**
STEK	SEK	SONT	SOT
PROOSH	POOSH	DOOLT	DOOT
FRISS	FISS	NAST	NAT
TWISH	TISH	HINK	HIK
DRAK	DAK	FESP	FEP
SPOL	SOL	BILK	BIK
TREAP	TEAP	SEMP	SEP
GLEB	GEB	JAST	JAT
BLOACH	BOACH	LOASK	LOAK
PLEEM	PEEM	WELP	WEP
DWIB	DIB	SHAND	SHAD
SMUP	SUP	PELF	PEF

Appendix F: Nonword Repetition Task BALI Pilot Study

Nonwords	**Number of Syllables**
Melpersai	3
Nimutyco	4
Fibeshanoll	4
Nisiverupy	5
Stamundingper	4
Raksibledercon	5
Gionityclafiden	6
Tudryonippor	5
Clirestaibispion	6
Stradiadtionmoy	5

Appendix G: Rapid Automatized Naming Digits Task (Block 1)

7	**2**	**9**	**3**	**6**	**9**	**2**	**6**
9	**3**	**7**	**6**	**3**	**7**	**3**	**2**
7	**9**	**2**	**6**	**2**	**3**	**6**	**7**
9	**6**	**3**	**7**	**9**	**2**	**3**	**9**
7	**2**	**6**	**2**	**7**	**3**	**9**	**6**

Appendix H: Rapid Automatized Naming Objects Task (Block 1)

Appendix I: 1 Minute Reading Task for BALI Main Study

from	very	effect	service	direction	quality
said	idea	hotel	without	limited	continued
that	many	level	different	reaction	industry
were	into	result	trouble	following	president
They	about	series	working	committee	national
Some	father	figure	determine	department	performance
Have	under	little	personal	general	temperature
With	always	higher	evidence	provided	property
Time	table	matter	afternoon	however	education
More	better	never	division	material	society
This	degree	system	natural	together	available
Must	power	nature	possible	beginning	literature
When	after	office	therefore	attitude	development
There	union	second	century	hospital	original
Which	island	children	radio	experience	particular
Kind	centre	number	family	religion	individual
Through	design	party	similar	important	population
Should	value	question	agreement	consider	understanding
Point	doctor	woman	secretary	business	professional
Know	future	entire	officer	government	situation
Public	human	college	anyone	involved	information
Morning	letter	feeling	company	equipment	administration
Complete	really	above	position	character	opportunity
Only	other	treatment	history	decision	university
Upon	money	because	average	production	responsibility

Appendix J: Stimuli Phoneme Deletion Task BALI Main Study

Block 1 CCVC – Words	**Correct Response**	**Block 2 CVCC – Words**	**Correct Response**
STEK	SEK	SONT	SOT
PROOSH	POOSH	DOOLT	DOOT
FRISS	FISS	GALN	GAN
TWISH	TISH	HINK	HIK
DRAK	DAK	FESP	FEP
SPOL	SOL	THOIST	THOIT
TREAP	TEAP	HESHP	HEP
GLEB	GEB	JAST	JAT
BLOACH	BOACH	LOASK	LOAK
PLEEM	PEEM	WELP	WEP
DWIB	DIB	SHAND	SHAD
SMUP	SUP	PELF	PEF

Appendix K: Stimuli Nonword Repetition Task BALI Main Study

Nonwords	**Number of Syllables**
Stamundingper	4
Thrasplabity	4
Tudryonippor	5
Stradiadtionmoy	5
Vizationlisi	5
Senpretaretive	5
Gionityclafiden	6
Breplaposenciary	6
Scriggaflumianal	6
Rystupicarelapto	7