Screening for Learning Disabilities: Development of a Multiple Deficit Model of Identification

Sarah Denise Callinan

B.A. (Hons).

April 2011

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy in the Faculty of Higher Education, Lilydale at Swinburne University of Technology
Abstract

The focus of this thesis is on the operationalisation and identification of Learning Disabilities (LD) in a classroom setting. Although cognitive processing deficits often form part of the definition of LD, they are rarely utilised in methods of identification (Kavale, Holdnack, & Mostert, 2005). For instance, the discrepancy method identifies students with a significant discrepancy between aptitude and achievement (e.g., Berk, 1983; Ferrer, Shaywitz, Holahan, Marchione & Shaywitz, 2010). The primary aim of this thesis was to pilot a freely available screening tool that can identify possible cases of LD in a classroom setting. The validation of the discrepancy method of identification of LD is often sought by testing the cognitive processing deficits of the students identified. In this study the reverse of this approach is taken so that the convergent validity of a method of identification based on cognitive processing deficits will be assessed with the discrepancy method of identification.

Phonological processing is the most widely accepted cognitive processing deficit associated with LD (van der Leij & Morfidi, 2006) and a nonword reading list was developed in this study to test for this. In order to test orthographic processing, which has a more opaque role in LD, an irregular word reading test was developed. Both word lists were refined and psychometric support for their validity was provided through both item response theory and classical test theory. Rasch analysis was used to demonstrate the good model fit of both of the refined nonword and irregular word lists. In order to account for the heterogeneity of LD, naming speed and verbal memory deficits were also used as indicators of the disorder. Two Rapid Automatic Naming (RAN) tests, using letters and colours, and preliminary norms for digit span tests, both forwards and backwards, were also developed. One of the most important attributes of these tests is that they are simple to administer and score, making them well suited to classroom teacher administration. In order to allocate students into discrepancy groups, the Raven’s Progressive Matrices (Raven, Court & Raven, 1998) and Progressive Achievement Test in Reading (Australian Council of Educational Research, 2008), as measures of aptitude and achievement respectively, were also administered to the entire sample. Finally, the South Australian Spelling Test (Westwood, 2005) was administered so that the relationships between spelling and the cognitive processes could be assessed.
Participants were sourced from five primary schools in the Eastern and South-Eastern suburbs of Melbourne, Australia. The sample was made up of 172 Grade 3 participants, as well as 58 Grade 1 and 57 Grade 2 students, who provided a reading age-based comparison for poor Grade 3 readers. Two cluster analyses, one with reading and cognitive variables and one with spelling and cognitive variables, indicated that these cognitive processes were more closely related to reading than spelling and as such the focus for the rest of the analyses was focused on reading-based, rather than spelling-based LD. All Grade 3 students were allocated into LD, Low Achievement (LA) or Regular Achievement (RA) groups as dictated by both the traditional and regression-adjusted discrepancy methods. The cognitive processing attributes of these two groups were examined. LA students, on the whole, had the same cognitive processing proficiency as their LD counterparts. The only exception was that, unexpectedly, LD students scored significantly lower on irregular word reading than their LA counterparts when the groups were dictated by the regression-adjusted discrepancy method.

Discriminant Function Analysis was used to investigate whether cognitive processes can be used to predict membership in both sets of discrepancy defined groups. In the first DFA, it was found that 77% of participants were correctly allocated into traditional discrepancy defined groups using only scores on nonword reading, RAN and digit span. In the second DFA, 82% of participants were correctly allocated into regression-adjusted discrepancy defined groups using the same three predictor variables. Interestingly, in the regression-adjusted discrepancy DFA no students were allocated into the LA group, despite the presence of poor readers in the RA group. Although there were still no differences between the LD and LA groups as allocated by the DFA model predicting traditional based discrepancy groups in phonological processing or naming speed, the LA group had significantly lower verbal memory scores than the LD group. This indicates that, contrary to the proposition that LD students have cognitive processing deficits that do not trouble an LA cohort; some LA students are facing the same deficits as LD students, in addition to verbal memory deficits that may reduce their aptitude. Furthermore, as verbal memory was utilised as a successful predictor variable in this model, these findings suggest that LD is best represented by a multiple deficits model in which at least three cognitive processing deficits may contribute to the disorder.
Finally, a new method of screening for LD students based on cognitive deficits alone is put forward. In this multiple deficits model, students were designated as possibly LD on the basis of reading age matched phonological deficits or chronological aged matched naming speed or verbal memory deficits. Using this method, 79% of students were placed in the same groups as they would have been through the regression-adjusted discrepancy method. As predicted, multiple deficits resulted in more severe reading deficits, despite reading scores playing no role in the allocation into these new groups. By utilising classroom friendly, freely available tests that automatically calculate results and generate student profiles, more students can gain access to screening tools that can predict the possibility of LD and hence enable more appropriate remediation.
Acknowledgements

First and foremost I would like to thank both of my supervisors, Associate Professor Everarda Cunningham and Associate Professor Stephen Theiler. I still cannot get over how lucky I was to walk into a new university and end up with supervisors who are both so good at what they do. Arda, thank you so much for taking a chance on me as a student and for being so generous in providing me with so much expertise, and assistance and so many opportunities. Steve, thank you for being so generous with your time and expertise and thank you for your guidance and calming influence, much needed at times. And to both of you, thank you for being so patient with my impatience!

I feel very fortunate to have worked with so many wonderful people in the research office. Jason Skues, your help with getting me started with both the project and recruitment was invaluable, well above and beyond anything that anyone would expect from a new colleague. Furthermore, Jason and Kay Munyard, thank you for your help with the professional development. There is no way I could have done that without you. To Kay, Jen Nichols and Macy Gray - so many problems in this thesis were solved over a drink or a walk with you three. Thank you for being so awesome. A big thank you to Nadine White for always looking out for all the research students and being generally wonderful. And finally, thank you so much to Jess Sharp for convincing me to do this in the first place - and of course for being such a cool lady!

This research could not have happened without the assistance of the schools, principals, teachers and students who participated in this study. I am extremely grateful to all of those involved in this project. Special thanks to Cath Buckland and Louise Ward for being so accommodating and helpful to someone new to the school system.

Thank you to my parents for being so supportive over the last few (30) years and for providing me with every opportunity to go on to do something like a PhD. And to my husband, Mark, for putting up with me whilst I got this thing done - I’m very lucky, thank you. I’m starting to think you’re not all bad.
Declaration

This is to certify that this thesis:

i. Contains no materials which has been accepted for the award to the candidate of any other degree or diploma, except where due reference is made in the text of the examinable outcome;

ii. To the best of my knowledge contains no material previously published or written by another person except where due reference is made in the text of examinable outcome; and

iii. Where work is based on joint research or publications, discloses the relative contributions of the receptive workers or authors.

Signature:.................................................................
### Chapter 1

Overview .................................................................................................................... 1

Background of the Study .......................................................................................... 1

Terminology ............................................................................................................. 3

  Learning Disabilities ............................................................................................ 3

  Cognitive processes .............................................................................................. 4

Focus of the Thesis ................................................................................................. 6

Purpose of the Thesis ............................................................................................... 7

Significance of the Thesis ........................................................................................ 8

Scope of the Thesis .................................................................................................. 9

Structure of the Research ....................................................................................... 10

### Chapter 2

Overview .................................................................................................................. 13

Literacy .................................................................................................................. 14

Learning Disabilities .............................................................................................. 15

  Nomenclature ..................................................................................................... 21

  Aetiology ............................................................................................................ 23

Operationalisation of Learning Disabilities ........................................................... 25

Discrepancy ........................................................................................................... 26

  Criticisms of the discrepancy method of identification .................................... 30

    IQ tests ............................................................................................................. 31

    Difference scores ........................................................................................ 33

    Learning disabilities and low achievement ................................................. 34

Low Achievement .................................................................................................. 37

Response to Intervention ....................................................................................... 38

  Criticisms of response to intervention ............................................................ 42

    Theoretical criticisms of response to intervention ........................................ 43
Overview .............................................................................................................. 101
Research Design ................................................................................................... 101
  Causal-comparative research ........................................................................... 102
  Applied research ............................................................................................... 103
Test Theory .......................................................................................................... 104
  Classical test theory .......................................................................................... 105
  Item response theory ........................................................................................ 106
Screening Tools .................................................................................................... 107
Research Instruments ........................................................................................... 109
  Tests of aptitude and achievement ................................................................... 109
    Progressive achievement test in reading ...................................................... 109
      Test of reading comprehension for Grade 1 students............................... 110
  South Australian spelling test ...................................................................... 111
  Raven’s progressive matrices ....................................................................... 112
Cognitive processing tests ................................................................................ 113
  Irregular word reading list ............................................................................ 114
  Nonword reading list .................................................................................... 116
  Rapid automatic naming ............................................................................... 117
  Digit span ..................................................................................................... 120
Test administration over grade levels .............................................................. 123
Sampling Framework ........................................................................................... 124
  Sampling method ............................................................................................. 124
  Schools ............................................................................................................. 124
  Students ............................................................................................................ 124
Procedure .............................................................................................................. 125
  Ethics ................................................................................................................ 125
  Administration ................................................................................................ 125
Grade 1 and 2 cognitive processing tests administration ......................... 127
Data Analysis ........................................................................................................ 127
Statistical programs ........................................................................................... 127
Statistical Analyses ............................................................................................ 127
Hierarchical data ............................................................................................... 128
Missing data ......................................................................................................... 129
Identifying students with learning disabilities ............................................ 129
  Traditional discrepancy .................................................................................. 129
  Regression-adjusted discrepancy ................................................................ 130
Rasch analysis .................................................................................................... 131
  Thresholds ..................................................................................................... 132
  Individual item fit ......................................................................................... 132
  Person fit ..................................................................................................... 133
Dimensionality .................................................................................................. 134
Targeting and reliability .................................................................................... 135
Cluster analysis .................................................................................................. 135
Discriminant function analysis ........................................................................ 136
Identifying students with learning disabilities via cognitive deficits ............ 137
Age-Comparisons .............................................................................................. 137
  Chronological age comparison .................................................................. 138
  Reading-age comparisons ......................................................................... 138
Summary .............................................................................................................. 142

Chapter 5 ............................................................................................................. 143
Overview ........................................................................................................... 143
Method .............................................................................................................. 144
Rasch Analysis .................................................................................................. 144
  Rasch analysis results for the irregular word list ........................................ 146
Chapter 6

Overview

Method

Summary Statistics

Descriptive statistics

Internal consistency

Test Re-Test Reliability
Homogeneity of the Grade 3 sample.................................................................179
Assumptions of normality..............................................................................180
Correlation matrix of all Grade 3 variables.....................................................181
Descriptive statistics summary......................................................................183

Cluster Analysis ..............................................................................................183
Reading cluster analysis................................................................................184
  Reading cluster analysis: Comparison of groups........................................186
Spelling cluster analysis...............................................................................188
  Spelling cluster analysis: Comparison of groups.................................188
Summary of findings from the cluster analyses.........................................189

Identifying students with learning disabilities with discrepancy-based methods 190
  Traditional discrepancy...........................................................................191
  Regression-adjusted discrepancy............................................................193

Discriminant Function Analysis.................................................................196
  Traditional discrepancy discriminant function analysis........................197
    Traditional discrepancy discriminant function analysis: Comparison of
groups.........................................................................................................199
  Regression-adjusted discriminant function analysis..............................201
    Regression-adjusted discriminant function analysis: Comparison of groups
.......................................................................................................................203

Summary of findings from the discriminant function analyses................204

Identifying Students with Learning Disabilities through Cognitive Processing
Deficits ...........................................................................................................205
  Identifying reading-age-based phonological deficits................................206
  Identifying chronological age matched deficits.........................................207
  Cognitive processing deficit groups.........................................................208

Summary ......................................................................................................212

Chapter 7 ......................................................................................................214
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>List of One Point Responses for the Full Irregular Word List</td>
<td>116</td>
</tr>
<tr>
<td>Table 2</td>
<td>Correct and Partially Correct Responses to Nonwords in the Full Nonword List</td>
<td>119</td>
</tr>
<tr>
<td>Table 3</td>
<td>Tally of Tests Administered to Different Grade Levels</td>
<td>123</td>
</tr>
<tr>
<td>Table 4</td>
<td>Number of Grade 3 Participants, Mean Ages and Response Rates per School</td>
<td>125</td>
</tr>
<tr>
<td>Table 5</td>
<td>Means and Standard Deviations of the Original Irregular Word List</td>
<td>147</td>
</tr>
<tr>
<td>Table 6</td>
<td>Reasons for Item Removal and Summary Statistics of the Rasch Analysis after each Item Removal for the Irregular Word List</td>
<td>153</td>
</tr>
<tr>
<td>Table 7</td>
<td>Means and Standard Deviations of the Final Irregular Word List</td>
<td>155</td>
</tr>
<tr>
<td>Table 8</td>
<td>Item Fit Residuals for the Final Irregular Word List</td>
<td>157</td>
</tr>
<tr>
<td>Table 9</td>
<td>Means and Standard Deviations within the Full Nonword List</td>
<td>162</td>
</tr>
<tr>
<td>Table 10</td>
<td>Reasons for Item Removal and Summary Statistics of the Rasch Analysis after each Item Removal for the Nonword List</td>
<td>166</td>
</tr>
<tr>
<td>Table 11</td>
<td>Means and Standard Deviations of the Final Nonword List</td>
<td>168</td>
</tr>
<tr>
<td>Table 12</td>
<td>Item Fit Analysis for the Final Nonword List</td>
<td>170</td>
</tr>
<tr>
<td>Table 13</td>
<td>Means, Standard Deviations, Skew and Kurtosis for all Grades, on all Tests</td>
<td>177</td>
</tr>
<tr>
<td>Table 14</td>
<td>Correlations for Grade 3 Participants on all Reading and Cognitive Processing Tests</td>
<td>182</td>
</tr>
<tr>
<td>Table 15</td>
<td>Descriptive Statistics and Results from the One-Way ANOVA Comparing Poor, Average and Good Readers</td>
<td>186</td>
</tr>
</tbody>
</table>
Table 16  Post-Hoc T-Test Results Comparing Poor and Average Cluster Analysis-Defined Reading Groups ..........................................................187
Table 17  Means, Standard Deviations and T-Test Comparing Low and Good Spellers on all Tests .................................................................189
Table 18  Means, Standard Deviations and ANOVA Results Comparing Traditional Discrepancy-Defined Groups for Cognitive Processing Tests ......................................................................191
Table 19  Planned Contrast Comparing Traditional Discrepancy-Defined LD Students to LA Students on Significant Cognitive Processing Tests ........................................................................192
Table 20  Means, Standard Deviations and Corresponding ANOVA Results of the Three Regression-Adjusted Discrepancy Defined Groups on Measures of Reading and Cognitive Processing ...195
Table 21  Planned Contrasts Comparing Regression-Adjusted Discrepancy Defined Groups on Irregular Word Reading, Nonword Reading, Digit Span Forward and Letter Rapid Automatic Naming......196
Table 22  Discriminant Functions Predicting Membership in Traditional Discrepancy Groups .........................................................................199
Table 23  Classification of Students into Traditional Discrepancy-Defined Groups through a Discriminant Function Analysis .............................199
Table 24  Means, Standard Deviations and Results from the One-Way ANOVA Comparing Traditional Discrepancy Discriminant Function Analysis Defined Groups ........................................200
Table 25  Results from the Planned Contrast Comparing Regular Achieving Students to Students with Learning Disabilities and Low Achievement in Traditional Discrepancy Discriminant Function Analysis Defined Groups ........................................201
Table 26  Discriminant Functions Predicting Membership in Regression-Adjusted Discrepancy Groups ............................................................................202
Table 27  Classification of Students into Regression-Adjusted Discrepancy
Defined Groups .................................................................203

Table 28 Means, Standard Deviations and Results from the One-Way
ANOVA Comparing the Discriminant Function Analysis
Defined Groups .................................................................204

Table 29 Cut-Off Scores for Nonword Reading Per Grade Level as
Dictated by Average Readers in Each Grade ..................207

Table 30 Means and Standard Deviations of Raven’s Progressive
Matrices, Progressive Achievement Test in Reading and the
South Australian Spelling Test Scores Within Multiple Deficit
Groupings .................................................................209

Table 31 Scores on all Tests for Students with Regular Achievement,
Learning Disabilities and Low Achievement as Identified via the
Multiple Deficits Model .................................................211

Table 32 Cross Tabulation of Students with Regular Achievement,
Learning Disabilities and Low Achievement via the Regression-
Adjusted Discrepancy Method and the Multiple Deficits Model
..................................................................................212
List of Figures

Figure 1  Sample Page from the Irregular Word List ...........................................115
Figure 2  Sample Page from the Nonword List ..................................................118
Figure 3  Copy of the Colour Rapid Automatic Naming Test .............................121
Figure 4  Copy of the Letter Rapid Automatic Naming Test ..............................122
Figure 5  Normal Distribution Curve of Grade 1, 2 and 3 Readers on a Grade
3 Percentile Axis ..........................................................................................140
Figure 6  Normal Distribution Curves of Grade 1, 2 and 3 Readers on a Grade
3 z-Score Axis with Cut-Points Designating Reading Age ..........................141
Figure 7  Scree Plot Comparing Eigenvalues from the Principal Axis Factor
Analysis of the Full Irregular Word List and the Eigenvalues
Randomly Generated in the Parallel Analysis .........................................148
Figure 8  Threshold Map for the Original Irregular Word List ............................150
Figure 9  Category Probability Curve for the Item frightened .......................151
Figure 10 Category Probability Curve for the Rescored Item frightened ....152
Figure 11 Threshold Map for the Final Irregular Word List ..............................156
Figure 12 Histogram for Total Scores for the Final Irregular Word List .............158
Figure 13 Person Item Interaction for the Final Irregular Word List .................159
Figure 14 Scree Plot Comparing the Eigenvalues from the Refined Irregular
Word List and from the Randomly Generated Parallel Analysis ..............160
Figure 15 Scree Plot Comparing Eigenvalues from the Principle Axis Factor
Analysis of the Full Nonword List and the Eigenvalues Randomly
Generated in the Parallel Analysis ..............................................................163
Figure 16 Threshold Map for the Full Nonword List ......................................164
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 17</td>
<td>Category Probability Curve for the Item <em>eero</em></td>
</tr>
<tr>
<td>Figure 18</td>
<td>Threshold Map for the Final Nonword List</td>
</tr>
<tr>
<td>Figure 19</td>
<td>Histogram of Total Scores for the Final Nonword List</td>
</tr>
<tr>
<td>Figure 20</td>
<td>Person-Item Interaction for the Final Nonword List</td>
</tr>
<tr>
<td>Figure 21</td>
<td>Scree Plot Comparing Eigenvalues from the Principal Axis Factor Analysis of the Full Nonword List and the Eigenvalues Randomly Generated in the Parallel Analysis</td>
</tr>
<tr>
<td>Figure 22</td>
<td>95% Confidence Interval of the Mean of all Variables for Groups of Poor, Average and Good Readers as Identified by a Cluster Analysis</td>
</tr>
<tr>
<td>Figure 23</td>
<td>Scatter Plot Displaying Aptitude and Achievement Scores of Grade 3 Students, Allocated into groups via the Traditional or Regression-Adjusted Discrepancy Method</td>
</tr>
</tbody>
</table>
Chapter 1

Overview

The research presented in this dissertation aims to increase the understanding of learning disabilities (LD) through the development and evaluation of a screening tool designed to identify the disorder. In order to do this effectively, a multiple deficits model which focuses on the causes, rather than the outcomes, of the disability has been developed. This chapter sets the context for the study and presents six main areas, including the general background of the study, clarification of terminology, the purposes of the research, as well as the significance and the scope of this research. The chapter concludes with an overview of the overall structure of the thesis.

Background of the Study

The adverse outcomes of poor literacy in general (Australian Council for Educational Research [ACER] 2000; McGee, Prior, Williams, Smart & Sanson, 2002) and LD in particular (Kortering, Braziel & McClannon, 2010; Undheim & Sund, 2008) are well established. Despite this, many Australian students with LD are leaving school early without knowledge of their LD and are then at greater risk for unemployment, juvenile delinquency, criminal conviction and mental health problems (Smart, Prior, Sanson & Oberklaid, 1996). Estimates of the prevalence of LD range from 2-3% (Kavale, Spaulding & Beam, 2009) to 10% (Prior, 1996).

In Australia, it is difficult to ascertain a specific prevalence rate of LD as there is no federal recognition of the disorder. Consequently, many students with LD are not formally diagnosed or identified. A Commonwealth enquiry into LD reported that teachers have limited knowledge of the phenomenon or of how to effectively respond to the needs of this group of students (Louden et al., 2000). Unless students and teachers have knowledge and understanding of their learning problems, it is difficult for them to take control of their circumstances through compensatory strategies that meet the student’s individual needs.

The need for new methods of combating literacy problems, such as LD, has been highlighted in recent changes to Australian policy. A new program provides
financial incentive to primary schools that implement innovative programs that improve literacy or numeracy in Grade 3 or 5 students (Department of Education, Employment and Workplace Relations, [DEEWR], 2010). Incentives like this are important as LD will be increasingly over-represented in low socio-economic groups. This is because those with literacy difficulties are more likely to end up living in poverty, with menial jobs on low pay or in prison (Kirsch, Jungeblut, Jenkins & Kolstad, 1993). Furthermore the genetic basis of LD indicates that those with the disorder and all its inherent disadvantages are more likely to pass these problems on to their children than those without LD (Berninger & Holdnack, 2008). Consequently, over time, children with LD will be increasingly over-represented in a low socio-economic demographic. This places the lack of affordable screening or diagnostic tools in a different light. It is important that LD identification, or at least screening, can be made freely available so that the students most likely to benefit from being identified as having a potential problem have access to the resources.

Unfortunately, LD identification currently requires highly qualified professionals for sizeable time periods, making many services inaccessible for those who do not have the means to obtain them (Goodman & Webb, 2006). Teachers lack accessible, reliable and easy-to-administer instruments that may indicate the presence of LD in their students. Such knowledge is critical for teachers so that the necessary accommodations that enable the student to participate more fully in school and beyond are developed (Abernathy & Taylor, 2009). Therefore, any method of LD identification developed by researchers must be easily integrated into a classroom setting or it is less likely to receive support from schools (D. Fuchs, Fuchs, Harris & Roberts, 1996).

However, it is not enough to simply designate this as a task for teachers simply because all school children need to have access to screening, without at least providing a realistic program for how this can be done. It is also imperative that the tests are relatively simple for teachers to administer and interpret. Therefore, the onus is on LD researchers to ensure that screening tools are developed with interpretation of results embedded into the model. By providing this resource it should be possible to empower teachers to identify reading problems without placing unrealistic expectations on their workload or training.
One of the barriers for educational professionals is that much of the research in this field is not accessible. This is in part due to the confusion in nomenclature in this field. The next section clarifies the use of nomenclature in this dissertation.

**Terminology**

**Learning Disabilities**

The sheer volume of special education research in LD is daunting. A panel was recently put together to nominate the most important articles in the field, in order to decrease the chance of the more important articles being overlooked (Scanlon, Boudah, Elksnin, Gersten & Klingner, 2003). This prolific output is made more confusing by inconsistent terminology. In 2006, an article on LD and nomenclature likened LD research to the Bible story of the Tower of Babel. In this story, God gave everyone different languages so they could not work together to reach their common goal. The authors of this article went on to say:

> Understanding LD and how to diagnose it does not, of course, represent the same lofty goal as that held by the Tower of Babel’s builders. Nevertheless, the task of delineating an LD definition and a corresponding set of diagnostic criteria suffers from the very same problem – failure to use a common language and an inability to communicate within and across disciplines using a common set of terms that mean the same thing to all who use them (Flanagan, Ortiz, Alfonso & Dynda, 2006, p.807).

This failure to develop a common language is reflected in the various terms that are used interchangeably in LD research. Learning disabilities, specific learning disabilities, learning disorders, learning difficulties, dyslexia, reading difficulties, reading disorders and reading disabilities have all been used interchangeably to represent difference facets of LD, often further confusing research. Sometimes in an attempt to clarify this, hierarchies are developed. For example, learning disabilities are a more serious subset of learning difficulties (Gartner & Lipsky, 1987); dyslexia is a reading specific learning disability (Mather & Gregg, 2006); or a reading disorder is a reading specific learning disorder (American Psychiatric Association, 2000). However none of these clarifications has been widely accepted, so the confusion continues.
The outcome of this confusion in LD terminology and definition is that there is no specific set of operationalisations or definitions attached to any of these terms that has not at one stage been attached to the others. For instance, what is operationalised as dyslexia in one paper, such as low achievement with an IQ score higher than a given cut-off (e.g. Jimenez et al., 2009), is different to another operationalisation of dyslexia using regression-adjusted discrepancy methods (e.g., Pennington & Lefly, 2001), which would also be used as the operationalisation of a reading disability in yet another paper (e.g., Stanovich & Siegel, 1994). This confusion is, of course, only recognisable when an operationalisation is provided and, unfortunately this is not always the case. A meta-analysis of LD operationalisations in Canadian journals found that 31% of articles with LD cohorts did not make it clear how this group was identified (Klassen, 2002). Naturally, this causes much confusion in the field, especially when one wishes to generalise results.

Because the focus of the current study is on operationalisation, not nomenclature, one term will be used consistently throughout this paper, namely Learning Disabilities (LD). This decision will not have an impact on the generalisability of the current study to previous research, as varying operationalisations cause difficulty in this field regardless of nomenclature. Throughout this study differences in nomenclature will only be noted when relevant to definitions or prevalence.

**Cognitive processes**

The issue of interchangeable nomenclature in this field is not specific to the description of LD. Many of the cognitive processes pertinent to LD also face this same issue, including the three tested in the current study. The terminology and definitions of phonological processing, naming speed and verbal memory, as used in the current study, will be clarified here.

The cognitive processing deficit most commonly referred to in studies on LD is phonological processing (Ramus et al., 2003). Phonological processing has been referred to as “the ability to perceive and deliberately manipulate the sound components of words” (Kirby, Roth, Desrochers & Lai, 2008, p.105). Often incorrectly used interchangeably with the terms phonological awareness, phoneme awareness or phonological decoding. In one clarification of the relevant nomenclature, phonological processing was considered an umbrella term that covers
phonological awareness, phonological recoding and phonetic recoding (Castles & Coltheart, 2004). The finer points in the definition of phonological processing, or any other relevant processes, are beyond the scope of the current study. The definition by Kirby and colleagues will be used in the current study and, as such, the term phonological processing will be used.

Another cognitive process that has been burdened with inconsistent nomenclature, albeit less than phonological processing, is naming speed. Often confused with processing speed, naming speed is specific to the speed at which a participant can name separate consecutive units, such as letters or colours. Therefore, although sometimes confused or linked with processing speed, there is actually little shared variance between the two (Wolf, Bowers & Biddle, 2000). The Rapid Automatic Naming (RAN) test is the widely accepted measure of naming speed. Due to the specificity of the test to the concept, RAN is often used in place of the term naming speed. However, in an effort to reduce confusion, the term RAN will only be used when referring to the test, not the construct.

Working and short-term memory are other cognitive deficits that are only recently gaining popularity in the literature surrounding the cause of LD. The differentiation between the two when digit span is utilised is not always clear. Forward digit span is widely regarded as a short-term memory measure, but backward digit span has been used to measure both short term (Conway et al., 2005) and working memory (Savage et al., 2005). Both processes are thought to be important for reading comprehension (Swanson & Howell, 2001) which in part is because both digit span measures are thought to tap into the phonological loop (Baddeley, Gathercole & Popagno, 1998). This phonological loop is where phonological information is kept for a very limited time, unless the sound is mentally rehearsed, however its relationship to short term and working memory is still unclear (Baddeley, 1986). This debate on short term and working memory is a complex one and not within the scope of the current study. Consequently, the term verbal memory will be used when referring to the construct tested by both forward and backward digit span, as has been done in other LD research (e.g., Helland & Asbjørnsen, 2004; Savage et al., 2005).
Focus of the Thesis
In this dissertation a new model of LD identification based on cognitive processing deficits is put forward. Currently, the manner in which LD is identified at any given time tends to influence the definition of the disorder, rather than the other way around (McPhail & Palincsar, 1998). Although cognitive processing deficits often form part of the definition of LD, they are rarely separately assessed and used as the sole benchmarks in LD identification (Kavale, Holdnack & Mostert, 2005).

What is proposed in the current study is that potential cases of LD be identified in the classroom via the cognitive processing deficits thought to cause the disorder. This is similar to other multiple deficit models have been developed in the past in order to identify LD (Fiorello, Hale & Snyder, 2006; Ho, Chan, Tsang & Lee, 2002), albeit with a different combination of variables. Furthermore in the current study the new model will be developed within an applied research paradigm making implementation of the model in the classroom possible. Although there are different deficits used in this study, the logic behind the multiple deficits models remain the same.

One of the major advantages of a multiple deficits model is that it will be possible to identify students before their reading is affected, instead of waiting for it to become problematic enough to be picked up by methods dependent on current reading proficiency. This is important, as some of these students are able to compensate for their difficulties early on in reading acquisition. Therefore, problems might not be recognised until as late as Grade 3 or 4, when the expectations placed on students begin to change (Woolley, 2007). The second major advantage is that it will be possible to identify potential LD without exclusionary criteria such as cultural disadvantage or low intelligence, as cognitive processing deficits should occur independently of these issues.

The ability to assess concurrent validity of a multiple deficits model of possible LD identification is dependent on having a yardstick with which to measure success. In the case of the current study, the aptitude/achievement discrepancy will be used. In this method a significant discrepancy between aptitude and achievement is thought to be indicative of LD. This method will be used cautiously as the benchmark for convergent validity as much of the criticism directed towards this method of LD
identification is justified. Although the discrepancy method should not be used as a definitive tool in identification, as a number of research papers testify (e.g., Aaron, Malatesha Joshi, Gooden & Bentum, 2008; Cahan, Fono & Nirel, 2010), it does represent the LD prototype well. Consequently, the discrepancy method can be used as an effective baseline measure with which improved measures of LD can be compared.

The multiple deficits method of identification, with simple tests for specific cognitive processes, will not only overcome some theoretical issues that plague LD identification, but also some applied ones.

**Purpose of the Thesis**

The primary purpose of the current study was to pilot a battery of screening tests that could be used to identify LD in a classroom setting. This is done by developing new tests of phonological and orthographic processing, naming speed and verbal memory. In addition, existing normed tests of reading, spelling and non-verbal intelligence are also used. As part of this model development, the following research questions and hypotheses will be tested.

1. Can the psychometric validity of the newly-developed irregular word and nonword lists be supported as per the tenets of both Item Response Theory (IRT) and Classical Test Theory (CTT)?
2. Can cognitive processing deficits be used to identify a similar cohort of LD students as would be identified via the discrepancy method?
3. Are the identification methods and cognitive processes currently researched in terms of reading-based LD relevant to spelling-based LD?
4. Are many students identified as Low Achievement (LA) by the discrepancy method of identification facing the same cognitive processing difficulties that are faced by LD students?

Furthermore the following hypotheses will also be tested.

1. That deficits in verbal memory will be indicative of LD in the same manner as phonological processing and naming speed.
2. That multiple cognitive processing deficits will be positively associated with more severe reading deficits.
3. That some of the students not identified as having cognitive processing difficulties will still be poor readers.

4. That orthographic processing deficits will provide an alternative explanation for poor reading if no indication of LD is found.

**Significance of the Thesis**

In the current study, a freely available new method of screening for possible LD has been developed. It was important to accommodate both theoretical and applied considerations when identifying LD students so that the screening battery could be incorporated into a school setting. Cognitive processes are simpler to test for than methods using intelligence tests or remediation following failure. Furthermore, this method is also more theoretically sound because it directly addresses the deficits that are thought to cause LD.

In line with this applied paradigm, the first contribution of the current study is the compilation of a screening tool, made up of six tests. These tests have been piloted in the current study, with normative data provided for each. Irregular word and nonword reading lists were specifically developed to respectively identify orthographic and phonological deficits, in a Victorian Grade 3 cohort. These word lists were analysed through the tenets of both IRT and CTT. The two tests of naming speed, Rapid Automatic Naming (RAN) of colours and letters and the two digit span tests of verbal memory, were also piloted and norms were developed for these single-item scales.

The second contribution of the current study was the integration of the new model of identification with the discrepancy model on which convergent validity was based. This integration occurred on two fronts. First of all, the justification for using the discrepancy method of identification as the basis for comparison is advocated, with the advantages and criticisms of that method reviewed. Secondly, as both methods have been used in the current study, any differences in classification are reviewed and a theory for these differences is suggested. By ensuring that the differences between the two models are recognised and made explicit it is possible to facilitate generalisability of results.

The third contribution of the current study is the addition of verbal memory deficits to a traditional double deficit paradigm based on phonological processing and
naming speed deficits as the main causes of LD. Verbal memory deficits, more so than phonological or naming speed deficits, resulted in low aptitude test scores in the current study. This in turn decreases the probability of having a significant discrepancy between aptitude and achievement and increases the probability of a student being designated LA. As a result, the role of verbal memory deficits in LD is less clear than phonological processing or naming speed deficits when viewed through a discrepancy paradigm, but made clearer through a multiple deficits approach.

The multiple deficit model developed within this dissertation is the fourth contribution of the current study. As noted earlier, although this model has a theoretical base, the screening tool itself has been developed with an applied paradigm. By identifying LD on the basis of the cognitive processing deficits thought to cause the disorder, rather than possible secondary outcomes, access to identification tools is increased, as well as accuracy.

**Scope of the Thesis**

To ensure that the aims of the current study are met, guidelines were put in place as to what will and will not be addressed. The primary focus of the current study will be on reading-based LD. This is so that the model can be based on the vast breadth of research on reading-based LD. A content analysis of the Learning Disabilities Research and Practice Journal found that 47% of the articles directly addressed reading and just 4% addressed spelling (Vostal, Hughes, Ruhl, Benedek-Wood & Dexter, 2008). It has been claimed that learning to read and learning to spell are very closely aligned (Ehri, 1997). It has even been suggested that early instruction in learning to spell could form an early intervention for reading disorders (Santoro, Coyne & Simmons, 2006). However, the bulk of research into LD is reading-focused. In the current study the primary focus will be on reading with a secondary, more exploratory focus on spelling. Therefore, the studies referred to in this thesis are focused on reading-based LD and any studies referring to difficulties in other areas will be explicitly noted as such.

Furthermore, the tests developed in the current study make up a screening battery of tests, rather than a diagnostic tool. This is important to note, not least because of the resulting inflated Type II error. This decision was made for two reasons. First of all,
this is because there is no federal recognition of LD in Australia and thus little perceived benefit in widespread diagnoses of the disorder. Secondly, a diagnostic tool would not be suitable for classroom administration. Therefore, the focus of the current study is on screening for possible LD, rather than making diagnoses.

Finally, the current study is focused on Grade 3 students only. This is so that the tests can be specifically tailored to a particular grade level, rather than trying to accommodate different stages of reading acquisition. Grade 1 and 2 students are tested in the current study in order to develop reading age norms, however the presence of LD will not be assessed in these cohorts.

**Structure of the Research**

This thesis is presented in seven chapters. This chapter, the first chapter provided an overview of the current study. Current operationalisations of LD are reviewed and discussed in Chapter 2. Although much of the debate in this field is centred on the discrepancy and response to intervention methods, both of these methods are based on possible outcomes of LD rather than causes. This chapter presents arguments that multiple deficit cognitive processing models explain LD. It is further argued that, for purpose of establishing convergent validity for the role of these particular cognitive deficits, that the discrepancy model should be used. This method was chosen as this idea of unexpected underachievement that discrepancy represents, despite psychometric issues, does represent the quintessential LD student well.

The cognitive processes utilised as predictor variables of LD are discussed in Chapter 3. Research into the role of phonological processing, naming speed and verbal memory deficits as causes of LD is reviewed. The debate about orthographic processing deficits in LD and, specifically their role as a cause or outcome of LD is discussed. It is argued that the applicability of programs reliant on professionally administered tests or extensive remediation programs preceding identification has historically not been well addressed. Consequently, many methods of identification in the past have not been readily integrated into a classroom setting. Applied issues in LD identification will also be reviewed in Chapter 3, including the role of teachers and how identification should inform remediation. Finally, previously developed screening tools of LD will be reviewed.
The methodology of the current study is detailed in Chapter 4. The research design is compatible with both causal comparative and applied research paradigms, both of which are reviewed. IRT and CTT are then discussed in the context of the validation of the word lists. The sampling framework, procedures, and descriptions of the instruments, both new and established, are outlined in this chapter. The method of designating students into both traditional and regression-adjusted discrepancy-defined groups is reviewed as well as the method of identifying reading age phonological deficits. Finally, the statistical analyses used in the current study, including assumptions surrounding the nature of the data and assumptions of the separate tests, are reviewed.

In Chapter 5, the psychometric investigation into the two newly developed word lists is detailed. Both lists are analysed via a Rasch analysis as well as having their unidimensionality and internal consistency tested through CTT. The decision making process for removal of items and iterative analyses after each item removal is presented. The chapter concludes with the presentation of the final word lists refined through Rasch analysis.

In Chapter 6 the internal consistency of finalised word lists along with the test-re-test reliability of the word lists and RAN tests are displayed, along with the descriptive statistics of all tests across all grade-levels. Cluster analyses are used to ascertain the relationship of the cognitive variables to reading and spelling. The Grade 3 sample is then allocated into LD, LA and regular achievement groups as per the traditional and regression-adjusted discrepancy methods. With the two discrepancy-defined groups allocated, ANOVAs were conducted to identify any mean cognitive processing differences between these groups. Two discriminant function analyses were calculated in order to see if tests of phonological processing, naming speed and verbal memory can predict membership in these traditional or regression-adjusted discrepancy-defined groups. Finally, a new method of identifying LD based on cognitive processing deficits is outlined and descriptive statistics of the final groups are displayed.

Chapter 7 opens with an overview of the results as they pertain to the research questions and hypotheses, followed by a discussion of the relationship between each cognitive process examined in this study to both reading and LD. The multiple
deficits method of LD identification is then compared with other methods, with a focus on the theoretical significance of the findings in the current study. The implications of the study for policy and practice, with an emphasis on how the interpretation of such a screening test could be embedded into a computer program are then discussed. Finally the limitations of the study are outlined and suggestions made for future research.
Chapter 2

Overview

In this chapter the definitions, aetiology and prevalence of Learning Disabilities (LD) are discussed and then the thorny issue of the operationalisation is examined. There are a wide range of definitions of LD and most of these acknowledge that cognitive deficits are inherent to the disorder (Kavale et al., 2005). Despite this they are rarely used in operationalisation and instead outcomes that are secondary to LD are often utilised in identification procedures.

In an effort to discriminate between LD and non-LD poor readers, a range of operationalisations have been devised. The discrepancy method of identification is based on the idea that students with LD can be identified by finding those with aptitude test scores significantly higher than their reading test scores. Although this method does seem to capture the essence of LD, that of unexpected scholastic underachievement, there are legitimate theoretical and psychometric criticisms of this method of identification that has resulted in it falling from favour in the relevant literature (Bradley, Danielson & Doolittle, 2005; Lyon, Fletcher, Torgesen, Shaywitz & Chhabra, 2004). These criticisms primarily centre on the lack of reliability of any kind of difference score and the use of IQ tests, both in general and as a predictor of reading achievement (Klassen, Neufield & Munro, 2005). Theoretical issues with the use of exclusionary criteria, which have not received as much attention in the literature, will also be discussed.

In response to these issues, another method, Response to Intervention (RTI), was developed with the aim of presenting a more pragmatic approach to LD identification. This method involves ascertaining the presence of LD on the basis of response to remediation that is provided to the lowest achieving students. However, once again there are criticisms of this method. These include the assumptions that LD students will not respond to intervention and that there is no reason other than LD that a student would not respond to intervention (C. Reynolds & Shaywitz, 2009a).

Both of these methods share many criticisms, primarily stemming from the fact that both are based on secondary symptoms that are likely outcomes of LDs, rather than the causes. Some models of identification have been developed based on the
cognitive processing deficits thought to cause LD. The first of these, the phonological core deficit model was based on the assumption that LD was primarily reflecting a deficit in phonological processing (Stanovich, 1988). However this single deficit model did not seem to be sufficiently addressing the heterogeneous nature of LD (Rack, Snowling & Olson, 1992).

In response to this, a double deficit model was put forward in which LD could be explained by deficits in phonological processing, naming speed or both (Wolf & Bowers, 1999). Although this model had more success in accounting for this diverse group of students, it still did not account for all the variation within the group (Lovett, Steinbach & Frijters, 2000). This chapter concludes by discussing the viability of a multiple deficit model of LD in which verbal memory deficits are also taken into account. Finally the argument that the convergent validity benchmark for a multiple deficit model of LD should be based on the discrepancy method of identification is put forward.

**Literacy**

Although reading is not innate (Frey & Fisher, 2010), the percentage of professions that require functional literacy has grown dramatically in the past century. In the past, those students that could not read had many more options ahead of them than do students who cannot read today (Jansen, 2001). The implications of early difficulty in developing reading skills are long reaching. Longitudinal studies have shown that early language impairment has been tied to both decreased academic outcomes and a higher incidence of LD (Young et al., 2002). In a longitudinal study of Australian and New Zealand students aged five to 18 years old, the long term effects of early poor literacy were examined while controlling for the confounding factor of hyperactivity. This study found that poor readers were less likely than their peers to complete school. Furthermore, very poor readers were more likely to show delinquent behaviour as teenagers (McGee et al., 2002). People with low levels of literacy are also more likely to be unemployed (Australian Council for Educational Research (ACER), 2000).

One of the reasons that early reading difficulty is such a problem is that it tends to compound, rather than decrease, over time. This problem is known as the Matthew Effect (Stanovich, 1986) in which students who are good at reading get better, while
students who are poor at reading get worse. This has been consistently supported in the literature, for instance neurological studies have shown that the more children read the more their brain adapts to be suited to reading (Frey & Fisher, 2010). Alternatively, a one sided Matthew effect has been suggested where the reading rich do not get richer, but the reading poor do get poorer (Morgan, Farkas & Hibbel, 2008), hardly providing much comfort.

Widespread support for the existence of the Matthew Effect has led to an increased focus on providing timely intervention for those who are struggling with reading, before these difficulties compound themselves. Students who are two years behind in reading by the time they reach Grade 4 are unlikely to make much progress between Grade 3 and Year 10 (Hill, 1995, cited in Hempenstall, 2001).

Unfortunately issues with reading can often go undetected until Grade 4, as the emphasis on reading comprehension sets in at Grade 3, as the text students read becomes more complex (Woolley, 2007). Furthermore, it has been found that students who struggle with foundation reading skills, such as mastering the alphabetical principle, have slower reading development later in life (National Institute for Child Health and Development, 2000).

One of the most confusing aspects of research into literacy is that there seems to be a group of students, qualitatively different from normal low achieving students, who do not seem to be able to learn to read despite having every opportunity to do so. If this inability to learn to read is innate, that student is considered to have LD, which is the focus of the current study. The next section will examine current definitions of LD, along with a discussion of the nomenclature and aetiology of the disorder.

**Learning Disabilities**

Learning Disabilities are “specific patterns (subtypes) of neuropsychological assets and deficits that eventuate in specific patterns of formal (e.g., academic) and informal (e.g., social) learning assets and deficits” (Rourke, 2005, p.111). The most commonly attributed symptom of LD is probably unexpected underachievement (e.g., Fletcher, Denton & Francis, 2005; Lyon, Shaywitz & Shaywitz, 2003; Mastropieri & Scruggs, 2005). Furthermore it is well documented that LD is a heterogeneous disorder that involves intra-individual cognitive deficits, presumably of a neurological basis (Scruggs & Mastropieri, 2002).
There are a variety of definitions throughout the world for LD and although on the whole they are fairly consistent, there are some differences between definitions. The American Psychiatric Association (DSM-IV-TR) has defined learning disorders (they do not use the term learning disabilities) as:

... when the individual’s achievement on individually administered, standardized tests in reading, mathematics, or written expression is substantially below that expected for age, schooling, and level of intelligence. The learning problems significantly interfere with academic achievement or activities of daily living that require reading, mathematical, or writing skills. A variety of statistical approaches can be used to establish that a discrepancy is significant. Substantially below is usually defined as a discrepancy of more than 2 standard deviations between achievement and IQ. A smaller discrepancy between achievement and IQ (i.e., between 1 and 2 standard deviations) is sometimes used, especially in cases where an individual’s performance on an IQ test may have been compromised by an associated disorder in cognitive processing, a comorbid mental disorder or a general medical condition, or the individual’s ethnic or cultural background. If a sensory deficit is present, the learning difficulties must be in excess of those usually associated with the deficit (2000, p.49-50).

What is important to take from this definition is that learning disorders have been operationalised as a discrepancy between aptitude and achievement. There are exclusionary criteria present that can be used as alternative explanations of this discrepancy and no specific method of defining what a significant discrepancy constitutes is prescribed. It has been suggested that generic LD terms are redundant and that they should specifically refer to the problem at hand, for instance reading (Stanovich, 1999). The DSM-IV has taken this on and although there is a definition for learning disorders, all of the diagnostic criteria for the different types of disorders under this umbrella have been split in to specific areas of difficulty. For instance the diagnostic criteria for a reading disorder are:
A. Reading achievement as measured by individually administered standardized tests of reading accuracy or comprehension is substantially below that expected given the person’s chronological age, measured intelligence, and age-appropriate education.

B. The disturbance in Criterion A significantly interferes with academic achievement or activities of daily living that requires reading skills.

C. If a sensory deficit is present, the reading difficulties are in excess of those usually associated with it (DSM-IV-TR, 2000, p.53)

Particularly noteworthy here is that firstly, the cause of these deficits is not mentioned and secondly, that aside from sensory deficits there are no exclusionary criterion, such as cultural disadvantage or low IQ. The contrasting definition and diagnostic criteria in the DSM demonstrates the importance of cognitive processing deficits. These are rarely skipped in the various definitions of LD, although these processes are often ignored in identification (Kavale et al., 2005), which tend to be based on more vague guidelines.

The Individuals with Disabilities Education Act of 2004 (IDEA) in the U.S defined LD as “a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which disorder may manifest itself in the imperfect ability to listen, think, speak, read well, or do mathematical calculations” (IDEA, 2004, 20 USC 1401 §30) However it is well accepted that in the U.S., that there are significant differences in diagnostic practices from state to state (Reschly & Hosp, 2004).

The importance of a nationally agreed upon definition of LD has been stressed in the Australian literature (Cunningham & Firth, 2005). However, there are important pragmatic considerations to consider in terms of policy and funding. In 2007, the New South Wales Government passed an Educational Support for Dyslexic Children Bill which requires dyslexia to be included within the Government’s disability criteria, thus making children with LD attending government schools eligible for extra assistance (“Educational Support for Dyslexic Children Bill,” 2007). However,
nationally there is no funding provided for students who are diagnosed with LD (Graham & Bailey, 2007).

Since the recent recognition of dyslexia in New Zealand, one suggested definition has been

   a. Persistent literacy learning difficulties
   b. In otherwise typically developing children
   c. Despite exposure to high quality, evidence based literacy instruction and intervention
   d. Due to an impairment in the phonological processing skills required to learn to read and write (Tunmer & Greaney, 2010, p.239)

One of the more interesting aspects of this particular definition is that a specific cognitive processing deficit, phonological processing, has been named. However aside from this, this particular definition could be considered fairly general in nature. In response to vague definitions of LD a new definition was put together by Kavale Spaulding and Beam:

Specific learning disability refers to heterogeneous clusters of disorders that significantly impede the normal progress of academic achievement in 2-3% of the school population. The lack of progress is exhibited in school performance that remains below expectation for chronological and mental ages, even when provided with high-quality instruction. The primary manifestation of the failure to progress is significant underachievement in a basic skill area (i.e., reading math writing) that is not associated with insufficient educational interpersonal, cultural/familial, and/or sociolinguistic experiences. The primary severe ability achievement discrepancy is coincident with deficits in linguistic competence (receptive and/or expressive) cognitive functioning (e.g., problem solving, thinking abilities, maturation) neuropsychological processes (e.g., perception, attention, memory) or any combination of such contributing deficits that are presumed to originate from central nervous system dysfunction. The specific learning disability is a discrete condition differentiated from generalized learning failure by average or above (>90) cognitive ability and a learning skill profile exhibiting
significant scatter indicating areas of strength and weakness. The major specific learning disability may be accompanied by secondary learning difficulties that also may be considered when planning the more intensive individualised special education instruction directed at the primary problem (2009, p. 45-46).

This definition is very specific, however there are points of contention. For instance no one with an IQ below 90 is eligible for diagnosis. Furthermore the three sets of deficits, in linguistic competence, cognitive functioning and neuropsychological processes, could still be considered too broad as no definitive criterion is outlined. The assumption that all children must be provided with high quality instruction could also be considered problematic, as this will not always be the case. This dictation of a scientific definition by pragmatic and political biases has been acknowledged and argued over for more than a decade (McPhail & Palincsar, 1998).

Unfortunately, calls for a definition that is not influenced by political or administrative demands may be unrealistic. For a definition to be successful, it has to be picked up by the wider community and be readily applicable for use in schools. Any definition or method of diagnosis that is too hard to implement will not be used and will thus be of no little or no benefit to anyone. This is especially true in Australia where there is no financial motive for identifying LD students, as there is no funding attached to diagnosis. This does not mean that research and debate over the definition and diagnosis should be led by more pragmatic needs, but that they should be taken into account. Even the utility of the existence of any kind of diagnostic term has been questioned. It has been argued that by differentiating between LD and remedial students you are merely disadvantaging those that are remedial (Spear-Swerling & Sternberg, 1998).

Arguments over the definition, or even existence, of LD have led to varying prevalence rates for the disorder. One Australian study found that 18% of students suffered from specific learning difficulty (McLeod & McKinnon, 2007). However, it should be noted that this was in the context of learning needs, not psychological diagnosis. In the US the lifetime prevalence of LD is thought to be 10% (Altarac & Saroha, 2007). A large scale longitudinal study, this time focusing on reading disorders specifically, found a 3-5% prevalence of reading-based LD (St Sauver,
It should be noted that the diagnostic rates of LD across American states are as, if not more, influenced by diagnostic practices in use as they are by the actual prevalence of LD (Lester & Kelman, 1997).

Prevalence studies are further confounded by conflicting studies on gender differences in LD. One found a gender ratio of 1.6-2:1 of boys to girls with LD (Chan, Ho, Tsang, Lee & Chung, 2007), however no gender differences have been found in another study (Undheim & Sund, 2008). Jimenez and colleagues hypothesised that certain methods of identification tend to result in the over-identification of boys, and that when identified in a research setting the ratio tends to be much more even (2010). Another population study found a gender ratio of 1.3:1 (Willcutt & Pennington, 2000) and attributed the different ratio they found in comparison to other studies to the fact that girls are less likely to be referred for LD type problems, as they are less likely to act out when struggling in class, not that they are much less likely to have them in the first place.

Share and Silva (2003) found that some of this gender difference can be attributed due to the IQ discrepancy method of identification. Boys’ scores have more variation, therefore they have higher representation than girls in both the highest and lowest scores. Consequently, they are more likely to have reading scores significantly lower than IQ scores. Once this confounding variance was controlled for, there were equal numbers of boys and girls with LD, with a comparable pattern of performance. Studies such as these demonstrate that one of the major issues for LD students is having a problem that is recognisable, not actually having a problem.

Interestingly, the persistence of LD is less controversial than the existence of the disorder. A large scale longitudinal study was conducted on a Victorian population (Smart, Prior, Sanson & Oberklaid, 2001) in order to ascertain the persistence of LD. It was found that LD was fairly persistent; 80% of those identified between seven and eight years old still had some form of LD at 13 to 14. Longitudinal studies can also attest to the persistence of LD throughout childhood and into adulthood (Undheim, 2009). This was confirmed in a large, longitudinal study conducted in Australia, 72% of the students identified as having LD in Grade 2, still had LD in Grade 6 (Waring, Prior, Sanson & Smart, 1996).
Results from the Connecticut longitudinal study are also sobering. Students who were identified as having a reading disability at the age of five were still lagging behind their peers into adolescence, even as growth in reading ability tapered off (Shaywitz et al., 1999). This has been supported by the recent finding that LD students make less progress in reading than their Regular Achievement (RA) counterparts (Judge & Bell, 2011). One longitudinal study found that students identified as LD had a slower rate of reading progress than their peers over a two year period (Katz, Stone, Carlisle, Corey & Zeng, 2008). It is of course possible that such issues could be a reflection of the Matthew effect, rather than the persistence of the disorder. It is also possible that both factors are contributing to this.

Despite long-term deficiencies, 16 year old LD students in mainstream education are more likely to aspire to higher education, employment and financial independence than other special education students (Casey, Davies, Kalambouka, Nelson & Boyle, 2006). Kortering and colleagues (2010) found that although there are no differences in the type of employment that LD and RA students aspire to, there is a significant difference in what they expect to end up doing. Aside from issues within the sphere of education, there are other serious adverse outcomes that are more likely for students with LD. Adolescents with LD have been found to have higher levels of reported depressive symptoms, school stress, lower grades and lower attachment to their parents than those without LD (Undheim & Sund, 2008).

**Nomenclature**

A lack of success in accurately identifying and defining LD is considered to be a major impediment to research in the LD field (Cunningham & Firth, 2005; de Fonseca, 1996). In 1986, the biggest problems that LD researchers faced were issues in defining and giving differential diagnoses of LD (Algozzine & Ysseldyke). It is possible that many of these nomenclature issues in LD research may stem from the expectations of how social issues will be dealt with in schools (Freebody, 2007). For instance in the case of LD, definitions are sometimes suggested that include anyone who may be struggling to gain literacy skills regardless of why. This is due to the well-intentioned motive of trying to give assistance to anyone who may need it. The problem with this of course is that it can then lead to a decrease in credibility for LD, which will in turn impact on the resources that are directed towards fixing the problem.
The debate within Australia about nomenclature continues, even though LD is not recognised by the federal government (Jenkinson, 2007). Some authors have tried to set out clear guidelines in order to facilitate more united and productive research, such as the previously mentioned Kavale and colleagues (2009) definition. However none of these issues have been picked up by the majority. For instance Spagna (1998) outlined some very clear and valuable marker variables of LD, that have not been used since. Many identification procedures do not take into account the heterogeneous nature of LD, as not all students are going to have all of the same markers. Therefore, there is a case for splitting the LD group into more homogenous groups.

This argument surrounding the splitting of an LD cohort into more specific groups, or lumping all of the similar groups under a wider umbrella, is one that is common to many scientific areas. The argument for both sides of the debate is fairly clear, split too much and you will end up with too many groups that will be difficult to keep track of and that will all require their own policy, research and treatments. Lumping groups of students together that are too different can result in faulty conclusions from research and policy that are not relevant to the whole cohort (Denckla, 1972a).

There are a wide range of proposed methods for splitting the heterogeneous group of LD students. For instance, dyslexia has been placed under the bracket of LD by a consortium of highly respected LD researchers (Lyon et al., 2003). Another group also found support for an auditory linguistic type, a visual perceptive type and a mixed deficit type of LD (Edwards & Hogben, 1999). Friedmann and Lukov found three subgroups based on interruptions at different points of the lexical route (2008). Other subtypes include letter position dyslexia (Friedmann, Dotan & Rahamim, 2010), where letters migrate within words and developmental attentional dyslexia, where letters migrate between words (Friedmann, Kerbel & Shvimer, 2010).

Another proposed group of definitions under the LD umbrella includes a definition of specific reading comprehension difficulties (being able to decode but not comprehend), dyslexia (being able to comprehend but not decode) and mixed reading disability (difficulty with both coding and comprehension) (Tunmer, 2008). As none of these has been widely accepted within the field, such distinctions tend to add to, rather than clarify confusion.
Many divisions within the field of LD research have been based on the differentiation between those students that have difficulty decoding and/or comprehending language. For instance, there has been a call made to differentiate between surface and phonological dyslexia (e.g., Stanovich, Siegel & Gottardo, 1997; Zabell & Everatt, 2002). In a comparison between the backgrounds of phonological and surface dyslexia, it was found that although phonological dyslexia had neurological origins, more of the aetiology of surface dyslexia could be accounted for by environmental factors (Castles, Datta, Gayan & Olson, 1999). This could of course bring into question the status of surface dyslexia as a neurological disorder. Shankweiler (1999) and colleagues found that a diagnosis of poor decoding was more robust than a diagnosis of poor comprehension. Research such as this points toward LD types incorporating a phonological deficit that is more robust and neurological in basis. Ultimately it is vital that the causes of and deficits behind LD be understood before an accurate system of classification can be devised.

**Aetiology**

The debate about nomenclature has lead to the reasonable objection that there may be no point to differentiating between poor readers and LD readers and as such there is no point to the construct of LD at all (Elliott & Gibbs, 2008). Others still have suggested that LD is nothing more than a different trajectory of learning in any given student and that given time those students identified as LD will catch up to their peers (Bender, 1963). This developmental lag theory has however been contradicted in longitudinal studies (Francis, Fletcher, Shaywitz, Shaywitz & Rourke, 1996).

This is why neurological studies focusing on LD are so important. By finding out more about the qualitative differences between LD and non-LD students, we can not only find out more about the problems that those with LD face, but also provide evidence for the very existence of LD. Genetic, psychobiological and neurological studies into the causes of LD provide continuing evidence that LD readers are qualitatively different to those who are simply low achievers.

The biological basis of LD has also been a point of contention with some researchers insisting that any biological difference is only at one end of a continuum and does not represent a qualitative difference (Spear-Swerling & Sternberg, 1998). However in a summary of research on brain imaging and LD it was found that there were
structural and functional differences between LD and RA readers (Hudson, High & Otaiba, 2007). It has been demonstrated that when students with LD and controls were compared on a reading task that was simple enough for both groups to perform well on, LD students utilised different areas of the brain than controls (Ingvar et al., 2002).

It has been demonstrated that when reading nonwords, LD student’s brain activation is qualitatively different to that of RA readers. LD readers were not so heavily dependent on the left hemisphere when phonologically processing words and instead had significantly higher activation in the right hemisphere when compared to RA readers (Bach et al., 2010; Pugh et al., 2001; Rumsey et al., 1999; Simos, Breier, Fletcher, Foorman et al., 2000). This reduced activation in the left hemisphere results in higher activation in the inferior frontal gyrus during reading and has been shown to not be due to dysfunction of this area of the brain (Simos, Breier, Fletcher, Bergman & Papanicolaou, 2000). A wide range of neuropsychological studies in LD have been reviewed in Johansson’s 2006 article.

Early individual differences in word reading are primarily heritable and the risk of inheriting impairments is continuous and not discrete (Harlaar, Spinath, Dale & Plomin, 2005). The increased likelihood of students with LD parents or siblings having LD themselves has been demonstrated in longitudinal studies (Muter & Snowling, 2009; Puolakanaho et al., 2007). Further evidence of the heritability of LD came from a study focussed on the children of those with LD (Gallagher & Frith, 2000). Although only 12% of the control sample was defined as literacy impaired, a full 57% of those whose parents had LD could be diagnosed with LD at six years of age. Furthermore, these children were showing verbal phonological language deficits at 45 months as compared to the more language based deficits that they displayed at age six.

The interaction of genetics and environment is also vital to this area of research (Berninger & Richards, 2010), as even in a neurological disorder, environmental factors still play a part (Robinson, 2002). Environmental issues in learning to read can result in a student mimicking the same problems that a genuine LD students has, this is a huge problem in LD identification. Vellutino and colleagues (1996) found that 67% of students identified as LD responded quickly to intervention and went on
to perform in a normal range. They interpreted this as resulting from the majority of people diagnosed with LD in this study actually struggling to read because of environmental disadvantage, not LD.

This confusion stems from the existence of the exclusionary criterion already discussed, namely low intelligence and socio-economic status. These criterion are in place because they could be responsible for LD-like symptoms, however it is difficult to pinpoint the cause of any given student’s reading problems. Therefore practitioners either need to not diagnose any of the students that meet these exclusionary criteria which is going to result in increased false negatives, or diagnose regardless, increasing false positives.

Genetic studies also indicate that LD has multiple causes that interact (Wood & Grigorenko, 2001). Having tested for, and finding genetic evidence for the heterogeneous nature of LD, some researchers recommend examining potential LD students at the individual level (Ziegler et al., 2008). Different chromosomes have been identified that seem to be at least partially responsible for some of the cognitive deficits in LD (Berninger, Raskind, Richards, Abbott & Stock, 2008). For a concise summary of what genes are thought to be involved in LD, please refer to Peterson and McGrath (2009).

Despite all this, the cognitive processing deficits thought to cause LD are usually not used in LD operationalisation and identification. Instead, as will be discussed in the next section, operationalisation is usually based on likely secondary outcomes of LD, such as a discrepancy between aptitude and achievement or an inability to respond to intervention. There are however some less popular methods of identification that do address these cognitive deficits such as the phonological and double deficit models, these will also be discussed in the next section.

**Operationalisation of Learning Disabilities**

Many studies that utilise LD samples do not clarify how they have operationalised LD, bringing the ability to generalise the results into question (Flanagan et al., 2006). As noted by Hoskyn and Swanson (2000), the inability to generalise within and between studies results in different papers allocating different participants into differently defined groups, even though all the results were subsequently compared.
The disparity in operational definitions of LD throughout the literature has slowed and confused progress in this particular area of research. It is difficult to base research on the vast number of papers available when LD has been operationalised in so many ways (Suhr, 2008). The wide range of options when operationalising LD confounds interpretation of results. Take, for example, a new study has results that contradict a previous study. It is a very real possibility that this difference is reflecting the different cohort of students that are selected to participate in the new study, rather than actually refuting the results from the previous study.

A comparison of four different models of LD diagnosis failed to find significant consensus between them (Proctor & Prevatt, 2003). It was found that although these four models diagnosed similar numbers of students in their sample, the percentage of corresponding diagnoses between groups was low, from 48-70%. This is concerning when different forms of identification are used constantly from country to country, state to state and school to school. Diagnosis may be just as dependent on the form of classification used as it is on the actual existence of LD. These differences in identified students also highlight the importance of understanding the most popular methods of identification used in the past, so that new results can be interpreted in light of the different groups of students identified. The rest of this section will look at some of the more common methods seen in recent academic literature, namely the discrepancy, RTI, phonological deficit models and the double deficit model. After this the tenability of a multiple deficit model will be discussed.

**Discrepancy**

In 1975, a large scale (n = 2300) study plotted the discrepancy between aptitude and reading achievement of all participants and a bump was found at the point in the graph where aptitude was significantly higher than reading achievement (Rutter & Yule). They attributed this bump to a group of students who were not reading at a level that would be dictated by their intelligence. The logic was that there were some students that were supposed to be showing poor reading achievement, the natural tail that occurs in the normal distribution, however those extra students reading at that low level, causing the bump, should not have been reading at this standard and the cause for this was proposed to be LD. In order to explain this bump they differentiated between poor readers who were reading at a level consistent with their aptitude and those who had unexpected under-achievement or a significant
discrepancy between IQ and reading ability. The logic used here is that those students with a significant discrepancy were not living up to their potential, as reading ability and intelligence are correlated in normally achieving students (Ferrer, Shaywitz, Holahan, Marchione & Shaywitz, 2010).

Those students who did have this significant discrepancy were thus designated as having specific reading retardation a term that has evolved into LD (amongst other terms) today. One of the most important aspects of this nomenclature was that LD students were separated from other poor readers, who did not have significant discrepancy between aptitude and achievement. This differentiation between these two groups led to the coining of the term garden-variety poor readers (Gough & Tunmer, 1986) to describe those students whose reading ability was low enough for LD diagnosis, but who did not have the high aptitude that was also required. However over time, this term has been replaced with Low Achievement (LA), providing a neat contrast to the LD terminology. This differentiation has been supported by a meta-analysis that found cognitive processing differences particularly in tests that measured automaticity, between LD and LA students (D. Fuchs, Fuchs, Mathes & Lipsey, 2000).

It has been found that IQ can explain a significant amount of variance in reading comprehension, even after listening comprehension, phonological decoding and naming speed are controlled for (Tiu Jr, Thompson & Lewis, 2003). This was taken as support for the idea that IQ could predict reading achievement and as such a discrepancy between the two could be considered abnormal. However it could be argued that this variance could be explained by other cognitive processes that are important to both reading and IQ, such as working memory. If a given cognitive process was important to both LD and IQ, then a correlation between the two could be expected (Siegel, 1989).

Stanovich (1991) argued that the verbal ability is a confounding variable as the aptitude measures used for LD diagnosis often have a verbal component. This argument is usually centred on the idea that the predictor of reading achievement, aptitude, should be strongly correlated with achievement, otherwise a discrepancy between the two is logically not cause for concern. However testing for a discrepancy between two tests that are effectively testing the same construct is
problematic. Although there have been arguments against using a non-verbal measure of aptitude, it is clear that a verbal intelligence measure could be affected by reading ability. This could lead an LD child being diagnosed as LA, as scores on the aptitude measure would decrease as reading ability decreased (Stanovich).

There are many ways in which significant discrepancies are defined: the most common of which is a difference of a standard deviation between aptitude and achievement, however the exact magnitude of the required discrepancy can vary. This inconsistency is cause for concern when the finding that the degree of the discrepancy between aptitude and achievement in the LD sample does not predict cognitive performance or deficits (Hoskyn & Swanson, 2000) is taken into account. It could be argued that is the existence of a discrepancy, not the magnitude of a discrepancy, which is important in LD identification. However even if this explanation was accepted there are implications of results such as these given the arbitrary nature of a one standard deviation discrepancy, or any other, arbitrary cut off.

Another discrepancy method involves identifying a given percentage of students with the greatest difference between aptitude and achievement (Shepard, 1980). However this was never widely accepted, presumably because it would not take the range of abilities of any given classroom into account, while not actually solving any of the issues with a more traditional discrepancy method. In another method both a minimum score for IQ and a maximum score for reading achievement must be met for an LD diagnosis (Gang & Siegel, 2002). This can of course lead to those students on the cusp of both criteria to have extremely similar aptitude and achievement. Furthermore, it can miss high and low aptitude LD students.

In the traditional method of identification, regression to the mean is the underlying cause of many problems with the method. Statistical probability dictates that when a person is given two normally distributed tests that correlate positively, the expected score on the second one, based on the score of the first, will always be a little closer to the mean than the score on the first. The expected value of reading achievement is always going to be closer to the mean than the attained score on aptitude. Therefore, it is statistically dictated that more students with high IQs are going to have a significant discrepancy between aptitude and achievement than students with low
IQs, regardless of the actual presence of LD (Dyck et al., 2004; C. Reynolds, 1984; Van den Broeck, 2002a).

In order to counteract some of these problems some models have been developed that do allow for this regression to the mean. This is done by accounting for the correlation between the aptitude and achievement tests used. This is because the likelihood of a discrepancy between two tests being found will decrease as the correlation between the two increases, as a function of the absolute score on the first test (Dyck et al., 2004). By factoring in the less than perfect correlation between the two tests it is possible to counteract the problems caused by regression to the mean and the subsequent under-identification of low IQ LD students.

It should be noted that in regression-adjusted discrepancy, the difference scores are not calculated from the difference between aptitude and achievement, but rather aptitude and achievement as it is predicted by aptitude (Evans, 1996). Efforts have been made to increase the accuracy of regression-adjusted discrepancy with formulas developed to account for more than one test (Evans, 1996). That this correction is required has led to the theory that the discrepancy criterion is invalid in the first place (Cahan et al., 2010).

In a comparison of a prediction model, true score regression model and another statistical model, pre-score partialling, it was found that the regression model was not only the most statistically sound but also the most user friendly (Willson & Reynolds, 1984). It should be noted that none of the three were particularly user friendly at the time. However the more confusing aspects of regression-adjusted discrepancy could now be easily programmed into a database. Although this regression method has been touted as a solution to some of the psychometric problems that the discrepancy method entails, this too has been thought of as invalid as it defeats the very purpose of a discrepancy in the first place (Van den Broeck, 2002a, 2002b) although others have defended its use (Willson & Reynolds, 2002).

Although there has been much written on the drawbacks of the discrepancy method of identification, as will be discussed in the next section, it is still one of the most common methods of LD identification and has in that respect stood the test of time. Furthermore, it has been stated that although research into alternative methods should continue, discrepancy is the best current method of identification of LD
(Scruggs & Mastropieri, 2003). However a large amount of research has been dedicated to discrediting this model of identification, this will be reviewed in the next section.

**Criticisms of the discrepancy method of identification**

In a review of literature, Lyon and colleagues (2004) concluded that using discrepancy as a primary identification tool of LD can pose more threat of harm to students than it can increase the chance of benefit. In the past decade, there have been many articles that have been very forthright in their belief that discrepancy is not sufficient for LD identification (Bradley et al., 2005). The extent of frustration with the aptitude-achievement discrepancy becomes apparent in articles primarily dedicated to discrediting the theory (e.g., Stanovich, 2005). However, the lack of difference in type of cognitive deficits in discrepancy-defined LD and LA students (Ellis, McDougall & Monk, 1996b; Fletcher et al., 1994; Share, 1996) and the lack of strong correlation between reading and intelligence (Vellutino, Scanlon & Lyon, 2000) indicates that much of this scepticism is justified.

Although some studies still utilise the discrepancy method (e.g., Johnson et al., 2010; Stoodley, Ray, Jack & Stein, 2008) arguments against discrepancy diagnosis have been made not just on scientific but more pragmatic, policy and fiscal grounds (Tanner, 2001). One of the most common is that aptitude and achievement measures alone cannot provide enough information to facilitate remediation (Silver et al., 2008). Furthermore, Aaron (1997) found that remedial reading programs were on the whole as effective for those with LD as they were for students without LD. As such, the authors argued that the method of identification was incorrectly identifying students and that the method as a whole should be discarded.

Of the major criticisms of the discrepancy method of identification, most can be summarised into a few broad groups. Firstly, the use of IQ tests is very controversial, not just in this area of study but in many others. Secondly, the psychometric issues inherent to difference scores also make the discrepancy method statistically problematic. The regression-adjusted discrepancy, utilised in order to address some of these psychometric issues has also met with a new wave of criticism. However the biggest issue with discrepancy method is that there is no real difference between LD and LA students. The failure to differentiate LA students
from LD students on the basis of cognitive processing deficits could be considered the primary reason for the fall from favour of the discrepancy method. These three broad criticisms, based on the use of IQ tests, difference scores and the lack of difference between LD and LA students, are discussed below.

**IQ tests**

It is common for articles arguing against the discrepancy method of identification to base their arguments on the inadequacy of the IQ test in not only predicting those that should be able to read but also as a test of intelligence (Spear-Swerling & Sternberg, 1998). A good example of the criticism directed at the discrepancy method of identification can be found in Gustafson and Samuelsson’s (1999) review article. Although the fuzziness of the concept of intelligence is discussed in the article, the focus is on the inability of IQ tests to accurately predict intelligence as well as the unestablished causality between IQ and reading skills.

Four primary issues with the use of IQ tests highlighted back in 1989 (Siegel) are still as relevant now as they were then, each one based on what was thought to be a faulty assumption. First, that IQ is a valid measure of intelligence; second, that intelligence and reading achievement are independent; third, that intelligence predicts achievement and, finally, that LD students will have cognitive deficits that will vary with their intelligence levels. All four of these assumptions are vital to the theory of the discrepancy method of identification, yet none have received much empirical support.

Stanovich (1999) launched a scathing attack on the discrepancy method of LD identification, however once again his criticisms were directed at the IQ test and not the method in general. One of the more psychometrically convincing arguments against the use of IQ tests in discrepancy diagnosis is that the IQ tests by themselves have so many psychometric flaws. Most pertinent to LD diagnosis, it has been shown that there are confounding factors in IQ scores based on both gender (Share & Silva, 2003) and race (Brooks-Gunn, Klebanov & Duncan, 1996; McLesky, Waldron & Wornhoff, 1990). This can result in students who are in groups more likely to receive a low IQ score becoming less likely to be diagnosed with LD and instead being identified as LA, when LD may be present.
These cultural biases of IQ testing also bring the cultural disadvantage exclusionary criterion into question. Not only are black students in America more likely to have a lower IQ, but also that this difference can be nearly fully accounted for by taking into account socio-economic status (SES) and neighbourhood poverty (Brooks-Gunn et al., 1996). Consequently, some students have less opportunity, socio-economically speaking, which is limiting their IQ, which in turn is limiting their chance of receiving assistance for their reading difficulties, which may be stemming from a lack of assistance in the first place. It seems illogical not to give financial assistance toward the education of those in low socio-economic areas, on the basis of their socio-economic status holding them back. However, this is effectively what happens, at least in countries with funding attached to an LD diagnosis.

Problems with the use of IQ tests in LD diagnosis include issues with not just different types of intelligence tests but also different editions of the same tests. For instance, students sitting the WISC-R intelligence test received scores five to six points lower than if they had sat the WISC-III, increasing the chance of them receiving an LD diagnosis (Schultz, 1997). This is an example of the Flynn effect in which IQ scores are increasing over time (Flynn, 1987). This could in theory result in a significant minority of students who would have been eligible for an LD diagnosis when using the WISC-III, becoming ineligible if the WISC-R was used instead.

More pragmatic issues involved in any kind of large scale screening involving IQ tests has led to a call for shorter intelligence tests that are less time consuming (Donders, 2001). This is due to the time and labour required for each IQ administration. Not only is IQ testing time consuming and expensive, but it also does not provide information needed for remediation. IQ testing has been described as a waste of time and money as the suggestion for remediation, regardless of IQ, is to simply begin a reading intervention and track the student’s progress. If the IQ test does not result in differing prescribed courses of action, then it could be easily argued that there is no point to it (Willis & Dumont, 2006).

The assumed knowledge inherent in the verbal IQ section has been often commented on and the need for tests of verbal ability without this problem has been outlined (Langdon, Rosenblatt & Mellanby, 1998). It has been pointed out that the subtests
on IQ tests were selected before the bulk of the discoveries about cognitive processes were made and are thus based on outdated research if any at all (Ardila, 1999). Many of the original questions in the first IQ tests were based on teacher recommendations of questions that poor learners would struggle with (Ceci, 1991). Not only is this a disturbing indicator of the circular logic inherent to intelligence tests (Ardila) but also has further ramifications specific to the diagnosis of LD, if it is made on the basis of a discrepancy between school performance and a test of aptitude that was originally based on school performance.

Tanner (2001) highlighted some of the major problems with utilising the aptitude achievement discrepancy as the sole diagnostic tool for identifying LD. Primarily, the shared variance between any two given intelligence tests can vary, resulting in the uncomfortable truth that in some cases, diagnosis of LD is dependent on the intelligence test chosen. With differences stemming from different individual tests, the use of a discrepancy between an intelligence test and any possible test of achievement can be seen as a precarious assessment tool.

Finally the validity of overall IQ scores has been questioned as more than 80% of students in one study had variable subscale scores that are then used to contribute to an overall score (Fiorello, Hale, McGrath, Ryan & Quinn, 2002). If an overall score is not suited to the majority of students, then a discrepancy between this overall score and another test score is a tenuous basis for a neurological disorder. Furthermore IQ scores have long been thought to be less accurate as IQ rises (Spearman, 1927) once again bringing in to question the ability of the discrepancy method to identify high IQ LD students.

**Difference scores**

LD diagnosis, via the discrepancy method, is based on a discrepancy, a difference score between tests of aptitude and achievement. Not only does this method has its own statistical issues that include the same problems with arbitrary cut-offs that most tests have, but also problems that are unique to difference scores as well. With normal test scores reliability is important, in difference scores abnormality must also be taken into account. Although a reliable difference is one that is unlikely to have occurred by chance, an abnormal difference is one that is so large that it would only be obtained by a relatively small proportion of the population (Silverstein, 1981).
This is an important distinction to make, as a reliable difference is not necessarily anything of note. It may instead be a relatively common difference score in a given population and thus not indicative of anything other than a normal pattern of scores.

Statistical issues with the use of discrepancy scores in LD diagnoses were highlighted a quarter of a century ago (C. Reynolds, 1984). Firstly the number of children that will be identified with LD is dependent on the correlation between the two tests used and their respective standard deviations. For instance, if using the Verbal and Performance IQ scores from the WISC, 12.5% of students will have a performance score one standard deviation higher than their verbal performance score. What is important to note is that this 12.5% is statistically dictated, regardless of the presence of LD. The importance of reliability in two tests being used also increases if the aim is gauge a difference between the two. Although test reliability is very important in any kind of diagnostic tool, it is even more important when using an abnormal difference between two scores as the primary indicator (Cotton, Crewther & Crewther, 2005).

Another criticism levelled at the use of difference scores in LD screening or diagnosis is that a similar discrepancy can be found in children with sensory or behavioural disorders (O’Donnell, 1980). Similarly, it would also be expected in students who have not overcome cultural disadvantage, this is why these were made exclusionary criterion in LD diagnosis. However, the problem with this is that although these disorders provide an alternative explanation for a difference score, the two explanations are not mutually exclusive. There is no reason why a student with cultural disadvantage could not have the cognitive processing deficits inherent to LD. Therefore many students designated LA on the basis of cultural disadvantage may actually have LD.

*Learning disabilities and low achievement*

If a discrepancy is indicative of LD then, logically, there should be qualitative differences resulting from the existence of a neurological disorder, between those designated LD and LA. The critique often put forward by critics of the discrepancy method is that there are no differences between LD and LA students. In 2000, Swanson posed the question "Which cognitive processes mediate the emergence of discrepancies and nondiscrepancies in low achievers?” (p. 39) and few studies have
been able to find a replicable answer. However, studies have found neurological (Decker, 2008), and cognitive (D. Fuchs et al., 2000), differences between the two groups.

However, when reviewing the research in this area as a whole, there is resoundingly little evidence for cognitive processing differences between the two (Stanovich & Stanovich, 1997). Some researchers have concluded that any differences are just a reflection of the IQ differences between the two groups and not related to reading or LD (Fletcher et al., 1998).

The social justice issues inherent to the assumption that someone with a high IQ deserves to have extra assistance with their reading difficulties has also been questioned (Stanovich, 1999). Furthermore, the impact of reading ability on intelligence also needs to be taken into account, as it could be responsible for changing the classification of any given student over time. For instance, a young LD student has an intelligence advantage over an LA counterpart, however as time progresses and the Matthew Effect takes hold, this advantage diminishes. Over time the chance of that LD student being identified as LA increases as a function of the decreasing discrepancy between aptitude and achievement.

In order to examine the differences between LD and LA students, Share (1996) matched participants on age, sex and word recognition skills, but with IQ differences that dictated their membership in the LD or LA groups. Any student whose reading comprehension or word recognition fell 1.5 standard errors below the achievement score predicted by their IQ were identified as LD. The two groups of students were then tested on 26 measures of reading ability. Only one test was significant at an alpha level of .001, leaving the author to conclude that there was no significant difference between the two groups. When Ellis, McDougall and Monk (1996a) found that there were no significant cognitive differences between LD and LA students, they received critical responses with some researchers doubting their results (Snowling & Nation, 1997). Ultimately Ellis and colleagues had to clarify their views; they were not stating that LD students did not have deficits, merely that LA students had them as well (Ellis, McDougall & Monk, 1997).

A more specific examination of the differences between LD and LA students, by Fletcher and colleagues (1994) identified eight groups of students. Six of these were
identified as LD or LA in one or both of arithmetic or reading. Another group was an Attention Deficit Hyperactivity Disorder (ADHD) comparison group and the final group consisted of children with IQs in the 60-79 range. Furthermore, students were identified as belonging to groups of traditional discrepancy, regression-adjusted discrepancy, LA definition, and anyone in the bottom 25% of reading scores. There was little evidence to suggest that there was a need to split LD and LA groups on the basis of cognitive differences. The authors concluded that traditional discrepancy is not the best method for identifying students and that another type of discrepancy could be better.

In a similar study, five groups of students were compared: LD via traditional discrepancy, LD via regression-adjusted discrepancy, LA, those who met the first two criteria, and those who met none of the criteria. These students were tested on ten neuropsychological tests and once IQ was accounted for there were little if any difference between groups (Fletcher, Francis, Rourke, Shaywitz & Shaywitz, 1992). However, it should be noted that four of these tests were language based and of the other six tests, none were directly testing phonological processing, naming speed or verbal memory.

Bone, Cirino Morris and Morris (2002) compared LD and LA undergraduate students using a regression-adjusted discrepancy method of LD identification. Both groups had similar phonological processing deficits however the LA group had additional deficits in naming speed and fluency. This finding is very interesting in that it is assumed that the LD cohort is facing processing deficits that the LA group does not possess and that these deficits are the cause of their discrepancy. However, in this study the LA group was actually more impaired than the LD group.

Vellutino, Fletcher, Snowling and Scanlon (2004) compared students who were actually eligible for membership in both the LD and LA groups to those that only fit into the LD group (high aptitude LD students), those that only fit into the LA group (low aptitude poor readers) and finally a control group. The LA only and LA+LD groups had significantly lower phonological processing than the control group and the LD only group, indicating that phonological processing is something specific to poor readers, not discrepant ones. It is studies such as these that have led to a call for
research that actually differentiates LD and LA students, or failing that, the
differentiation should cease.

A recent meta-analysis also found that there were no differences in measures of
literacy development, including phonological processing, between discrepancy-
defined LD and LA groups (Gresham & Vellutino, 2010). The inability to find
differences between LD and LA students has not resulted in a total dismissal of the
discrepancy method of identification. Despite finding no phonological differences
between LD and LA groups, Hoskyn and Swanson (2000) did not recommend
abandoning the use of discrepancy without further investigating the utility of the
distinction between the two groups. However, many of the reviewed studies are
indicating that discrepancy-defined LD and LA students have similar deficits and
respond to treatment in a similar manner, and that this could lead to the conclusion
that there is no point to utilising the discrepancy method of identification.

With all the issues inherent to discrepancy method of identification, especially the
difficulty in finding cognitive processing differences between LD and LA students,
the process of identifying a particular cohort of poor readers was questioned. Instead
it may be best to identify those students that are having the most trouble with
reading, regardless of aetiology and designate them LD. The issues surrounding this
theory of LD identification is discussed in the next section.

**Low Achievement**

One of the more popular alternatives to the discrepancy method of LD identification
is to use a percentile cut-off, for example the bottom 10%, and flag any students who
are under this cut-off as LD. One paper compared intra-individual discrepancy with
a simple low achieving cut off and a more specific low achievement cut off that was
calculated per school or per country or state (Peterson & Shinn, 2002). In this study,
the latter method of identification was found to be the most effective out of the three.

In a group of students with specific reading disability, as defined by low
achievement, measures of speech perception were tested and it was found that
although 70% of these children showed similar results to controls, the remaining
30% were significantly different (Adlard & Hazan, 1998). In theory, this could be
interpreted as the ‘real’ LD group making up the 30% of students showing
qualitatively different results, however it is not possible to be sure. Reports such as these highlight the difficulty that is caused by the different identification methods.

There are two major theoretical issues with this method. First and foremost, there is no getting past the undeniable fact that one in ten students is going to be under a 10\textsuperscript{th} percentile cut-off, regardless of the presence of LD. In fact, all this method does is identify remedial readers, nothing is done to identify those who have a neurological disorder. It is likely that many of those identified are low achieving readers and that on the flip side many bright children with LD who have found effective methods of compensation, appropriate to the time of testing, not necessarily for life, will go undetected. Therefore there does, on face value at least, seem to be an unacceptable level of both false positives and false negatives with this method.

Issues surrounding LD becoming synonymous with underachievement, or developmental delay, due to putting in place a simple reading achievement cut-off as a means of diagnosis, have been explicitly raised (Mather & Gregg, 2006). One of the primary concerns is that the dissolve of the boundary between LD and LA students will result in the recently gained acceptance of LD as a neuropsychological disorder being undermined. Every student within a certain percentile range on reading achievement cannot be diagnosed with neurological disorder, without that disorder no longer being taken seriously. As stated by Kavale and colleagues, “What should not happen is a designation of LD for a slow learner; the LD concept should not be sacrificed to resolve a long standing school problem” (2005, p.5).

In an effort to discriminate between poor readers with a neurological disorder and those struggling for other reasons, without resorting to discrepancy methods of identification, it was suggested that response to instruction be utilised as an alternative method for discriminating between the two groups (Vellutino & Scanlon, 2002). RTI has since become a very popular method for identifying LD students, and research on this method will be reviewed in the next section.

**Response to Intervention**

The first step of Response to Intervention (RTI) basically involves ensuring that students receive effective literacy teaching. After identifying those that are struggling to respond to this literacy program, students are given more explicit training as an intervention. Those who do not respond to this, and possibly further
interventions, are assumed to have LD (D. Fuchs, Mock, Morgan & Young, 2003). The idea is that by exposing all students to these interventions it will be possible to rule out environmental causes for reading difficulties (Gelzheiser, 1988). Because of this, there are often different theories about how the effectiveness of interventions should be assessed, for instance with a simple benchmark, by the progress made throughout the intervention or a combination of the two (L. Fuchs, 2003).

Due to the wide range of variations possible in the implementation of this framework, there have been calls made for a unified model of RTI. One proposed model provides an apt example of how RTI could be implemented in a school setting. The first tier, primary prevention, should be administered by general educators. The second tier, secondary prevention, consists of small group instruction sessions of 20 to 40 minutes, three to four times a week, for 10 to 15 weeks. The prescriptive nature of this step dictates, in theory, that it can be conducted by a paraprofessional. Those not responding to this level would then go on to the third tier, tertiary prevention, that should be government funded (L. Fuchs & Fuchs, 2009). The confusion inherent to the different tiers at the moment has been acknowledged by proponents of RTI. However they maintain that this method is more effective than the discrepancy method at identifying those students with a neurobiological disorder (L. Fuchs & Fuchs).

In support of this, one longitudinal study found that when all students at-risk for reading disabilities are given assistance, those that are most resistant to intervention scored lower on individual difference measures (Case, Speece & Molloy, 2003). Simos and colleagues (2007) examined students who did not respond to Tier 1 intervention in an imaging study and found that they had qualitatively different brain activation to normal readers. After more intensive intervention, the responders had similar brain activation patterns to controls. Furthermore, although non-responders showed different patterns there were indications that they had developed compensatory behaviours that were not as effective as the patterns displayed by responders and controls (Simos et al.). A meta-analysis of students who did not respond to intervention found that the majority of these students had low phonological processing skills and performed more poorly on tests of naming speed than their responding counterparts (Al Otaiba & Fuchs, 2002).
Another advantage of RTI is that it is quite proactive; the testing involves providing extra assistance to those who need it, and is more practical than ‘wait to fail’ models such as discrepancy (Vaughn & Fuchs, 2003). Another perceived benefit is that RTI would provide intervention for all students struggling with poor reading skills not only those with LD. However, the big perceived advantage does seem to be that the biases associated with IQ testing, required for the discrepancy method, are removed from the equation.

It has been stressed that there are many questions yet to be answered about this method of identification (D. Fuchs & Deshler, 2007). Yet, RTI has received support on the educational front lines. Machek and Nelson (2007) found that the majority of school psychologists surveyed endorsed an RTI approach to LD diagnosis. With RTI being accepted as a mainstream method of diagnosis, the next issue was how a student’s response to intervention should be measured (D. Fuchs & Fuchs, 2006).

A recent article concluded that RTI could provide an ethically and legally sufficient method of LD assessment, however the authors concluded with the warning that the relationship between LD and non-response needs to be more carefully analysed (Burns, Jacob & Wagner, 2008). How RTI could be considered sufficient in LD diagnosis, without this relationship already thoroughly investigated and determined, is unclear. With multiple interventions used as part of the interventions designed to identify students, the questions of what is expected to occur once a student is identified was also asked (Reschly, 2005).

Many researchers have stressed that the implementation of RTI does not take the place of clinical evaluation and note that “the failure to have objective, repeatable criteria that clinicians can understand and apply will lead inevitably to chaos in the field” (C. Reynolds, 2008, p.14). However, some argue that the success of RTI has done away with the need for neuropsychological testing at all (Hiscock & Kindsbourne, 2008), or that cognitive processes are not relevant to the identification of LD (Fletcher & Vaughn, 2009a). An example of a cognitive assessment free method of RTI identification includes an age-based performance discrepancy combined with an age-based growth rate discrepancy in response to an intervention (McMaster, Fuchs, Fuchs & Compton, 2005). These tests would of course face
many of the psychometric issues that led to RTI proponents to dismiss the discrepancy method in the first place.

RTI opponents have pointed out however, that at the very least a summary of strengths and weaknesses as would be assessed by testing, is required for determining the best kind of intervention, and that actual diagnoses can not be achieved using RTI alone (C. Reynolds & Shaywitz, 2009b). Troia suggested that the judgement of speech pathologists in conjunction with RTI should be the basis for LD diagnosis (2005). Another suggestion was the use of measures of growth made at intervals throughout intervention (Speece, 2005). Although this would be thorough and informative, there is of course the need for multiple tests, something that is not easily put in place in the classroom.

Another suggestion for assessment in RTI is dynamic assessment (Grigorenko, 2009). This involves a teacher being free to give feedback to the student they are testing, as they are assessing the process that students use to get the answer rather than the answer itself. This involves repeated one-on-one testing that may not always be feasible, especially in public schools (Caffrey, Fuchs & Fuchs, 2008). It has been suggested that dynamic assessment will reduce the amount of students that have to wait until their reading skill decreases sufficiently in order to be picked up for Tier 1 assessment (D. Fuchs et al., 2007). However in order to gain this benefit, wide scale dynamic assessment would be required, once again probably not feasible in a public, or even private, school system.

There have however, been calls from some prominent RTI proponents for the use of standardised measurement in conjunction with RTI (L. Fuchs & Fuchs, 2006). Most have conceded that RTI cannot take the place of evaluation and that some kind of assessment tools would be required even in a setting where a full RTI system was employed (Bradley, Danielson & Doolittle, 2007). A two stage screening process has been suggested, where the first gate is a screening tool that identifies those that might be at-risk and the second gate is the response to intervention itself (Compton et al., 2010). This of course leads to the question of what the first screening tool would consist of.

One suggestion is neurological testing that occurs in conjunction with Tier 3 intervention (Schmitt & Wodrich, 2008). Another is for curriculum-based
measurement that complements RTI in which teachers analyse the rate of improvement of individual students and then decide what changes in instruction are necessary (L. Fuchs, Fuchs & Hamlett, 2007). Once again, this would be extremely time-consuming as well as not differentiating between those with or without a disorder.

Cognitive hypothesis testing has also been suggested as a solution to the assessment difficulties in RTI (Fiorello et al., 2006). Some researchers have pointed out that cognitive assessment and RTI are not mutually exclusive and that the two could be complementary (e.g., Anastasiou & Polychronopoulou, 2009; Fiefer, 2008; Hale, Kaufman, Naglieri & Kavale, 2006; Holdnack & Weiss, 2006; Hughes, 2008; Willis & Dumont, 2006). However, a student who does not respond to the first tier of intervention would then receive a cognitive assessment. Consequently, a student needs to perform poorly in order to receive initial testing. Furthermore it would be less time-consuming and costly to use the test as a screening tool instead of the intervention.

This idea of using cognitive testing in conjunction with RTI brings up troubling issues about the role of RTI as a diagnostic tool. First, if testing cognitive processing anyway, then why is RTI needed as anything other than a method of allocating resources? Secondly, if a student does not respond to intervention, once the diagnosis on the basis of cognitive processing has been made, should the failure to respond be attributed to the efficacy of the training program rather than the abilities of the student? Ultimately the biggest issue with RTI as a diagnostic tool is that there is so much debate, even among prominent proponents, of what diagnostic methods can accompany its use. If a method of assessment is required in conjunction with RTI, and this has not yet been decided, then how has RTI solved the problem of how to assess for LD? This and other criticisms of RTI are discussed in the next section.

**Criticisms of response to intervention**

First and foremost it should be noted that the process of RTI, as a method of allocating resources and training to students who need assistance is rarely criticised. Instead, it is the use of RTI as a diagnostic tool that is controversial. C. Reynolds and Shaywitz (2009a) in an article critical of the use of RTI based many of their
criticisms around the assumptions inherent to RTI: namely that all LD students will not respond to intervention and that all students without LD will. In a comprehensive review of RTI as a diagnostic tool, C. Reynolds conceded that RTI could be useful in reducing the number of false positives that are referred for diagnostic testing if testing was still the final diagnostic procedure (2008). However, if RTI was to be used as the sole tool of identification, then he was concerned that it will under-identify high aptitude students, and over-identify low achievers. Furthermore, he stated that the tenets of RTI ignore the processing deficits that are inherent to the disorder and that model assumes that all classrooms are receiving up-to-date and scientifically validated teaching programs.

Overall the criticisms of RTI as a diagnostic tool can be divided into two groups, theoretical and applied. Theoretical drawbacks to RTI mainly centre on the assumptions that must be made in order to accept RTI’s effectiveness, and applied drawbacks to RTI involve the heavy costs in both time and money that would be required if RTI was to be implemented as a diagnostic tool for LD.

*Theoretical criticisms of response to intervention*

One of the primary reasons for the emergence of RTI was that discrepancy was thought to be inaccurate as a method of LD identification and/or diagnosis. One of the main reasons for this was that LA and LD students were thought to respond to remediation in the same way (Wheldall & Pogorzelski, 2003), thus making the need to differentiate between the two a moot point. However, RTI is based on the theory that it is only the LD readers that will not respond to intervention, which is a troubling misstep in logic.

Within an RTI framework, the diagnosis of LD does not come until after intervention. If a child receives an educational intervention that is effective for most students and does not benefit from it, then this child is designated LD. There are a multitude of issues with this assumption. First of all there is circular logic inherent to the assumption that the training program will be effective for all students who do not have LD, when LD is diagnosed on the basis of a failure to respond. Secondly, the assumption that all students would respond to training, if it was not for a disorder, in the same way “is contrary to 100 years of science associated with education, individual differences, and knowledge of pedagogical outcomes” (C.
Finally, the aim of early intervention of students with LD is surely based on the assumption that there is a point in identifying them and that there is a training program that will be able to assist these students, despite their neurological disorder. If LD students are immune to quality intervention, a primary assumption of RTI, then there is little point in identifying them. Finally, and most importantly, none of this has anything to do with the cognitive processing deficits that differentiate LD students from their low achieving counterparts (Ofiesh, 2006).

There are two aims that RTI endeavours to meet (Fletcher & Vaughn, 2009b): the first of these is to prevent and remediate both academic and behavioural difficulties, which is rarely, if ever, contested. The second aim however, providing relevant information that could be utilised for LD diagnosis has not been met. Although the hit rate of identifying those who need special education will be good, the hit rate of accurate LD identification will not even be calculated, as there is no way of ascertaining the true positive (Hale et al., 2010; Kavale, Kauffman, Bachmeier & LeFever, 2008; Kavale & Spaulding, 2008). It has been suggested that RTI can only diagnose low achievement in reading, not LD (Kavale et al., 2005).

Although RTI does circumvent some of the issues in early identification that are faced by the discrepancy method, it has been argued that the early providence of intervention could result in the delayed identification of LD students (Torgesen, 2009). This is because it is only those students who are displaying low achievement in reading that are picked up for the first tier of intervention. Lipka, Lesaux and Siegel (2006) demonstrated that the reading difficulties of many students are not apparent until the fourth grade. Furthermore, 36% of the students identified as poor readers in the fourth grade had above average reading scores from kindergarten all the way through to the third grade. They also found a subgroup of students who had low reading scores in kindergarten who, without intervention, had average scores by the time they reached the fourth grade. These results highlight the dangers of identifying students on the basis of low achievement alone, which is how the first tier students, and therefore any students eligible for subsequent tiers, are identified through RTI.

Results such as this have led to concern about the identification of gifted LD students within an RTI paradigm (McKenzie, 2010), especially those whose
strengths in other areas may serve to mask their LD enough to keep them out of the low achieving cohort (Volker, Lopata & Cook-Cottone, 2006). One of the primary concerns here is how much a student’s progress would have to decline before they could be picked up and given assistance, and how much this wasted time could damage the education of these students (Mather & Kaufman, 2006).

The poor track record of utilising a response to treatment, or lack thereof, in medicine has been noted as an indicator of the inappropriateness of RTI in educational psychology (C. Reynolds, 2008). The assumption that it is only the LD students that do not respond to remediation has been challenged (e.g., Hale, 2006). As noted by Berninger and Holdnack, “failure to respond to instruction simply signals that another approach to instruction should be brainstormed, implemented, evaluated, and if necessary, revised” (2008, p.76). A direct additional empirical test of how failure to respond to intervention could be measured also seems to be lacking (Scruggs & Mastropieri, 2002).

Recent studies have demonstrated that LD students can be responsive to intervention (Swanson, 2008a). Both comprehension based and fluency based remediation programs have been found to be beneficial to students with LD (Vaughn et al., 2000). Another remediation program was found to be more beneficial for LD students than students with mild mental retardation (Caffrey & Fuchs, 2007). Even an aphasic patient with deep LD has responded to intervention, albeit not fully (Davies, Cuetos & Rodriguez-Ferreiro, 2010).

Neuropsychological studies also support the claim that LD can be remediated (Shaywitz & Shaywitz, 2005). Intervention in LD students that had abnormal brain patterns pre-intervention, resulted in the same students showing similar brain activation to controls after intervention (Richards & Berninger, 2008). Studies such as this one are particularly important, as they demonstrate that those students with neurological deficits, genuine LD, can not only be remediated in terms of a testable response, but also that with training they can process information in a manner similar to that of their RA peers.

Berninger (2000) developed a successful LD treatment program. The seven participants in the study were selected on the basis of test scores and abnormal brain activation during phonological processing. After moderately intensive remediation,
the students no longer had different processing patterns as compared to their matched controls when phonologically processing information. A brain imaging study that tested adults with LD before and after phonological training, demonstrated that remediation can affect the activation of the bilateral parietal and right perisylvian cortices (Eden et al., 2004). These results indicates that the assumption that LD is fully resistant to intervention may be incorrect, although longitudinal studies would be needed to confirm that these changes remain consistent over time.

The flip side to the assumption that LD students cannot respond to appropriate intervention is the assumption that only those students without LD will respond to intervention. For instance, it has been found that problem behaviour is a significant predictor of a child’s response to intervention (Al Otaiba & Fuchs, 2006). This provides an alternative explanation for failure to respond to intervention beside LD.

It could be said that the theoretical concerns with RTI are somewhat balanced by the fact that RTI was designed for implementation in school, unlike many other theories that stem from academic research that are not suitable for an applied setting. However, there have also been problems of an applied nature with the implementation of RTI, these are discussed in the next section.

Applied criticisms of response to intervention

A wide range of questions about the applicability of a multi-tiered RTI program have been asked, including who would assume ultimate responsibility for the success and considerable time outlay involved in a fully implemented program (Mastropieri & Scruggs, 2005). Prominent proponents of RTI have asked for caution in its implementation stating that “the importance of asking questions . . . about RTI implementation is underscored by recently released federal regulations that are bereft of procedural detail” (D. Fuchs & Deshler, 2007, p. 130).

There has been much confusion as to how to implement RTI; there is little consensus amongst special education directors as to what the procedure for LD identification is within an RTI paradigm, let alone who is responsible for the said intervention (Gessler-Werts, Lambert & Carpenter, 2009). Six decisions that a school needed to make when implementing RTI were listed by L. Fuchs and Fuchs (2007), in an article touted as providing a framework for RTI. These were: how many tiers of intervention would be required, how to target students for preventative intervention,
the nature of that intervention, how to classify the response to that intervention, the nature of evaluation before referral to special education and finally the function and design of said special education. This could reasonably lead one to question what aspects of LD identification and remediation that this RTI framework provided.

The vagueness within the RTI paradigm about what constitutes comprehensive assessment is a point of contention (C. Reynolds & Shaywitz, 2009a). With only one tier separating general education and special education in a recommended model of RTI (L. Fuchs & Fuchs, 2007), the consequences of not responding to the intervention supplied at Tier 2 are severe. One suggestion, made by L. Fuchs and Fuchs, was that criteria could include meeting a cut point at the 25th percentile for overall achievement, as well as not being in the bottom 50% in the magnitude of response to the intervention. The arbitrary nature of such categorisations, especially as a tool for deciding who is entered into special education, was a major criticism of the discrepancy method of identification. Furthermore these cut-off scores do not take into account the varying standards of both teachers and students in different schools.

Variation in standards of teaching that will always exist between classes, schools, districts and countries, has been brought up as concern in implementing RTI (Gerber, 2005), as well as the fact that students will react to increased motivation or teacher intervention in different ways. This has led to the question of whom or what should be considered responsible for not responding to intervention: teachers, instruction, students or the measures implemented (Hale et al., 2006)? The role of general teachers, special education teachers and diagnosticians, would presumably change within an RTI paradigm, but it is not clear how (Mastropieri & Scruggs, 2005).

The varying standards of teaching, combined with different possible methods of applying RTI could lead to the identification of RTI students becoming more dependent on the measures put in place than the actual presence of LD. One review of the implementation of RTI in the US found that most states have implemented a three tier model, but in vastly different ways. Furthermore each state has not necessarily enforced the research based practices aspect of RTI (Berkeley, Bender, Peaster & Saunders, 2009), an assumption inherent to the model.
One of the major reasons for differences in application of RTI from school to school is financial cost. The money required to pay for the teachers for the extra remediation alone is prohibitively expensive (Gessler-Werts et al., 2009) and many principals believe that RTI would be too expensive to implement (Fletcher & Vaughn, 2009a). Furthermore, as noted by Swanson “a mere call for evidence based studies and/or progress monitoring does not yield uniformity in practice” (Swanson, 2008a, p.30). Basically, the assumption that all, or even the majority, of environmental issues can be overcome by RTI has not been justified.

**Shared problems with response to intervention and discrepancy**

One of the most striking aspects within the literature surrounding the RTI-discrepancy debate is that many of the criticisms levelled at either method of identification are actually relevant to both. For instance, one of the primary empirical weaknesses of RTI is that there have been a range of methods of implementation of the RTI program that vary between schools, states and countries, which affects generalisation (Swanson, 2008a). This was of course also a primary issue with the discrepancy method, the variability of how a significant discrepancy was defined and tested between schools, districts and countries (McKenzie, 2009).

One review of the implementation of RTI in the US found that most states have implemented the three tier model in vastly different ways and that the states have not necessarily enforced the research based practices that are a pivotal assumption of RTI (Berkeley et al., 2009; Mastropieri & Scruggs, 2005). This is similar to the criticism levelled at the discrepancy method of identification in that the varying methods of determining the significance of an aptitude achievement discrepancy reduced the reliability of that method (Sternberg & Grigorenko, 2002).

Although, early screening has been described as a cornerstone to LD prevention (Barnes & Wade-Woolley, 2007) this is an area that has been considered lacking in both methods of identification. Proponents of the discrepancy method are accused of suffering from a wait to fail mentality (Bradley et al., 2007; D. Fuchs & Fuchs, 2006), something that has been described as morally unacceptable (Nicolson, 1996). Meanwhile, RTI proponents have been accused of a ‘watch them fail’ mentality (C. Reynolds, 2008). In both cases these accusations occurred because reading has to get to a certain low standard before it can be identified as a problem: in the case of
RTI as compared to classmates, in the case of discrepancy as compared to IQ. This is important when the Matthew effect is taken into account, as once students get behind, it becomes increasingly difficult to catch up.

Many of the psychometric issues inherent to the arbitrary nature of the discrepancy method must surely apply to RTI (Scruggs & Mastropieri, 2006), primarily in the operationalisation the concept of response to intervention. However, such criticisms of RTI guidelines have not been made explicit as of yet, as no clear guidelines or standards have been put forward by proponents of RTI (C. Reynolds & Shaywitz, 2009b).

Both methods also have issues with timing for effective use, it would be ideal for a method to be appropriate at all ages, however this is not the case with discrepancy or RTI. Proponents of RTI have acknowledged that it may not be appropriate for those in Grade 2 or older (Semrud-Clikeman, 2005). Meanwhile it has been found that the use of the discrepancy model is more appropriate in Grade Three students than in Grade One students (Shaywitz, Escobar, Shaywitz, Fletcher & Makuch, 1991).

Another example of shared criticisms is how little each method informs remediation. It has been argued that aptitude and achievement measures alone cannot provide enough information to facilitate remediation (Silver et al., 2008; Willis & Dumont, 2006). However, the same has been said of RTI as not responding to intervention gives no clue as to why the intervention was not successful (C. Reynolds & Shaywitz, 2009a).

It has also been argued that a flaw in the discrepancy method is that it overlooks students who have compensated for their shortcomings in say, phonological processing to date and thus do not show a sizeable discrepancy (Gustafson & Samuelsson, 1999). However, the same could be said for those who are compensating well enough in an RTI environment to avoid original selection for remediation. In fact, a common criticism of RTI is that it fails to identify gifted LD students (Volker et al., 2006).

Ultimately when looking at these shared criticisms it becomes clear that both sides of the argument have merit. With a discrepancy method of identification, students have to wait for their difficulties with reading to become severe enough to result in a
significant discrepancy before identification. Meanwhile, a student in an RTI paradigm won’t even qualify for the first tier of intervention, if they do not fall in to the low achievement range. In both cases this is a real concern. Some LD students have an ability to memorise the small number of words that they are expected to read in the early years of education, and don’t have their problems recognised until three or four years into their schooling when the number of words they are expected to read rises exponentially (Woolley, 2007).

These shared criticisms directed at the two methods for identification of LD can both be summarised with one issue. Both methods test for the existence of LD by testing for a likely, but not necessary, operational definition of LD, rather than testing for the deficits that are causing the reading difficulty in the first place. As such both are inaccurate to a certain extent, and provide little information for remediation. However, there are methods of identification that are identifying LD in terms of directly testing the deficient cognitive processes underlying the disorder, namely the phonological deficit and the double deficit models.

**Phonological deficit model**

Given the consistent finding that phonological deficits are common in students with LD, it has been put forward that LD can be best represented by a phonological deficit (Shaywitz, 1996; Snowling, 1981). In 1998, a phonological core deficit model was put forward in which phonological deficits were the primary marker variable in LD operationalisation (Coltheart & Jackson, 1998) and in Stanovich’s phonological-core variable difference model (Stanovich, 1988). More recently this theory has been put forward again as the phonological-core variable orthographic deficits model (van der Leij & Morfidi, 2006), that co-opts research in English and orthographically shallow Dutch, stating that phonological deficits lie at the core of LD in both languages.

This model has been supported by multiple studies that have found phonological deficits at a younger age were a good predictor of later reading problems (Jorm, Share, Maclean & Matthews, 1986; Shaywitz et al., 1999). These studies are an example of how the discrepancy method of identification can be used as a starting point for assessing the efficacy of other methods of identification. Phonological processing deficits were thought to provide an explanation for the discrepancy that is found in LD students, as taking the ‘unexpected’ out of underachievement. This is
because “the reading failure of a high IQ individual is expected if the person is low in phonological awareness” (Stanovich, 1998, p.18).

In 1980, Snowling wrote an article on the difference between LD and LA students that is still relevant today. One of the methods used in this article was a reading matched design in which LD students were compared to a reading matched sample, rather than a chronologically matched one. The type of match in the design is important because, up until this point, a persuasive argument in the field was that LD students did not have phonological deficits. Instead, it was thought that they were merely lagging behind their classmates in reading and all the processes that contribute to reading. As such, the argument went, when a student reached the same reading level as their classmate they would also gain a commensurate level of phonological processing ability. By matching LD students to reading matched controls, for instance matching an eight year old who read at a six year old standard with a six year old who read at an expected level for his or her age, it should be possible to see if the eight year old had reading age-appropriate phonological processing levels. In this paper, Snowling found that there was indeed a deficit and that LD students did have significantly lower phonological processing skills than their reading age matched counterparts. Successful replications of this finding (Stanovich, 1988; Stanovich et al., 1997) led to the theory that LD could be best represented by a phonological deficit.

Snowling, Goulandris and Defty (1996) tested 20 discrepancy-defined LD students with chronological age and reading age controls and administered a variety of measures over two years. All LD students had an IQ of 85 or more and had a reading age of at least eighteen months below their chronological age. At the initial time of testing, there was no significant difference between LD students and their reading age counterparts, on reading proficiency measures. Two years later however, the reading age controls were significantly better at reading nonwords than LD students.

Interestingly, phonological processing gaps do seem to close over time, but without a commensurate closing of the gap in reading ability. Foster and Miller (2007) found that the gap in decoding between poor and good readers in Grade One took two years to close. However, most importantly, reading comprehension deteriorated in the
poor reading group while the decoding gap was narrowing. This indicates that although poor readers do eventually overcome their decoding issues, this is done at the expense of comprehension. This theory could also provide an explanation for the Matthew effect: if a student has to compensate for a deficit in a skill that the teacher is now assuming they have, they will not be able to make the most of the training that their classmates are receiving in a new reading-related skill.

Although the phonological deficit theory of LD has received empirical support (Fletcher et al., 1998; Shaywitz & Shaywitz, 2005), it has been found in other studies that the neurological deficits in phonological processing that could be found in LD students could also be found in LA students that would not have fit the discrepancy based requirements for an LD diagnosis (Stanovich & Siegel, 1994). Whether this is a failure on the part of the discrepancy or phonological deficit model of LD identification is unclear.

However the major obstacle for the phonological deficit theory is that phonological processing is not low in all readers that could be considered LD (Pennington, Cardoso-Martins, Green & Lefly, 2001). Although reading comprehension problems are often thought to stem from problems with phonological processing, this has been shown to not always be the case. Some students with poor reading comprehension are good phonological decoders and seem to have problems stemming from processing semantic information (Nation, Snowling & Clarke, 2007). Once again, one of the issues that has to be constantly grappled with, is whether or not these other poor readers also have LD, or whether the absence of phonological processing deficits indicates that they should receive a different diagnosis.

In a response to Siegel’s (2006) suggestion that LD can be diagnosed on the basis of poor phonological processing as measured by nonword reading, it was argued that any unidimensional model of reading is not going to be complex enough to explain the heterogeneous nature of LD (Katzir, Kim, Wolf, Morris & Lovett, 2008). The importance of remaining open to the idea that there are cognitive deficits playing a role in LD aside from phonological processing has been stressed (Rack et al., 1992).

From this point, the obvious avenue of research is in investigating what other cognitive processing deficits could be responsible for LD. One theory put forward to answer this question was the double deficit theory of LD in which identification
could be based on not only phonological processing deficits but also deficits in naming speed. The next section will cover research into this theory of LD identification.

**Double deficit model**

The double deficit hypothesis, put forward by Wolf and Bowers (1999) states that LD can be explained by deficits in not just phonological processing, but also in naming speed. This results in three potential groups of LD students; those with naming speed deficits, those with phonological processing deficits and those with both, the double deficit group. The double deficit group would have more severe reading problems than the two groups with only one deficit, which in turn would have more difficulty than those without LD at all.

In two exploratory studies utilising cluster analysis, a four factor structure was considered appropriate for the analysis and these clusters corresponded with those that would be predicted by the double deficit hypothesis, namely the double deficit, naming speed deficit, phonological processing deficit and no deficit groups (King, Giess & Lombardino, 2007; Pierce, Katzir, Wolf & Noam, 2007). In both studies, the sample was made up of poor readers and the no-deficit groups consisted of those who were the better readers in the sample.

Even after controlling for phonological processing and nonword spelling ability, naming speed is still a unique predictor of spelling, leading to the conclusion that the double deficit theory was possibly applicable to spelling as well as reading (Savage, Pillay & Melidona, 2007). Naming speed and phonological processing skills were thought to be a predictor of numeracy as well as literacy, possibly indicating global, rather than reading specific deficits of LD should be examined (Wise, Pae, Wolfe, Sevcik & Morris, 2008).

A longitudinal study based on intervention success found the double discrepancy method of identification preferable to the simple low achievement or IQ discrepancy methods (Speece, Case & Molloy, 2003). Results from a Greek study suggested that naming speed deficits in kindergarten and Grade 1 did not result in marked deficits in reading until Grade 3, when fluency was more of an issue (Papadopoulos, Georgiou & Kendeou, 2009). It should be noted that the same study found that phonological processing was a less robust predictor of future reading ability than
naming speed. This was possibly attributable to the orthographic consistency of the Greek language. The double deficit theory is also applicable in non-alphabetic languages such as Chinese, even though the non-alphabetic nature of the language dictates a reduced focus on phonological processing. Chinese students with LD have phonological and naming speed deficits when reading in English (Chung & Ho, 2010).

The double deficit hypothesis was utilised successfully in a longitudinal study in which both processes provided separate variance in predicting future reading ability (Torppa, Lyytinen, Erskine, Eklund & Lyytinen, 2010). Deficits in phonological processing and naming speed are also supported by neurological theories of reading impairment (McCandliss & Noble, 2003). Neurological studies have shown that phonological deficits and naming speed deficits may both stem from reduced activation in the left occipito-temporal, region indicating that they might have a common biological basis (McCrorry, Mechelli, Frith & Price, 2005).

It has been argued by proponents of the double deficit theory that phonological processing and naming speed are both unique contributors to reading ability despite the significant correlation between the two (Frith, Wimmer & Landerl, 1998; Yap & van der Leij, 1993). Others (e.g., Schatschneider, Carlson, Francis, Foorman & Fletcher, 2002) consider naming speed as simply a part of phonological processing and not a separate construct. In order to test these two theories, Powell, Stainthorp, Stuart, Garwood and Quinlan (2007) used structural equation modelling to assess the independence of the two constructs. They found that naming speed is at least partially independent of phonological processing. This was because models depicting the two as separate deficits accounted for the data collected from a large sample of British Grade 3 students, whereas a model utilising one factor did not.

In order to ascertain whether or not the moderate correlation between these two variables had had an effect on studies to date Schatschneider and colleagues (2002) examined the shared and unique variance in reading ability that is attributable to phonological processing and naming speed. One of their findings was that the shared variance of these two variables was at least as effective in predicting reading as the unique variance of both variables combined. This finding could be interpreted in a number of ways, but the shared variance between the two was so important for
predicting reading variance, this could indicate that these deficits have a cumulative affect, rather than that the two tests are representing the same construct.

Furthermore Schatschneider and colleagues (2002) also pointed out that one of the difficulties with the double deficit hypothesis, at least in terms of interpretation of results, was the correlation between the two variables. If two variables are correlated positively, then the mean on variable A will decrease as the mean of variable B decreases. Therefore, it is possible that results that have been attributed to naming speed could have actually been a reflection of the importance of phonological processing, or vice versa. As such, it could be that it is not the double deficit that is causing so much difficulty for students, but rather the decreased phonological processing abilities of students afflicted with both deficits. However they went on to stress that the role of naming speed in reading was something that warranted further study.

Critics of the double deficit hypothesis often claim that the correlation between phonological processing and naming speed could be making results seem more significant than they are (Vellutino et al., 2004). However, it has been found that naming speed does not even correlate significantly with most phonological processing tasks (Krasowicz-Kupis, Borkowska & Pietras, 2009). Pennington and colleagues (2001) suggested that one of the reasons that many LD students struggle in naming speed tests is because their letter knowledge is poor, not their naming speed. However their results were mixed, regardless of how they were interpreted. Although phonological processing was easily the strongest predictor of measures of reading and spelling, tests of naming speed were also significant predictors to all measures.

Further to this debate is the evidence from a genetic study that suggests there are unique and shared genes responsible for variation in naming speed and phonological processing (Naples, Chang, Katz & Grigorenko, 2009). This could explain why the two processes can be found to be different in some students while sharing so much variance. If indeed this is the case, then it is important that both are tested for, as the joint causes of the deficits can be assessed in both tests but the individual causes can only be tested through each individual assessment.
The temporal instability of the double deficit theory has been put forward as a possible issue. In one longitudinal study (Spector, 2005) only 50% of students remained in their original grouping over a six month period. However, it should be noted that 80% of students in one of the deficit groups were still in one of these three groups at the six month follow-up. Furthermore, individual deficits were more stable over the time period with 74% and 70% of those students with naming speed deficits and phonological processing deficits at Time 1, still having them at Time 2 respectively.

One of the primary advantages with the double deficit theory is the information it provides for remediation. In a longitudinal study students were designated into deficit groups before receiving remediation. The phonological deficits group benefited the most from the phonologically-based remediation (Lovett et al., 2000). This is an indicator that remediation programs directed at the cognitive processes behind LD, can be of use.

Further support for the double deficit theory can be found in a study of 158 students that were either assessed as LD or LA, via the discrepancy method, before being allocated into double deficit prescribed groups (Katzir et al., 2008). Within the LD group, 63% had a double deficit, 16% had a phonological deficit, 11% had a naming speed deficit and 10% had no deficit. In contrast, within the LA group, 41% had a double deficit, 36% had a phonological deficit, 18% had a naming speed deficit and 5% had no deficit. What is particularly fascinating about this study is that phonological deficits, thought to be the deficit that separates LD from LA students, were present in 77% of LA students. Furthermore there were twice as many students in the LD group without any deficits than there were in the LA group. Once again, this was a strong indication that students designated as LA may actually be facing the same issues as their LD counterparts.

In a meta-analysis of 19 studies comparing LD and LA students, LD students outperformed LA students on measures of nonword reading and naming speed (Hoskyn & Swanson, 2000). This was thought to be due to LA students scoring low in these measures, not LD students scoring high. As such, the double deficit hypothesis does seem to have the same difficulty that both the discrepancy and RTI methods of identification have, in that it cannot differentiate between LD and LA.
students. However, what needs to be stressed here is that the theory does not
differentiate between LD and LA students as prescribed by the discrepancy method.
As with the vast majority of research in the field, there is no guarantee that the LD
cohort identified via the discrepancy method is actually the genuine LD cohort.

Some of the more interesting studies in this field are those that find that although
there is evidence for the three groups prescribed by the double deficit group, there
are often LD students that do not fit into any of these categories (e.g., Lovett et al.,
2000; Pennington et al., 2001). This could also be interpreted as the double deficit
not adding enough cognitive processing deficits to their model. It could be that there
are more than two cognitive processes that could cause LD.

Further support for the theory was provided by a large scale study of readers aged 8
to 18 years (Compton, DeFries & Olson, 2001). In this study, phonological and
naming speed deficits were found to be additive in students with a double deficit.
Furthermore, as in previous studies (Lovett et al., 2000; Pennington et al., 2001), not
all of the LD cohort in the study could be classified. By reducing the deficit criterion
from one and a half to one standard deviation below an age-based mean, the study
increased the number of readers classified from 61% to 78%. However this does not
necessarily diminish the double deficit theory. Many researchers agree that there are
many different types of LD and that it is unlikely that all cases have a single
underlying cause. The remaining unclassified students with LD could be those with
deficits in other cognitive processes.

From this discussion of competing theories of LD, two common themes emerge. In
the case of RTI and discrepancy, there are common issues inherent to testing for LD
on the basis of secondary symptoms as compared to testing the direct causes. In the
case of the phonological deficit hypothesis and double deficit hypothesis, despite
addressing this issue there are groups of students not accounted for. Given the
generally accepted heterogeneous nature of LD, this leads to the suggestion that,
although the double deficit hypothesis provides more insight, there are other deficits
involved that need to be taken into account in the identification of LD. Primarily
there is evidence that verbal memory deficits may play a role in LD. A multiple
deficits model, incorporating the verbal memory deficits with a double deficit model,
along with the theoretical rationale for the current study, is put forward in the next section.

**Convergent Validity and the Multiple Deficit Model**

In a comparison of students with and without LD, Ackerman and colleagues came to the conclusion that although phonological processing and naming speed did separate the two groups, there were other cognitive processing deficits involved that needed to be taken into account (Ackerman, Holloway, Youngdahl & Dykman, 2001). For instance, Vukovic and Siegel (2006) divided an LD cohort into double deficit groups, and found that four out of five students with naming speed only deficits had digit span deficits as well. A review article of neurological and genetic studies on LD found that alongside phonological processing, short term and working memory should be considered as deficits inherent to LD (Shapiro, 2001). It is possible that some of the unexplained variance in LD research could be explained by verbal memory deficits. The current study aims to examine if this is the case by comparing this multiple deficit model to the discrepancy theory.

Support for a multiple deficit model has also come from studies conducted in orthographically consistent Spanish (Jimenez, Siegel, O'Shanahan & Ford, 2009). In this study naming speed was significantly faster in RA readers than LD readers. Furthermore naming speed did not improve with IQ. Phonological processing was also significantly lower in LD than RA readers, however this did decrease with IQ. Finally, working memory differences were also found between the LD and RA reading groups.

Support for multiple cognitive processing deficits being utilised as a criterion in LD identification has been sought and gained (Johnson, Humphrey, Mellard, Woods & Swanson, 2010). In this study the discrepancy method of identification was used as a baseline for measuring effectiveness and found that cognitive processes could be used to identify discrepancy-defined LD students. Large effect sizes in verbal working memory, phonological processing and processing speed amongst others were found between LD and non-LD students.

Hatcher, Snowling and Griffiths (2002) utilised a discriminant function analysis using nonword reading, digit span and writing speed and were able to correctly identify 96% of LD students in a tertiary education sample. They went on to use
these processes to develop a model of identifying LD students. It should be noted however that, as already discussed, these processing deficits often lead to reading difficulties that may not be evident in primary school, but would presumably be clear by the time a student reached tertiary education. Consequently, such a high success rate may not be attainable in a younger sample.

It is appropriate to make one caveat at this point. The theory that verbal memory and naming speed deficits exist in LD students is accepted by some proponents of the phonological deficit theory of LD: the difference is that they consider both of these processes to be part of phonological processing (Carroll & Snowling, 2004; Ramus, 2003). Ultimately whether or not phonological processing, naming speed and verbal memory are considered separate processes or sub-groups of one process has little bearing on the current study. What is being argued is that all three processes, be they separate or part of a wider process, provide some unique variance that is relevant to the identification of LD.

There has however been much support for the claim that LD is a heterogeneous disorder (Cain & Oakhill, 2006). Some studies have found a variety of cognitive profiles within a cohort of students diagnosed with LD (Fiorello et al., 2006; Pierce et al., 2007). The heterogeneity of LD has resulted in difficulty generating a neuropsychological profile of the cognitive processes affected in the disorder (Kohli, Malhotra, Mohanty, Khehra & Kaur, 2005) and could be confounding many of the results found in studies focusing on an LD cohort (Rack et al., 1992). It has been also been demonstrated that this wide range of intra-group differences remain through adulthood (Rack, 1997). Furthermore, Pennington (2006) has put forward the theory that multiple deficit models of LD could help explain the complex comorbidity between LD, ADHD and speech sound disorder.

There are a number of multiple deficit models that are currently attempting to account for the heterogeneous nature of LD. For instance Pennington (2009) introduced a multiple deficits model incorporating phonological processing, naming speed and auditory deficits. Meanwhile Ho and colleagues have developed a multiple deficit model based on phonological processing, naming speed, visual and orthographic processing deficits (Ho, Chan, Tsang & Lee, 2002). Whilst models such as these, and the one that will be put forward in the current study, could be seen
as conflicting, they could also be seen as complementary. With multiple, overlapping causes, the testing of varying multiple deficit models will assist in ascertaining which deficits are pertinent and which ones are not.

Group results summarising LD profiles are often incongruent with individual profiles of LD students, even those forming a part of the group used to develop the profile. Consequently, the LD cohort is often too heterogeneous to have one, two or even three patterns of scores to utilise as a diagnostic tool (Spagna, 1998). It has been recommended that individual tests of aptitude be utilised so that individual deficiencies can be examined (Berk, 1983). Of course, if there were three deficits, such as phonological processing, naming speed and verbal memory, then there would be seven subtypes and more if additional processing deficits were included in the model. This is further confused by evidence for overlap between these cognitive processes, with each processing deficit providing both unique and shared variance (Pennington et al., 2001).

The disconnect between definition and operationalisation has been noted as a real issue in this area of research. Cognitive processes are mentioned in the definition of LD but not in the methods with which it is identified (Kavale, 2005). Using multiple combinations of cognitive deficits as marker variables in LD identification is not new, it has however been ignored in the wake of the RTI-discrepancy debate. In 1994 Stanovich and Siegel stated that

If there is a special group of children with reading disabilities who are behaviourally, cognitively, genetically, or neurologically different, it is becoming increasingly unlikely that they can be easily identified by using IQ discrepancy as a proxy for the genetic and neurological differences themselves (p. 48)

One of the most important assumptions about LD students is that they have a range of cognitive processing deficits and in the current study, decisions will be made based on this very assertion. Current procedures misclassify LD because of an inability to accurately assess aptitude and differentiate between LD and LA students (Swanson, 1996). In a study of the opinions of LD academics, 96 and 88 percent of respondents did not believe that RTI and discrepancy, respectively, were sufficient for LD identification (Hale et al., 2010). Others are concerned that the rise of RTI
and concurrent decrease in the use of cognitive processing measures could result in
the erosion of the concept of LD altogether (Mastropieri & Scruggs, 2005).

One of the reasons for this rise was the dismissal of the discrepancy method due to
identification in LD research was this method’s failure to differentiate between LD
and LA students. RTI was touted as an improvement on the discrepancy method as
there was little need in an RTI model to do so. However, from this point it is logical
to assume that either the discrepancy method can not differentiate between the two
and is thus not an effective identification tool, or alternatively, that there is no point
to differentiating between the two and that as such this failure has no bearing on the
efficacy of the discrepancy method as a tool of identification. What has instead been
taken away from this is that there is no point to differentiating between the two, and
that this failure is indicative of the discrepancy methods inability to do just that. In
the current study it is argued that there is a point to separating LD and LA students
and that while the discrepancy method cannot accurately do this, an improved model
is desirable.

One of the major advantages of using cognitive processing deficits to identify LD is
that it will be possible to identify students early, before their reading is affected.
This is instead of educators waiting for reading difficulties becoming enough of a
problem to be picked up by RTI or discrepancy. This is important, as some of these
students are able to compensate for their difficulties early on and do not have their
problems recognised until the expectations in reading ability change around Grade 3
or 4 (Woolley, 2007). The second major advantage would be that such testing could
be implemented without exclusionary criteria such as cultural disadvantage or low
intelligence. There has never been any suggestion that the existence of phonological
decoding difficulties and low intelligence are mutually exclusive, nor has it been
demonstrated that those with low intelligence would not benefit from intervention. It
could be easily hypothesised that LD students with low intelligence would benefit
more from intervention as they presumably do not have as many ‘back up’ processes
to draw on when developing compensatory skills. Yet these students are not
traditionally thought to have LD as their underachievement is expected, due to their
aptitude.
The same argument could easily be applied to SES. The negative effect of low SES on IQ aside, the exclusionary criterion of low SES is certainly not based on any kind of link between cognitive processing and SES. There is no reduced likelihood of a student from a low SES background having a neurological deficit in phonological processing. Thus using SES as an exclusionary criterion, for what is becoming widely accepted as a neurological disorder, is flawed. What is required instead is the ability to differentiate between low SES poor readers with LD, and those whose poor reading is due to their environment. By assessing the cognitive processing abilities of such a student it should be possible to ascertain the cause of any reading problems.

The development of a new model based on multiple cognitive processing deficits needs a benchmark against which its efficacy in accurately identifying LD students can be measured. One of the primary issues in this area is that true validation is difficult if not impossible without a well-defined and tested standard for comparison can be made. For example, if a researcher wished to put together a predictive screening tool for future obesity they could utilise cross sectional studies in order to develop a set of predictor variables and then validate the developed framework with a longitudinal study. In the study, people would be considered accurately identified if they were predicted to be obese and then went on to have a Body Mass Index (BMI) over the cut level for obesity. However in the case of LD, we are not just looking for those who are poor readers, we are trying to identify those who struggle with literacy for a given set of reasons that as yet have not been well defined, or at least agreed upon. Therefore the biggest difficulty lies in establishing convergent validity of the variables being tested, as there is a need for a recognised standard to which results can be compared.

When given the choice of the two most popular forms of identification, RTI and discrepancy, the choice for discrepancy as a comparison for convergent validity is clear. By contrast, RTI does not differentiate between poor readers and LD students, except in as much as they respond to remediation. Without evidence that the response to intervention is dictated by the cause of the reading problem, there is little to be gained from using RTI as a base. This is not to say that there are no benefits to the use of RTI, this is rarely disputed, just that it’s accuracy in identifying, rather than remediating, LD has not been ascertained. The discrepancy method of
identification does represent LD quite well: the student who cannot read despite not having any obvious cause for this. In a review of the identification and definition problems that plague LD research, Scruggs and Mastropieri concluded that “discrepancy is the most objective indicator” (2002, p.164) of LD and that the elimination of this construct would result in an increase in many of the problems plaguing LD identification.

Although the discrepancy method has been widely dismissed in the academic literature as an inaccurate way of identifying students with LD, there are still compelling reasons for selecting it as the basis for convergent validity for a multiple deficits hypothesis. Although the logic of the phonological and double deficits model is sounder, they can not be used as the basis for comparison as convergent validity will be automatically reduced with the addition of extra variables.

It is therefore important to look at why the discrepancy method of identification was dismissed. The primary reason for moving on from the discrepancy based method of identification was the psychometric instability of a significant discrepancy. When two factors have a moderate correlation the likelihood of a discrepancy over one standard deviation occurring changes as the level of aptitude increases (Dykman & Ackerman, 1992). Secondly, the logic behind predicting a neurological disorder on the basis of a discrepancy between two constructs that only correlate moderately is dubious at best (Fletcher et al., 1994). However, the reason the discrepancy method was so popular in the first place was that it represented that often quoted idea that there are students out there who are bright but just can not read. The discrepancy method was quite good in this respect: critics do not claim that the discrepancy method was totally inaccurate in identifying LD students, just that it was not accurate enough.

To come back to the obesity analogy, if it were not possible to measure muscle mass or body fat percentages, researchers may resort to other correlates of obesity in order to establish that whatever measure they have put in place is the most appropriate measure of, for instance, heart problems. However, obesity is not the only risk factor involved in heart health; genetics, lifestyle, smoking and a range of other factors play a part and as such you cannot measure the predictive validity of an obesity measure by how well it predicts heart health.
The same goes for LD and reading. LD is one reason why a student may not be able to read, however there are a wide range of other reasons for this including socio-economic factors, scholastic aptitude, parent’s literacy, quality of schooling, motivation and emotional behavioural issues (Coutinho, Oswald & Best, 2002; Davis-Kean, 2005; Metsala, Sweet & Guthrie, 1996). As such, it is not enough for a LD screening tool to be able to predict those that will not be able to read in the future. One needs to identify those who will struggle with literacy due to a neuropsychological disorder. This is why the discrepancy method as baseline for comparison is important.

Neurological qualitative differences between LD and LA students (Pugh et al., 2000), is evidence that there is a point to further examining differentiating the two groups. Although the discrepancy method of identification has failed to accurately do so, it is an operational definition that can provide a benchmark. In this case a difference between the two groups, given the research reviewed in this chapter, could be in cognitive processing deficits. As such one of the explanations for the failure to find differences in the LD and LA discrepancy-defined groups is that both of these groups, as they are currently defined, contain LD students. Therefore what is being proposed in the current study is that instead of ascertaining a methods efficacy in differentiating LD and LA students by testing for cognitive differences between the two, that the cognitive differences be used as the test that separates the two groups in the first place. In this scenario, the concept of LA would still be applicable, but refined. It would represent those students whose reading difficulties cannot be explained by cognitive processing deficits, rather than those students who are reading at a level that is similar to that of their aptitude.

In order to address these issues, to ‘untangle’ ourselves so to speak, the same methods that these assumptions have been based on need to be examined through a new paradigm so that previous results can be appropriately interpreted. For instance, within the discrepancy method of identification studies have been conducted utilising traditional discrepancy methods and regression-adjusted discrepancy methods. The results from these studies have been utilised as a basis for further work, most disturbingly in areas where low achievement is considered a sufficient criterion for identification. How these results could be interpreted through a cognitive processing framework, as compared to a discrepancy or RTI based one, is important and needs
to be investigated. With these issues fully investigated, it may then be possible to even integrate the discrepancy and cognitive deficit models, such as that found in the concordance-discordance model of LD identification (Hale & Fiorello, 2004).

This brings us back to the assumptions that a study requires in order to be able to proceed. The ability to assess concurrent validity is dependent on having a yardstick with which to measure success by and in the case of the current study, the aptitude achievement discrepancy will be used. However, it will be used cautiously, intelligence measures will be utilised so that the discrepancy method of diagnosis can be examined. However, any final decisions will not be made on the basis of intelligence testing, the area of the discrepancy method that has been shown to be the most psychometrically and theoretically flawed. Although the discrepancy method should not be used as a final word in identification, as a slew of research papers would testify, it does represent the LD prototype well. Consequently, it can be used as an effective baseline measure to which improved measures of LD can be compared.

**Summary**

The importance of ensuring that each student is able to develop adequate literacy skills before leaving school is widely accepted. However there are some students who, regardless of quality of education, can not learn to read. In this chapter a brief summary of the definition, nomenclature, prevalence and aetiology of LD has been put forward. A more detailed exploration into the debate into the operationalisation of LD and the implications has also been put forward.

Discrepancy based identification methods were discussed, along with low achievement, RTI, phonological deficits and the double deficit models. Although the double deficit model seems to be gaining moderately successful results in the literature, the numbers of LD students that cannot be classified under this model suggest that a more heterogenous array of deficits may be involved.

A multiple deficits model of LD operationalisation was discussed, with the possibility of putting forward verbal memory as a third deficit suggested. However, in order to assess the success or failure of this method to identify students with LD, an already established method of identification is required in order to evaluate convergent validity. An argument for using the discrepancy method of identification
as this basis for comparison was put forward. This was because although there was
debate over the accuracy of the discrepancy method, the claim that it represented the
essence of LD well was rarely debated.

With a multiple deficit hypothesis proposed, the ability of these specific cognitive
deficits to differentiate LD students from their non-LD counterparts is discussed in
Chapter 3. Furthermore the importance of meeting the applied ethos when
researching within an educational psychology field is discussed, along with methods
of ensuring that any screening battery to come out of this study can be easily applied.
Chapter 3

Overview

In Chapter 2 it was argued that one of the possible ways forward in the identification or prediction of Learning Disabilities (LD) is to directly test for the cognitive processing deficits that are thought to cause LD. In a review of current practices for diagnosing LD in adults, it was concluded that cognitive processing measures were more accurate because models such as discrepancy or RTI were fraught with problems. This chapter will review recent studies on the cognitive processes thought to cause LD, as well as relevant aspects of applied research suitable for the classroom.

Phonological processing has been long considered the primary cognitive deficit that causes reading-based LD. Previous research in regards to phonological processing and LD will be discussed with a focus on studies utilising nonwords and reading age-matched controls. Naming speed is another deficit that is thought to be partly responsible for LD, often discussed within a double deficit paradigm. Relevant literature on naming speed and LD will also be reviewed. Verbal memory is less commonly touted as a cause of LD, but a review of the relevant literature suggests that it may indeed be another deficit that could contribute to LD.

The theory that multiple deficits in these processes has a cumulative effect on reading originally stemmed from the double deficit theory, where it was suggested that students with both phonological and naming speed deficits have more trouble with reading than those with a deficit in only one of these processes. Recent research on multiple cognitive deficits within LD suggests that this is indeed the case, even in a model that incorporates other processing deficits. The conflicting literature around another cognitive process that is implicated as a cause of LD, orthographic processing, is also reviewed.

After the discussion of the cognitive processing deficits used in this study, previous issues with conducting applied LD research are reviewed. In this section, research conducted involving teachers in LD identification and how cognitive deficits have been the successful focus of LD remediation in the past will be reviewed. Finally, a review of the challenges inherent to the development of screening tools, similar to
the ones developed in the current study, along with a review of why it is important to target the screening battery to a narrow age range, is included.

**Identifying Learning Disabilities**

The heterogeneous nature of LD is one of the primary causes of confusion in the definition and operationalisation of the disorder (Stanovich, 1988). There has been limited neurological support for different groups of students with different neurological deficits existing within one LD cohort (Pernet, Poline, Demonet & Rousselet, 2009). A large twin study with 8000 participants found support for the theory of genetic overlap between the different types of learning and cognitive abilities that are required for reading (Haworth et al., 2009). This finding is consistent with a multiple deficit model of LD, as there are a number of cognitive deficits that seem to have some overlap, as well as contributing unique variance to reading and LD.

The argument that there are a range of cognitive processes that can contribute to LD has been supported by a study that utilised both Chinese and English languages (Ho & Fong, 2005). On average, Chinese students with LD have more trouble learning English than their non-LD counterparts. As LD was diagnosed in a non-syllabic language, phonological processing deficits were not as important to the original diagnosis. However, the concomitance of LD in both languages was high, possibly due to a set of core of processing problems such as verbal memory or naming speed. Furthermore, the overlap of causes for these deficits could result in an increased likelihood of students with verbal memory or naming speed deficits also having phonological processing deficits.

It has also been consistently found that there are different levels of importance in cognitive predictors of reading proficiency in LD and Regular Achieving (RA) readers (Chan et al., 2007). It is important that this finding is taken into account in any study hoping to identify LD in the general population, especially when considered in conjunction with the well documented difficulty in finding a consistent operationalisation for LD. If, for instance, verbal memory deficits are an underlying cause of LD and, as such, are a predictor of reading ability in an LD cohort, this does not necessarily mean that higher levels of verbal memory proficiency in non-LD students will predict higher reading proficiency. Thus, it is plausible that verbal
memory could be confounding results. However in a method of identification of LD that was based on cognitive deficits alone, this would be less of an issue.

According to the double deficit hypothesis, multiple cognitive processing deficits will cause more reading difficulty than would be caused by only one deficit (Graham & Bellert, 2005). That there might be various possible combinations of cognitive deficits within LD is supported by a longitudinal growth model study in which ten different types of reading development were identified (Boscardin, Muthen, Francis & Baker, 2008). These groups were based on phonological processing, naming speed and orthographic processing, and the authors suggested that, in future, other cognitive variables could be utilised to increase predictive validity.

Cognitive processes such as phonological processing are consistently found to be important to LD identification throughout the research literature. However, there have been many theories of LD and its causes that have been less popular and not received empirical support. Some of these will be reviewed in the next section.

**Contentious cognitive processing deficits in learning disabilities**

When attempting to develop a model of LD identification based on cognitive processing deficits, it is important to examine some of the other processes that have not received much empirical support in previous research. Many types of processes have been put forward as a cause of LD in the past. However, many of these theories have since been discredited, or remain controversial.

One of the more popular explanations for LD in earlier research was that it was due to a visual impairment (e.g., Casey & Ettlinger, 1960; McKeever & Van Deventer, 1975). However recent studies have been fairly consistent in the finding that this is not the case (Goulandris, McIntyre, Snowling, Bethel & Lee, 1998; O’Brien, Mansfield & Legge, 2000). A large group of respected ophthalmologists went so far as to write an article making it clear that vision problems were not a cause of LD and that treatment programs based on vision difficulties were not supported by current research (Lueder et al., 2009).

Another theory put forward was that auditory processing deficits are the cause of the phonological processing deficits that cause LD (Goswami, 2003). Similarly, another theory, the auditory temporal deficit theory (Tallal, Stark, Kallman & Mellits, 1980)
stated that students with LD will be less effective in discriminating between any types of auditory stimuli, not just speech. However, once the presence of Attention Deficit Hyperactivity Disorder (ADHD) or attentional deficits is controlled for, auditory processing deficits are no longer present in an LD cohort (Breier, Gray, Fletcher, Foorman & Klaas, 2002; Willburger & Landerl, 2010).

Another theory is that the deficits behind LD are actually in perception, attention and memory (Ahissar, 2007) and while this theory has received some support (Chandrasekaran, Hornickel, Skoe, Nicol & Kraus, 2009) the confounding factor of attention is yet to be accounted for. Furthermore, studies that have tested this have garnered results that directly contradict the foundations of this theory (Di Filippo, Zoccolotti & Ziegler, 2008). The PASS (Planning, Attention, Successive and Simultaneous) cognitive processing theory of LD was less successful at predicting LD than the double deficit theory (Das, Georgiou & Janzen, 2008) and has received little attention since.

A cerebellar deficit has also been suggested that states that cerebellar tests differentiate between discrepant and non-discrepant readers (Fawcett, Nicolson & Maclagan, 2001). This was further supported by a Positron Emission Tomography (PET) imaging study that found decreased cerebellar activation in the right hemisphere during a memory based task (Nicolson, Fawcett & Dean, 2001). This hypothesis however is tied to the once popular theory that LD might stem from motor or balance deficits (e.g., Nicolson & Fawcett, 1994), a theory that has since been refuted consistently (e.g., Savage, 2007; Savage & Frederickson, 2006).

Some theories on the deficits underlying LD are more direct and measure reading itself, for instance syntactic difficulty (Scarborough, 1991). However this theory was not supported by a study that found that syntax is not an issue for LD readers (Leikin & Assayag-Bouskila, 2004). Another theory was that LD could be identified on the basis of word recognition, comprehension or vocabulary deficiencies (Aaron et al., 2008), or that those who have good decoding skills but poor comprehension with no other deficits also be designated LD (Catts, Adlof & Weismer, 2006). Theories based on any kind of measure of reading performance task alone run the risk of only identifying poor readers, not those with a neurological disorder. Such a method would be at odds with the widely accepted notion that LD has a neurological
basis (Bach et al., 2010; Pugh et al., 2001; Rumsey et al., 1999; Simos, Breier, Fletcher, Foorman et al., 2000).

Each of the processes that will be used in the current study to predict the presence of LD will be addressed in the following sections of this chapter. These processes are phonological processing, naming speed and verbal memory. In addition, the difficulties inherent to the use of orthographic processing as a cause, or reflection, of LD will be discussed.

**Phonological processing**

Phonological processing is “the ability to perceive and deliberately manipulate the sound components of words” (Kirby et al., 2008, p.105) and is probably the most commonly discussed cognitive deficit in LD literature. In a review of studies on reading and LD, Shankweiler and Fowler (2004) concluded that the role of phonological processing, despite the massive amount of attention it received in the relevant literature, was not overrated. Studies consistently show that the primary deficit in LD is of a phonological rather than semantic nature (Hanly & Vandenberg, 2010) and the role of phonological deficits in LD has been supported by neurological studies (e.g., Pugh et al., 2000).

Longitudinal studies have provided support for the genetic base of phonological processing deficits (Carroll & Snowling, 2004; Castles et al., 1999). Once general language skills are controlled for, phonological processing could be used to differentiate between those with a hereditary risk of LD and those without (Puolakanaho et al., 2007). The gene responsible for the phonological buffer deficits that contribute to LD has been identified (Bates et al., 2010). Studies that divide LD into phonological and surface dyslexia subtypes can also provide information on the importance of phonological processing in LD. Phonological dyslexia has also been shown to have a stronger genetic component than surface dyslexia, which is thought to have a stronger environmental aetiology (Castles et al., 1999).

Neurological studies have validated the theory that phonological processing is an important aspect of reading for both RA and LD students (Perfetti & Bolger, 2004). Recent neurological evidence suggests that phonological deficits are the core deficit within LD (Goswami, 2011). In one case study, a nine year old girl with left hemisphere damage had difficulty with phonological processing and reading that was
very similar to the difficulties experienced by students with LD (Pitchford & Funnell, 1999). Cerebral volume, planum temporale asymmetry and the size of Heschl’s gyrus can accurately predict the existence of either LD with phonological deficits, or specific language impairment, a disorder that is also thought to be in part caused by phonological deficits (Leonard et al., 2001). In reading tasks that did not involve phonological assembly, LD and RA readers displayed similar left hemisphere activation. In contrast, tasks that were dependent on phonological processing resulted in marked differences in left hemispherical functional connectivity between LD and RA readers (Pugh & Mencl, 2000). Furthermore, event related potential studies have been used to demonstrate qualitative neurological differences between LD and non-LD students when processing phonological information (Fosker & Thierry, 2005).

Longitudinal studies have demonstrated that early phonological processing can predict later reading achievement or the presence of LD (David, Wade-Woolley, Kirby & Smithrim, 2007; Linklater, O'Connor & Palardy, 2009; Muter & Snowling, 2009; Storch & Whitehurst, 2002; Torgesen & Wagner, 1994). One Australian longitudinal study found that phonological processing tests in kindergarten could be used to predict poor reading skills in Grade 2 (Jorm et al., 1986). It should however be noted that the predictive power of phonological processing is not always consistent with results from some studies indicating that other cognitive processes need to be utilised in order to increase accuracy (e.g., Johnson, Jenkins, Petscher & Catts, 2009).

It is important to take the language of studies into account, as the importance of phonological processing may vary as a function of the orthographic depth of the tested language (Silver et al., 2008). However, there has been evidence that phonological processing is still important in orthographically shallow languages such as German (Mayringer & Wimmer, 2000), Korean (Pae, Sevcik & Morris, 2010) and Czech (Caravolas & Volin, 2001). The role of phonological processing in character based languages such as Chinese is less clear as there is evidence that other processing deficits such as naming speed seem to play a larger role (Ho, Chan, Lee, Tsang & Luan, 2004). Nine-year-old Finnish students with a reading disability were significantly poorer at recognising differences in phoneme discrimination than their RA counterparts (Hamalainen et al., 2009). Studies such as these indicate that
phonological processing, although very important to literacy and LD research, may not be the only cognitive processing deficit that is of importance.

Phonological processing also seems to be predictive of future reading ability or problems at a wide range of ages. Nonword reading at Grade Three actually predicts a student’s performance on a Year 12 Wordchain test better than the performance on the same Wordchain test in Grade Three (Svensson & Jacobson, 2006). Nonword reading at age 11 is also predictive of reading ability at age 14 (Conti-Ramsden & Durkin, 2007). Furthermore, LD children between 7.5 and 9.5 years of age consistently had phonological deficits, regardless of comorbidity with arithmetic or behavioural disorders (Shankweiler et al., 1995).

Furthermore, it has been proposed that the relationship between phonological processing and reading is not strictly linear, with phonological processing less important once reading reaches a certain proficiency level (Schatschneider et al., 2002). This leads to the suggestion that the correlation between phonological processing and reading ability is actually higher in poor readers, but lower in good readers because the relationship between the two is not as strong in proficient readers (Anderson, 2008). This proposed lack of linearity also has statistical implications as most analyses assume linear relationships between variables.

Another interesting aspect to the relationship between reading and phonological processing is that phonological processing may present a type of ‘hurdle requirement’ when learning to read. Over twenty years ago it was suggested that some factors in a child’s reading progress could not have any effect until phonological processing reached a certain level before higher level reading related skills could improve (Juel, Griffith & Gough, 1986). In theory this has been supported by a longitudinal study that found a causal relationship between phonological processing and word decoding (Wagner, Torgesen & Rashotte, 1994). It could be possible that when learning to read, certain skills such as phonological processing need to reach a certain level before the next skill, say word decoding, can be perfected or learned. Students that linger on one of these stages are then of course prone to the Matthew Effect.

The impact of reading on phonological processing could in part be explained by models incorporating both top down and bottom up theories of reading (Verhoeven,
Reitsma & Siegel, 2010). Phonological processing influences reading, but it is possible that reading also influences phonological processing. This is why studies utilising reading age controls are so important. They can, to a certain extent, control for the reciprocal relationship between phonological processing and reading.

**Reading age-matched controls**

Despite the use of tests of phonological processing over the decades, how to best utilise them has remained an issue. Age-based norms do not take into account the different rates at which children learn how to read (Chard & Dickson, 1999). This can in turn, affect phonological processing. One of the most interesting aspects of phonological deficits in LD is that LD students have lower phonological processing skills than not only their peers, but also those younger students who are reading at the same standard as them. This reading level matched design, used by Snowling (1980), supported the theory that LD readers are qualitatively different to their RA peers and not just lagging behind. The existence of reading age phonological deficits in LD students has been supported by other studies that have found reading age-matched nonword reading deficits (Siegel & Ryan, 1988).

Reading-age comparisons can provide valuable information on qualitative differences in the reading ability of students. For instance, consider a Grade 3 student who is reading at a Grade 1 level, but does not have the phonological processing abilities of a regular achieving Grade 1 student. This is an indication that the problems that the student is facing are not simply developmental and instead may be due to a processing deficit (Stanovich, 1988). In a study with LD students and controls matched on reading ability and intelligence, students with LD read both regular and nonwords significantly worse than their controls (Kochnower, Richardson & DiBenedetto, 1983). Siegel and Ryan (1988) compared the phonological abilities of those with LD to younger reading matched controls and demonstrated that the LD students still had significantly poorer phonological skills than controls.

An excellent example of the effectiveness of reading-matched designs can be found in a study on orthographic and phonological processing (Martin, Pratt & Fraser, 2000). A group of LD students was compared to chronological and reading matched controls. The LD group’s phonological processing was commensurate with that of
reading matched controls. However their orthographic processing was considerably better than reading matched controls and closer to the level of the chronologically matched controls.

In a summary of nonword reading deficits in LD students, Rack and colleagues (1992) point out the issues inherent in different reading matched studies, namely the method by which they are matched. For instance, if they are matched on regular word reading and some of the regular words used in the study are not frequently seen by the participants, then the students in the study will be phonologically processing the words. Consequently, differences in nonword reading ability between matched groups would not be expected. Additionally, studies matched on paragraph reading accuracy report smaller deficits than those that are matched for reading accuracy (Herrmann, Matyas & Pratt, 2006). It should be noted that in this study reading age nonword deficits are still significant for LD readers, but reduced in magnitude, for these students.

**Nonword reading**

There has been debate over what tests of phonological processing are the best predictors of reading. Muter, Hulme and Snowling (1998) used a modified parallel analysis, confirmatory factor analysis and item response theory in order to assess the unidimensionality of tests of phonological processing. Overall they found that these tasks were all tapping into a single construct (Schatzneider, Francis, Foorman, Fletcher & Mehta, 1999). This indicates that there is a range of tasks that could be reliably selected to test phonological processing.

One thorough longitudinal study by Muter, Hulme and Snowling (1998) investigated the impact of phonological processing, measured by a phoneme deletion test and a rhyming test, on reading ability. They found that although different tasks accounted for different ratios of unique variance over the years, phonological processing as a whole accounted for much of the variance in reading ability over time.

Phoneme segmentation is another task that is often used to assess phonological processing in LD research. However, it is quite a difficult task, even to those who are beginning to read well and, therefore, may not be well targeted to poor readers (Murray, 1998). Other tests have been developed that involve investigating the types of phonological errors made within the tasks (Moseley, 2004), but these have not
attracted widespread attention, presumably due to the length of time required to interpret the results.

It has been proposed that tests involving nonwords may be a better gauge of word-level reading than tests using real words (Spear-Swerling, 2004). Nonwords are made-up words with no meaning that are pronounceable. For instance *blurk* is a nonword while *thdfaxc* is not. LD students have a well established difficulty in reading nonwords. This has been attributed to the fact that the nonwords are novel stimuli that the reader could not have memorised from previous exposure and must instead be processed phonologically (Snowling, 1981). Although a quick examination of the literature will produce many studies utilising nonwords as a measure of phonological processing, it is important to differentiate between nonword reading and repetition. Nonword reading is where a student is visually presented with nonwords to read out loud and nonword repetition requires both the presentation and response to be oral.

Although it has been claimed that nonword repetition “probably represents the most effective predictor of language learning that is currently known” (Gathercole, 2006, p. 213) this claim was made in the context of specific language impairment, rather than LD. This is an important distinction, as specific language impairment does not have the reading component that is so important to this area of study. Furthermore, it is important to take into account that nonword repetition, as compared to nonword reading, is meant to be a test of short term memory, as well as phonological processing (Santos, Bueno & Gathercole, 2006). This would provide a confounding factor if short term memory is also tested. This concern is supported by an event related potential study that found evidence for the hypothesis that nonword repetition is actually assessing the phonological loop (Barry, Hardiman & Bishop, 2009). As will be discussed later in the chapter, this is a construct that has been linked to verbal memory.

Not only is nonword reading thought to be the “primary index of phonological coding ability” (Stanovich & Siegel, 1994, p. 28), it is also considered to be the best single test of the presence of LD (Siegel, 1989; Stanovich, 1999). Nonword reading skills seem to be not just quantitatively different in LD and non LD populations, but also qualitatively different. For instance, while there is a strong correlation between
nonword reading and age for RA readers, there is no significant correlation between nonword reading and age for LD readers (Griffiths & Snowling, 2002).

The make up of nonwords is important and needs to be taken into account when developing a new list. The number of syllables (Rispens & Parigger, 2010), word length (Martens & de Jong, 2006; Ziegler, Perry & Coltheart, 2000), consonant clusters (Kelly, 2004), letter frequency (Lupker, Perea & Davis, 2008) and number of phonemes (Perry, Ziegler & Zorzi, 2010) all mediate the difficulty of reading nonwords. Consequently, there needs to be a variety of combinations of these factors in any newly-developed nonword list.

In this section the evidence for the causal role that phonological processing plays in LD has been reviewed. Furthermore, research on reading age-matched phonological deficits has also been discussed. Finally, various tests of phonological processing have been reviewed with a focus on nonword reading, which will be used in the current study. In the next section, the role of naming speed in LD and reading will be reviewed.

**Naming speed**

The discovery that students with LD were significantly poorer than their RA counterparts at Rapid Automatic Naming (RAN), a task that measures naming speed, was made in the 1970s (Denckla, 1972b; Denckla & Rudel, 1976a, 1976b). In these RAN tasks, students were asked to read out loud a list of presented stimuli, such as letters or colours, as quickly as they could. However, it was not until the double deficit theory was put forward in 1999 (Wolf & Bowers) that this particular cognitive processing deficit was given much attention. Naming speed is thought to be a representation of the ability to form letter strings in memory for the recognition of words. This relationship between naming speed and reading ability can not be attributed to either semantic processing or traditional processing speed (Conrad & Levy, 2007). Although processing speed measures are related to measures of naming speed, they do not seem to have any influence on the relationship between naming speed and reading (Powell et al., 2007).

There is neurological support for naming speed as an organic cause of LD. For instance there is a strong positive correlation between reading skill and activation of the occipito-temporal region of the brain, an area activated in naming speed tasks.
There have also been genetic studies that have supported the role of naming speed deficits in LD. For instance, a gene thought to affect naming speed could also be a candidate for causing LD (Couto et al., 2008). However, some regard the role of naming speed in reading and LD as part of phonological processing ability (e.g., Phinney, Pennington, Olson, Filley & Filipek, 2007).

Despite this debate, the theory that naming speed has predictive value that is unique to phonological processing is well accepted (de Jong & van der Leij, 2002; Schatschneider & Torgesen, 2004). In a study by Kirby, Parrila and Pfeiffer (2003), phonological processing and naming speed both predicted reading ability, with each contributing unique variance. Furthermore, Kirby and colleagues also found that the relative importance of naming speed in this longitudinal study increased over time. Waber, Forbes, Wolff and Weiler (2004) actually found that naming speed deficits were more important than phonological processing deficits in identifying LD. When they attempted to find double deficit driven groups in an LD population they did not find any participants that fit into a phonological processing deficits only group.

There is some evidence for the two constructs each contributing unique variance to reading (Kirby, Georgiou, Martinussen & Parrila, 2010; Savage & Frederickson, 2005). It is possible that even if naming speed is testing a portion of phonological processing, it is assessing a specific aspect of it that is of importance, such as phonological retrieval (Catts & Hogan, 2003). Furthermore, naming speed has also been linked to orthographic processing (Kirby et al., 2010) supporting the theory that, similar to many of the cognitive processes involved in reading, the relationship could involve overlap between these processes, as well as each process contributing unique variance.

Not all studies however have found that naming speed is a deficit relevant to LD. Some have had mixed results (Waber et al., 2001) while others have found that naming speed is a better predictor of age than of reading ability (Stanovich, 1988), or an outcome, rather than cause of LD (Wong & Ho, 2010). Furthermore, unlike phonological deficits that have been shown to be significantly lower than that of reading age-matches, naming speed deficits are found when LD students are compared to students of their own age (Chung et al., 2008; Griffiths & Snowling,
LD students have slower naming speed than chronological controls, however they address the naming speed tasks in a manner similar to their reading age controls (Wong & Ho), making reading age deficits unlikely. This does not mean that the deficit is irrelevant to LD. Instead it is more difficult to demonstrate that the problem is a deficit and not a developmental lag.

Some of the conflicting results around the role of naming speed and reading might be because naming speed is thought to be a better predictor of reading in poor readers than in good readers (Kirby et al., 2010; Neuhaus, Roldan, Boulware-Goorden & Swank, 2006; Truman & Hennessey, 2006). It is possible that as long as naming speed is at a sufficient level, then it is not important for reading ability, it is only important if there is a deficit. In support of this proposal it has been found that the predictive value of naming speed in reading is thought to decrease with time, only providing unique variance in primary school students (Ferrer et al., 2007; Meyer, Wood, Hart & Felton, 1998). Furthermore, the rate of improvement in LD and non LD students in naming speed tasks between the ages of seven and eleven years is consistent, suggesting that whatever deficits existed, were in place before the age of seven (Weiler, Forbes, Kirkwood & Waber, 2003). This does not change the importance of the process, but is important to note any studies looking at developing a screening tool for an older cohort.

There is evidence to suggest that naming speed deficits play a more important role in LD in languages that do not rely on phonological processing, such as Mandarin (Ding, Richman, Yang & Guo, 2010) and even Greek (Georgiou, Parrila & Liao, 2008). Furthermore it has been demonstrated that Chinese LD students had deficits in orthographic processing and naming speed, rather than phonological processing (Ho et al., 2004). Naming speed also tends to be a bigger problem in LD students who speak a language with a more transparent orthography than English, such as Spanish (Serrano & Defior, 2008).

In addition to the importance of naming speed deficits in LD crossing over languages, naming speed deficits also do not seem to be exclusive to LD readers with high IQs. When poor readers are split into average and low IQ groups, approximately half of both groups display naming speed deficits (Catts, Gillispie, Leonard, Kail & Miller, 2002). Results such as these are difficult to interpret as
there is no way of knowing which of these students actually have LD and which of
them are poor readers.

**Rapid automatic naming**

Unlike phonological processing, which can be measured through a number of tests,
there is really only one commonly used test of naming speed. This is in part because
the deficit was noticed specific to the task. Denckla found that boys with LD had a
pronounced difficulty performing Rapid Automatic Naming (RAN) tasks when
compared to their RA peers (1972b). This RAN test require students to name a list
of stimuli, such as letters, numbers, colours or objects, as quickly as they can and
scores are represented by the time it takes to complete the task.

The methods of administering RAN tests are fairly consistent, however there are
some variations. For instance it has been found that non-alphanumeric stimuli such
as colours or shapes are better predictors of those in the process of learning to read,
such as in the first year or two of primary school, while letters or numbers are more
appropriate once some proficiency in reading can be assumed (Misra, Katzir, Wolf &
Poldrack, 2004; Wolf et al., 2000). This could depend on the target language as
well, for instance objects are a better predictor of reading in Chinese (Ding et al.,
2010), a character based language. Other studies have found neurological evidence
to suggest that letters and objects were processed the same way in either adults or
children (Turkeltaub, Flowers, Lyon & Eden, 2008). Because of this, it is better to
test both alphanumeric and non-alphanumeric stimuli, especially in a young cohort,
in order to ascertain which test is best suited to any given population of interest.

Not only are the target stimuli important, the manner in which RAN tasks are
administered also needs to be taken into account. For instance continuous formats,
such as those with all the stimuli on one page, are preferable in most cases to discrete
formats where one letter is presented at a time. This is because the extra resources
the continuous format draws on are similar to those required for reading and are thus
more relevant (Denckla & Cutting, 1999).

**Verbal memory**

The moderate success of the double deficit theory has resulted in both phonological
processing and naming speed deficits becoming widely accepted as inherent to LD.
However, the role of verbal memory has received much less attention. One of the
issues here may possibly be that there is overlap in the research on, and operationalisation of, working memory and short term memory. Although both types of memory are associated with the immediate retention of a small amount of information, working memory also commonly denotes the manipulation of that information but short term memory does not (Richardson, 2007).

It has been suggested that it will be difficult to untangle the respective roles of phonological processing and working memory in reading (Gathercole, Briscoe, Thorn, Tiffany & Team, 2008) and some have suggested that, similar to naming speed, such deficits are actually part of phonological processing (Gupta, 2006). Tasks that refer to verbal information and response that are testing short term or working memory, are thought to utilise the phonological loop, and are referred to as verbal memory tasks (Helland & Asbjørnsen, 2004; Savage et al., 2005). This phonological loop is thought to be made up of the phonological store and articulatory control processes (Baddeley, 1990). Verbal memory deficits are thought to stem from inefficiency in the phonological loop (Steinbrink & Klatte, 2008).

Working memory is thought to play a big part in both reading comprehension (Siegel, 1999) and vocabulary acquisition (Jones, Gobet & Pine, 2007) and short term memory has been linked to linguistic functioning (Kibby et al., 2004). It has been suggested that poor readers are less effective at storing phonological information in short term memory than good readers (Macaruso, Locke, Smith & Powers, 1996). Furthermore the role of short term memory in specific language impairment has been long established (Conti-Ramsden & Durkin, 2007) and memory deficits in a range of disorders with high comorbidity such as Asperger’s Syndrome and ADHD also exist (Alloway, Rajendran & Archibald, 2009). It is even possible that working memory may be an underlying deficit of all types of LD, with deficits also found in those at-risk for serious math based LD (Swanson & Beebe-Frankenberger, 2004).

Although there is debate about the respective roles of short-term and working memory in LD, it has been found that in an LD cohort that memory deficits are specific to the phonological loop (Kibby & Cohen, 2008). As such the memory deficits in LD can be conceptualised as being specific to verbal memory, and this is how these deficits have been conceptualised in some LD research (e.g. Cornwall,
1992; Helland & Asbjørnsen, 2004; Savage et al., 2005; Scarborough, 1998). This is not to say that memory deficits would be specific to language based LD. Studies have shown that the phonological loop is also important in mathematics (Geary, Hoard, Byrd-Craven, Nugent & Numtee, 2007). Rather it indicates that memory deficits may be important to all types of LD, rather than those specific to speech. Some have suggested that the link between phonological processing and verbal memory is so strong that verbal memory does not provide unique variance after accounting for phonological processing (Savage, Lavers & Pillay, 2007). Others have argued that, if testing both naming speed and phonological processing, then tests of verbal memory are not required (Schat Schneider & Torgesen, 2004). However, it has been demonstrated by others that LD students do have deficits in verbal memory (Gang & Siegel, 2002; Siegel & Ryan, 1988). This has also been found in an Australian study (Brown & Greaves, 2001). Furthermore, a meta-analysis, conducted in order to ascertain if memory deficits were representative of a lag or a deficit, found support for a deficit rather than a lag (Swanson, Zheng & Jerman, 2009).

Additionally, verbal memory problems increase over time. For example Gaultney (1998) found that Grade 3 LD students had lower verbal memory scores than their non-LD counterparts and that the gap between the two groups grew over Grade 4 and 5. Jeffries and Everatt (2004) found in a later study that not only did an LD cohort have significantly poorer verbal memory than their RA peers, but also poorer than a group of students with special education needs other than LD.

There is genetic evidence for the role of working memory in LD (Berninger et al., 2008). Despite the fact that these deficits may be inherited, there is also evidence that they are treatable. Compensatory paths for those with phonological memory deficits can be developed, albeit with lengthy remediation (Gathercole et al., 2008). Further evidence for the role of memory deficits in LD came from Elbro’s (2010) study of French students with LD. Although there was no consistent neuropsychological profile in the LD cohort, the Working Memory Index was consistently low. It has also been found that students with LD have significantly lower IQ scores if the test used has a higher focus on verbal memory than if the test does not (Dickerson Mayes & Calhoun, 2007).
The importance of working memory in relation to IQ needs to be taken into account in any LD research utilising the discrepancy method of identification. Working memory is thought to affect both crystallised and fluid intelligence (Swanson, 2008b). Furthermore, it was also the best predictor of g, latent intelligence, in a structural equation model (Colom, Rebollo, Palacios, Juan-Espinosa & Kyllonen, 2004). Therefore if g, the concept that IQ was developed to measure, is strongly correlated with working memory, which in turn shares properties with verbal memory, then it could be suggested that students with deficits in verbal memory would have a reduced IQ score. This, in turn, could decrease the probability of a student receiving an LD diagnosis via the discrepancy method of identification (Siegel, 1989).

If this theory is taken in conjunction with finding that the many studies that have indicated that LD and Low Achievement (LA) students have similar deficits in phonological processing and naming speed (Ellis et al., 1996a), then it is possible that many in the LA cohort actually have LD as well, but with an additional verbal memory deficit that reduces their measured IQ. This is not to say that all students in both groups have LD. As discussed in Chapter 2, there are reasons other than LD that could cause a discrepancy or low reading achievement. Instead it could be that both groups contain students with genuine LD and that this has been confounding results.

Support for this theory that many discrepancy-designated LA students actually have LD can be found in a study with a very liberal definition of LD. Students with a two year lag in reading ability and with an IQ over 80 were designated LD and were significantly poorer than chronological age controls in a verbal memory task but they did perform in a similar manner to reading-age controls. Furthermore the same study found that verbal memory deficits in LD students cannot be explained by deficits in naming speed (de Jong, 1998). The fact that the IQ exclusionary criterion was lower than what is often used could result in students that would not usually meet a discrepancy definition of LD being identified as LD in de Jong’s study. Consequently those LD students whose IQ has been reduced by verbal memory deficits were included and therefore deficits in verbal memory were significantly worse than controls. This lowering of IQ scores as a result of cognitive processing deficits is known as the Mark Penalty (Dumont & Willis, 1999).
**Digit-span**

Deficits in phonological or verbal memory in LD students can be easily assessed by digit-span tests (Service, 2006). In these commonly-used tests, participants recite back random numbers read out to them either in the same order or backwards from the order that they were read out to them. Forward digit-span is a well accepted measure of short term memory (Conway et al., 2005; Gathercole, 1999; Kauffman, 1994; Savage, Lavers et al., 2007) but there is debate over what backward digit-span is testing.

Some researchers have stated that backwards digit-span is a measure of working memory, as the act of reversing the order of the numbers involves more than rote repetition (e.g., Baddeley, 1986; Baddeley et al., 1998; Gathercole, 1999; Kauffman, 1994; Savage, Lavers et al., 2007; Savage et al., 2005). Others, however, believe that the task is not taxing enough to warrant the use of working memory and is instead a measure of short term memory similar to the forward digit-span task (e.g., Cantor, Engle & Hamilton, 1991; Conway et al., 2005; Engle, Tuholski, Laughlin & Conway, 1999; Richardson, 2007). Others still believe that digit span backward results are influenced by attention and executive function (Hale, Hoeppner & Fiorello, 2002). This debate is beyond the scope of the current project, however as the role of memory in LD has received less research than that of naming speed and phonological processing this is both tasks are worthy of further investigation. One of the only points of agreement in this argument is that the two should not be added into a total score as there is some difference between the two tasks (Hale et al; C. Reynolds, 1997) and the two tasks tap into different parts of the brain (Sun et al., 2005). Because both forward and backward digit-spans have been described as measures of verbal memory (Helland & Asbjørnsen, 2004; Savage et al., 2005), analogous to the phonological loop, the term verbal memory will be used instead in the current study when referring to results.

LD students are thought to have deficits in both forward and backward digit-span tests (Helland & Asbjørnsen, 2004). LD students with a family history of the disorder also tend to score very low on digit-span tests (Berk, 1983). However, similar to naming speed and unlike phonological processing, these deficits are thought to exist when compared to chronological and not reading age controls (de Jong, 1998; Snowling, 1996). One of the greatest advantages of using the digit-span,
especially in this field, is that it is commonly used and thus invites generalisability of results, a rare occurrence in LD research.

**Orthographic processing**

Orthographic processing is “the ability to form, store, and access orthographic representations” (Stanovich & West, 1989, p. 404). These representations when testing reading are usually words. There are claims that orthographic processing has been ignored in LD research and that there are links between phonological processing, naming speed, short term memory and orthographic processing with reading and spelling (McCallum et al., 2006). One theory, based on extending the double deficit theory, involved adding orthographic processing deficits to make a triple deficit (Badian, 1997; Stage, Abbott, Jenkins & Berninger, 2003).

Recent research has indicated that one of the primary issues in LD may actually be the integration of orthographic and phonological information (Cao, Bitan & Booth, 2008; Desroches et al., 2010). However, how this integration would be operationalised is currently unclear. There has also been debate about a phonological-orthographic trade off, suggesting that the importance of the relationship between the two processes may not be consistent over different levels of phonological or orthographic processing proficiency (Foorman, Francis, Fletcher & Lynn, 1996).

One of the ways in which the role of orthographic processing in LD can be viewed, is through a model incorporating both surface and phonological dyslexia as subgroups of LD. Those with surface LD do not have the typical phonological processing deficits, but have marked orthographic processing deficits (Friedmann & Lukov, 2008). Neurological studies have shown that surface dyslexics have increased activation in areas associated in phonological processing and that phonological dyslexics have increased activation in areas associated with orthographic processing, possibly as a means of compensation for their deficits (Price & Mechelli, 2005). However, as discussed in Chapter 2, previous studies have indicated that surface dyslexia is more indicative of environmental difficulties than it is of a neurological disorder (Castles et al., 1999). As such it may be that use orthographic processing deficits may be a reflection of poor reading, rather than a cause of LD.
High correlations between phonological and orthographic processing should not be considered surprising as it has been shown that people refer to their orthographic knowledge even when performing phonological tasks (Castles, Holmes, Neath & Kinoshita, 2003). However, some students with LD have phonological deficits and orthographic processing as a compensatory strength (Lavidor, Johnston & Snowling, 2006). This has been supported by studies that have found significantly higher orthographic processing abilities in LD students than in reading age controls (Martin et al., 2000; Siegel & Ryan, 1988). Further confounding the high correlation between the two processes, a small cohort of LD students that have significantly better phonological than orthographic processing skills has also been identified (Lachmann, Steinbrink, Schumacher & van Leeuwen, 2009).

Long term issues stemming from orthographic processing deficits may be less serious than those stemming from phonological processing deficits. A longitudinal study by Byrnes, Freebody and Gates (1992) split their initial cohort into four groups on the basis of nonword or irregular word reading, representing phonological and orthographic awareness. Those good at both tasks formed one group, as did those who did poorly in both tasks. Finally a Phoenician group, good at nonword reading but poor at irregular word reading, and a Chinese group, good at irregular word reading but poor at nonword reading were identified. The Chinese group had stronger reading comprehension skills than the Phoenician group at Time 1 but, over time, this reversed and the Chinese group got progressively worse. This seems to indicate that orthographic processing deficits are not representative of a permanent neurological disorder.

This notion has support in the literature. It has been suggested that phonological processing deficits are indicative of a neurological disorder and that orthographic processing deficits are, instead, attributable to exposure to print (Griffiths & Snowling, 2002). Other studies have also attributed orthographic deficits to environmental factors (e.g., Braten, Lie, Andreassen & Olaussen, 1999; Castles et al., 1999). Therefore, deficits in orthographic processing will not be utilised in the current study as a LD marker variable. Instead, poor readers who are also poor in orthographic processing, but not any of the other processes, could be earmarked in the classroom as those who are most likely to respond to intervention. This is because the reviewed research indicates that orthographic processes do not have the
same organic base as other deficits. Although this is something that still needs to be tested, it may provide an alternate explanation for reading difficulty if LD is not identified.

**Irregular word reading**

Irregular words are words that are not pronounced the way they are spelt, such as *yacht*, and make up approximately a quarter of the words in the English language (Coltheart, 2005). Irregular word reading is considered a good measure of orthographic processing (Kirby et al., 2008) as it requires readers to recognise the whole word rather than reading it phonetically.

The relationship between irregular and nonword reading is complex and not necessarily consistent across all students. For instance, the correlation between irregular word reading and nonword reading is higher in a RA group than in an LD cohort (Griffiths & Snowling, 2002). Furthermore the high correlation between reading irregular words and nonwords has been explained by the fact that people still rely on orthographic processing when reading novel words (Castles et al., 2003). However, nonwords can not be recognised in the same way that irregular words can and, as such, it is important to test both processes separately.

It has been found that although LD students perform as well as reading age-matched controls on general spelling tests, they perform significantly worse in experiments that involve factors such as consonant clusters (Bruck & Treiman, 1990). Furthermore, letter frequency (Mason, 1975) and the number of syllables and phonemes (Treiman & Zukowski, 1996) have also been shown to be mediating variables in irregular word reading. Consequently, variables such as this need to be taken into account in the development of new lists.

**Multiple deficits**

As discussed in Chapter 2, LA students with the cognitive weaknesses inherent to LD are often used as the example of why IQ tests are inappropriate for LD diagnosis. However this argument is usually made with the follow up claim that RTI be utilised instead (Meyer, 2000). Rarely is it suggested that those same cognitive processing deficits be used as a diagnostic tool instead, despite the fact they are the barometer that current success or failure are often compared to.
There have been specific LD phenotypes based on cognitive processing put forward of late, such as the phonological processing and short term memory deficit phenotype (Bonifacci & Snowling, 2008). There is also a suggestion that, as per the double deficit hypothesis, multiple deficits result in compounded difficulties (Shaywitz & Shaywitz, 2005). For instance, readers with naming speed deficits rely more heavily on working memory than those who do not (Baddeley, 1986). If a student had deficits in both of these processes, this would surely cause more difficulty.

There are successful examples of utilising multiple cognitive processing deficits to predict the presence of LD. Researchers have developed cognitive processing measures to identify Chinese adolescent LD readers within a multiple deficit paradigm (Chung, Ho, Chan, Tsang & Lee, 2010). Phonological processing, short term memory and naming speed tasks were amongst those successfully used to identify LD. Due to the character base of the Chinese language, orthographic processing was more important than phonological processing in this study. This flexibility between languages could be another advantage of a multiple deficit model.

Phonological processing and naming speed in kindergarten can predict reading in Grade 2, and impressively, Year 8 (Adlof, Catts & Lee, 2010). Similar constructs were also able to successfully predict the presence of LD with moderate power (Puolakanaho et al., 2007). The most effective way of predicting reading ability in LD and RA cohorts may be different. A longitudinal study by Scarborough (1998) used reading ability in Grade 2 to predict reading ability in Year 8 in a RA cohort. However, in order to improve the accuracy of reading prediction in the LD group, measures of verbal memory, naming speed and phonological processing were required.

Other multiple deficit models have been developed such as the multiple deficit hypothesis (Ho et al., 2002) in which naming speed, visual, phonological and orthographic deficits in an LD cohort. Another model put forward the cognitive hypothesis testing model tested used a number of processes to identify LD within an RTI paradigm (Fiorello et al., 2006). Finally a triple deficit model with phonological, orthographic and naming speed deficits has also been put forward.
The current study will be conducted on a similar premise to these studies but the processes tested will be different. Ultimately the primary aim of developing a more accurate LD identification paradigm is so that more effective remediation can be facilitated. It is vital that groups of students are identified correctly before appropriate remediation for these groups can be developed (Wodrich, Pfeiffer & Landau, 2008). This aim has caused difficulty in LD research as LD students make up such a diverse cohort. However, if students can be identified on the basis of their deficits, relatively homogeneous groups can be developed on the basis of this and appropriate remediation formulated.

In this section the role phonological processing, naming speed and verbal memory as predictors of LD have been discussed. The role of orthographic processing in LD was also examined, but currently research dictates that is more appropriate as an alternate explanation for poor reading. In the next section, the pragmatic considerations that need to be taken into account when developing a screening tool for LD will be discussed.

The Integration of Theoretical and Applied Learning Disabilities Research

The integration of relevant school psychology and neuropsychology research is important (Cleary & Scott, 2011) but slow (Decker, 2008). This could be due, at least in part, to the fact that researchers often employ methods that make sense within a research context but are difficult to apply (Freebody, 2007). The research-to-practice gap in this field is a sizable impediment to assisting students with LD (D. Fuchs & Fuchs, 1996). This has been highlighted by an intervention program put in place in the US which worked well when implemented by researchers. However, when control of the program was handed over to the teachers, it was abandoned within a year (D. Fuchs, Fuchs, Harris & Roberts, 1996). D. Fuchs and colleagues concluded the main problem was that the program promoted did not fit in well with the funding and structure that was already in place in the school.

It is important that considerations such as applicability and cost are taken into account in research in this field. While literacy levels in Australian schools dropped slightly from 1975 to 1998, the expenditure per student has risen significantly (Leigh & Ryan, 2008). It is reasonable to suggest that a part of this rise in costs is representative of the rise of students identified as in need of special education and
the cost of assisting these students. It has been noted in the US that much of the money put towards special education students is spent on screening and diagnosing students, rather than actually assisting them (D. Fuchs & Fuchs, 1995).

Much of this difficulty in identifying LD students can be attributed to the many causes of reading difficulties beside LD. This is why there is exclusionary criterion such as cultural disadvantage in most operationalisations of LD. Some longitudinal studies have identified home life risk factors for reading achievement (Chard et al., 2008), others have demonstrated that parent’s beliefs and behaviours influence reading achievement (Davis-Kean, 2005), or that grammar skills are affected by upbringing and education that influence reading development (Muter, Hulme, Snowling & Stevenson, 2004).

There are other external factors, aside from those of a socio-economic nature, that need to be taken into account. For instance, the increasing importance of accommodating students for whom English is a second language has been recognised (Singleton & Vincent, 2004). The issues involved in LD screening for multilingual children have also been raised (Smythe, 2003). It should be possible to avoid many of these problems by developing a screening tool that taps into the underlying cognitive processes that are said to cause LD rather than the mastery of the language itself.

Although research on environmental factors is very important to reading research as a whole they should not have any bearing on LD research into identification if it is accepted that LD is a neurological disorder. Instead these factors need to be recognised so that their confounding influence on LD identification can be minimised. If the research on the genetic nature and employment outcomes of LD is considered concurrently, it is likely that many students identified will be socio-economically disadvantaged. However these environmental factors should not influence the diagnosis of the disorder in the first place.

**Teachers in learning disability screening**

Australian literacy specific teacher-training in general, let alone with a focus on LD, is not currently sufficient given the importance of such matters in a classroom (Freebody, 2007; Munyard, Sullivan, Skues & Cunningham, 2008). Nelson and Machek (2007) surveyed school psychologists in the US and over 90% of those
surveyed stated that they would benefit from more training in reading assessment. This is not surprising given that nearly half of them had never been required to attend a reading specific class in their training. Despite this lack of assistance in training, current student teachers are positive about their abilities to help LD students (Gwernan-Jones & Burden, 2010). However, this finding could be interpreted as indicative of the good intentions of teachers, as compared to the training that they receive.

What is important here is bridging the gap between ensuring that the knowledge and expertise of classroom teachers is not wasted or ignored in LD assessments, while also ensuring that there is not further addition to already heavy teacher workloads. One of the best ways to do this is to ensure that any potential issues in a test are clearly explained to teachers. One example of this is the overlap in reading ability between grade-levels, where there is more difference within each grade-level than between them. This is something that is often underestimated by teachers, who can overestimate the difficulties of a student who is reading at a grade-level one year below their own (Izard, 1998). Additionally, neuropsychological assessments of students are too jargon filled to be of use to teachers (Westwood, 2001). By ensuring that the interpretation does not need to be done by teachers, and that the information imparted is accessible, many of these issues can be avoided.

When it comes to reports stemming from assessments, it is important that the level of knowledge required to understand them is pitched at a suitable level for all parents and not fully trained, experienced, teachers. A full third of students that were put forward for educational psychology assessment in one school were nominated so that the parents could get more information on how to effectively give their children help at home (Long & McPolin, 2009). If a test with appropriate recommendations embedded into the results could be given uniformly in schools, the demand for superfluous testing may be reduced.

Although automatically generating reports with interpreted results would have been considered unattainable in the past, many of the considerations surrounding testing have recently been made moot by advanced technology. By using the possibilities afforded by readily-available software programs, such as Microsoft Excel (2007), it is possible for more complicated statistical decision making processes to be
imbedded into a program so that there is no onus on teachers administering a test battery to fully understand the statistics involved. This idea of using popular programs in this way is not new and has been successfully used before (e.g., Olm-Madden, 2008). In the case of LD, the aim would not be to actually diagnose LD on the basis of an automated program. Rather it would be to highlight those students who are at risk of LD. Furthermore in the case of Australia, where there is no incentive for diagnosis, these same reports can be used to give teachers information on the deficits that their students are facing when trying to master language-related skills. It is hoped that by making assessment accessible then the focus can move on to the area that is most important, namely that of the best remediation of LD students.

Remediation
Assessment of students with disabilities should provide useful information to parents and teachers, with the primary aim of improving student performance (Victorian Curriculum and Assessment Authority, 2005). This is a realistic aim as there are remediation programs that have claimed to be successful with both poor readers and LD students (McMaster, Fuchs & Fuchs, 2007). Others have even shown that LD students with qualitatively different brain function pre-remediation can be remediated and then display brain function similar to their peers (Richards et al., 2008).

Although there is much debate over the best type of intervention for poor or LD readers some have argued that any one intervention based on the possibility of meeting the needs of such a heterogeneous group with one program is based on a false premise (Hay, Elias, Fielding-Barnsley, Homel & Freiberg, 2007). It is important that as much information as possible is available for those people administering the intervention.

Mather and Gregg (2006) made note of the need for norm-referenced tests of cognitive abilities that can identify areas of difficulty, thus highlighting the areas in need of remediation. Remediation programs are starting to focus on the specific cognitive processing difficulties causing LD (Hayward, Das & Janzen, 2007). For instance, a training program that focuses on skills associated with working memory
can result in improved performance in this area in LD children with working memory deficits (Swanson, Kehler & Jerman, 2010).

There are other examples of programs that have successfully targeted cognitive processing deficits. For instance, longitudinal studies have demonstrated the efficacy of well-implemented phonologically-focused remediation (Foorman, Francis, Fletcher, Schatschneider & Mehta, 1998; Hempenstall, 2008; Scanlon, Vellutino, Small, Fanuele & Sweeney, 2005). What is still unclear is whether this success was a reflection of a successful remediation program overcoming LD, or if it was simply those students who were misdiagnosed responding to intervention. Once again, this brings us back to the problems of different operationalisations of LD and how this affects the generalisability of studies.

There does seem to be an assumption that early intervention and remediation is the gold standard in this area of research. Indeed the sooner problems are identified and remediated, the less impact the Matthew effect can have. However, there have also been some promising results with older cohorts. Students in Grade 6 to Year 8 received similar benefits from remediation as students in Grade 3 and 4 (Tressoldi, Lorusso, Brenbati & Donini, 2008). Successful remediation programs also been implemented in the Grades 4 and 5 (Vadasy & Sanders, 2008) as well as Year 8 (Calhoon, 2005).

Ultimately, all of these remediation programs rely on the identification of the students who will benefit from them. As such it is important that screening tools are put in place to ensure that those students who would benefit from assistance are identified before their difficulties have an insurmountable affect on their education. One of the best ways to ensure that more students have access to LD testing without financial or timing issues is by using a screening tool, rather than a diagnostic one. The next section starts with a review of currently available LD screening tests, both international and Australian.

**Global Screening Tests**

When identifying LD the judgement of neuropsychologists has been found to be both reliable and valid (Garb & Schramke, 1996), however this does not help when large scale testing is required. Therefore, even when acknowledging the increased accuracy of individual testing, the pragmatic advantages to group screening tools
make them preferable over individual professional testing, in the right circumstances (C. Reynolds, 1984). Furthermore the objectivity large scale screening provides can be an advantage in some settings. For instance, diagnosis of LD can be affected by race and socio-economic status, even when all other factors, including test scores, are controlled (Frame, Clarizio & Porter, 1984). This is not to say that screening tools are not affected by geographic socio-economic factors, just that they are more consistent and thus easier to control. For instance, the higher proportion of socio-economically disadvantaged children in the United States (US) can make Australian readers appear stronger than they actually are (Hempenstall, 2009). Furthermore, the norms on US reading tests are often grade-based and Australian year levels fall halfway between respective American year levels (Galletly & Knight, 2006). Both of these are real issues that can be controlled with the development of Australian-specific tests and normative data.

Fortunately, there are long standing well developed Australian-specific tests commonly administered in the classroom, such as the South Australian Spelling Test (Westwood, 1979) and the Progressive Achievement Test in Reading (Australian Council for Educational Research, 2004). However, these are tests of reading and spelling achievement, and not screening tools developed with the purpose of identifying LD.

There are a wide range of well developed screening tools of LD, both in Australia and internationally. For instance the Wheldall Assessment of Reading Lists (M. Reynolds, Wheldall & Madelaine, 2009), Queensland Year 2 Diagnostic Net (van Kraayenoord, Luke, Elkins & Land, 1999) and a screening tool designed for middle school children (Speece et al., 2010) have all had some success in identifying LD. However, these tests are all based on reading skills alone and thus may not have as much predictive validity as would be provided by cognitive processing tests that could be more sensitive to future, rather than present, difficulties.

It should be noted that there are currently cognitive profiling screening tools available for young children. For instance, the cognitive profiling system (Singleton, Thomas & Leedale, 1996) is an excellent screening tool that can provide valuable information to classroom teachers about their students. Unfortunately this test
requires interpretation by a trained professional, making it unsuitable for large scale administration.

A group-administered measure of nonword reading was developed by Martin and Pratt (2001), making it possible to test for phonological processing ability in a classroom setting. However, the cost of the test may not be financially realistic for wide-scale implementation. Furthermore, only phonological processing is assessed by this measure and not any of the other processes thought to be inherent to LD.

Over a decade ago the need for freely available word lists that measure both orthographic and phonological processing was recognised and three word lists of regular, irregular and nonwords were developed (Edwards & Hogben, 1999). Surprisingly, considering their potential application in schools they were not widely used. One of the possible reasons for this is that it was quite skewed for most age groups despite psychometric validation being sought through a classical test theory paradigm. Another reason might be that the scale was aimed to an age range of 7 to 12; as such it was not specifically targeted to one population. This means that the scale had to be targeted to a wide range of age groups as well as the wide range of reading proficiencies within those age groups. The importance of appropriately targeting scales to a specific population is discussed in the next section.

Targeting

It is important when psychometrically validating a test that the sample used forms a representative part of the population for which the test is developed for (Gorsuch, 1997; Willson & Reynolds, 1984). Tests need to be targeted to specific populations as even within a grade-level there are age specific variations. Therefore, the chance of students receiving assistance decreases by 2 to 5% for every month older a student is in their classroom (Dhuey & Lipscomb, 2010). Furthermore neuro-imaging studies have demonstrated that there are age-related changes in phonological performance (Simos et al., 2001), so the consistency of the importance of different variables over time cannot be assumed. By developing screening tools specific to each year level confounding factors such as these can be reduced.

Identification of LD students at pre-school age seems to be the ultimate goal of many researchers, but some suggest that this may not be a realistic aim. When looking at conflicting studies of familial risk of LD it has been found that those conducted on
students before the age of five were inconclusive although those conducted past the age of seven were quite successful (Snowling, Gallagher & Frith, 2003). Grade 3 is considered a pivotal role in reading acquisition (Jarmulowicz, Taran & Hay, 2007) and is often the year that bright students with LD start to noticeably fall behind their peers as reading proficiency starts to be assumed (Woolley, 2007). It is also an important year as standardised testing, both individually and in groups, has become more commonplace and, as such, the test provides less of a confounding factor than it would in younger years. For instance, the PAT-R is not recommended for group administration with a Grade 1 sample (ACER, 2008).

Further to the topic of appropriate targeting, not only is it important that a test is age-appropriate but also that it provides the most discrimination between those participants who are of interest, in this case those who perform poorly. Some researchers in the past have developed tests that have been negatively skewed for this very reason (Edwards & Hogben, 1999; Nicolson et al., 2001) but this has made them difficult to psychometrically validate and as such the skew has been slight. Item response theory has been identified as important for developing measures for identifying LD (Hogan & Thomson, 2010) in part because it is more lenient on issues such as skew. Therefore measures can be appropriately targeted to the population of interest.

**Aims of the Current Study**

As discussed in Chapter 2, not only are there theoretical problems with the use of IQ tests in LD diagnosis, but there are also pragmatic ones, in that they are expensive and time-consuming to administer. Despite this, a national inquiry in Australia into teaching literacy made a call for comprehensive diagnostic assessments of every child in school (Coltheart & Prior, 2006). Directly testing for the cognitive processing deficits that are thought to cause LD is not only theoretically sound, it also involves simple to administer tests, making this request a realistic one. Therefore, what is proposed in the current study is that a screening battery of tests be put together to investigate these cognitive processing deficits.

As discussed in Chapter 2, the yardstick by which the success of this method will be assessed is the discrepancy method of LD identification. The cognitive processing deficits that will be used to predict the presence of LD, as per the discrepancy
method of diagnosis, are phonological processing, naming speed and verbal memory. These will be tested via nonword reading, RAN and digit-span tests respectively. Only those students with deficits in these processes will be identified. This will involve a dichotomous deficit/no deficit paradigm, due to the inconsistency of the relationship between reading and these processes as proficiency increases. In order to make the tests as useful as possible, an irregular word list will also be developed in the hope that low scores on this test can provide an alternative explanation for poor reading.

By testing for LD in this manner, a few issues in LD screening can be circumvented, first and foremost those of a financial nature. Many established methods of identification require expensive tests and professional assistance that is not necessarily financially viable for all. Although the fact that these resources are out of financial reach for many schools and families hoping to address LD and may seem unjust, financial inequity is hardly a problem that is specific to this sphere of education, or even disability. However, it is argued here that it is particularly important that LD resources be freely available.

The problem is in the accepted genetic nature of LD, combined with long term outcomes of those with the disorder. This results in what could be described as a generational Matthew Effect. Students with literacy difficulties are more likely to end up living in poverty, in prison and in menial jobs on low pay (Kirsch et al., 1993). Furthermore the genetic basis of LD dictates that those with the disorder, and all its inherent disadvantages, are more likely to pass these problems on to their children than those without LD (Berninger & Holdnack, 2008). Consequently, over time, it follows that children with LD will be increasingly over-represented in a low socio-economic demographic. This puts the lack of affordable screening or diagnostic tools in a different light. It is vital that LD identification can be made freely available so that the students most likely to benefit from them can do so. In order to ensure that this is possible, a number of decisions throughout this study will be made within an applied research paradigm.

Another issue that can be circumvented, this time through the use of technology, is that of the time and knowledge expectation placed on teachers. Ensuring that students, teachers and parents are informed about LD is vital to how LD students
experience school (Singer, 2008), and the onus to ensure this happens is usually placed on the teacher. All three of the tests developed for this study are easy-to-administer and score and could be administered by a teacher or teacher’s aide. Furthermore, the burden of interpretation and generation of reports is removed from teachers, as results can be programmed into a database so that reports for parents and teachers can be automatically generated.

Overall, the aim of the current study is to develop an easy-to-administer screening battery for LD that assesses the cognitive processing deficits of the disorder, rather than secondary symptoms. There are four primary steps that this study will attempt to complete in order to meet this aim.

The first step is to develop psychometrically-sound measures of the cognitive processes thought to cause LD. It is important to develop these independently so that they can be made freely available due to the already discussed socio-economic issues in LD. In the case of the RAN and digit-span tests this can be done confidently as both are well accepted measures. Therefore norms will be developed for the new stimuli, but the tests themselves have already been validated on numerous occasions.

In the case of the word lists, there is much more variation possible in newly generated lists and as such these tests will require more in-depth psychometric scrutiny. The targeting of poor readers in the scale also dictates that the scales should be negatively skewed so that there is more differentiation between poor readers. There is less need to differentiate between good readers, as they are not of interest in this screening battery. As such, item response theory, as compared to classical test theory, will be utilised to validate these scales. The reasons behind this will be discussed in more depth in Chapter 4.

The second step is to establish the links between reading and these cognitive processing tests, and show that those who are low in the processing tests also are, generally speaking, poor readers. It is important to differentiate between measures that compare these processes to poor readers and those that compare them to measures of LD. The majority of LD students are going to be current poor readers. However, this does not mean that the majority of poor readers will have LD.
Therefore, demonstrating that these cognitive processes are correlated with reading ability is an important first step in this area of research, but is only that, a step.

The third step is to then compare these cognitive processing deficits with the presence of LD. As discussed in Chapter 2, the standard to which this will be compared is the presence of an aptitude-achievement discrepancy. What will be explored is if these cognitive processes can be used to predict membership in the groups dictated by the discrepancy method. In order to help generalise these results to those of previous studies, this analysis will be done with both traditional discrepancy methods and regression-adjusted discrepancy methods.

The final step is to identify those students who possibly have LD as per a multiple deficit model. As discussed previously, the measures being developed in the current study are making up a screening tool, not a diagnostic one. Consequently, the identification process will be fairly liberal in order to reduce false negatives within the screening tool. Furthermore as already discussed, the method of identification will be based on absolute deficits and not regression based equations, as it has been shown that the correlation between LD and these processes is weakened as ability increases.

In the process of completing these four steps, four research questions will be asked

1. Can the psychometric validity of the newly-developed irregular word and nonword lists be supported as per the tenets of both classical test theory and item response theory?
2. Can cognitive processing deficits be used to identify a similar cohort of LD students as would be identified via the discrepancy method?
3. Are the identification methods and cognitive processes currently researched in terms of reading-based LD relevant to spelling-based LD?
4. Are many students identified as LA by the discrepancy method of identification actually facing many of the same cognitive processing difficulties that are faced by LD students?

Furthermore the following four hypotheses will also be tested.

1. That deficits in verbal memory will be indicative of LD in the same manner as phonological processing and naming speed.
2. That the more processing deficits will be positively associated with more severe reading deficits.

3. That some of the students not identified as having cognitive processing difficulties will still be poor readers.

4. That orthographic processing deficits will provide an alternative explanation for poor reading if LD is not found.
Chapter 4

Overview

This chapter outlines how the current study was conducted. Initially, the two research paradigms that guided decision making, namely causal-comparative research and applied research, are discussed. Secondly, Classical Test Theory (CTT) and Item Response Theory (IRT), which guided the psychometric validation of the word lists in Chapter 5, are briefly outlined. The psychometric issues involved in screening tools are also reviewed.

The research instruments are also presented along with their respective scoring systems. In order to test phonological and orthographic processing, nonword and irregular word lists, respectively, were constructed. Two Rapid Automatic Naming (RAN) tests were developed to test naming speed and two digit span tests were developed to assess verbal memory. Established tests were used to measure reading comprehension, spelling and non-verbal intelligence. These tests were the Progressive Achievement Test in Reading (PAT-R), the South Australian Spelling Test (SAST) and Raven’s Progressive Matrices (RPM), respectively.

The recruitment and demographics of the participants are outlined. The types of comparisons utilised within groups and the procedure for collecting the data for the results presented in both Chapters 5 and 6 are then discussed. A description of the statistical analyses utilised, including the method for designating students into discrepancy-defined groups and calculating reading age norms, is included.

Research Design

As discussed in Chapters 2 and 3, there are two primary aims in the current study. The first aim is to ascertain if cognitive processing variables can be used to identify students with Learning Disabilities (LD) in the classroom. The second aim is to pilot these tests to see if they can be developed in a manner that make them easily applied in the classroom. The first aim is addressed within a causal-comparative research paradigm and the second aim requires an applied research approach.

Only quantitative data was collected in the current study. This was important, firstly, because of the type of data required for causal comparative research. Secondly, as per the applied research ethos of the study, it is vital that the final method of
identifying students does not require extensive interpretation by the test administrator. As discussed in Chapter 3, new programs installed into a school system need to fit into the schools’ existing programs and training, or they will not be utilised by staff (D. Fuchs et al., 1996). The analyses used to ascertain the appropriateness of these tests to this purpose have been made within the paradigm of causal-comparative research.

**Causal-comparative research**

In an experimental design, participants are allocated into groups randomly so that differences can be analysed. In causal comparative designs these differences already exist naturally in the groups of interest (Schenker & Rumrill, 2004). Consequently the dependent variable, in this case the existence of LD or Low Achievement (LA), or Regular Achievement (RA), does not need to be allocated as it is already in place. Therefore, any research into the differences between the groups is done retrospectively and, because of this, caution must be taken when drawing conclusions from the results.

For over half a century, the need for experimental psychology that investigates causes and treatments, and also correlational designs that investigate individual differences has been widely accepted (Cronbach, 1957). By collecting data on students with and without LD, it is possible to then compare the cognitive profiles of the two groups (Jacobson, Mulick & Foxx, 2005). However, the nature of survey research dictates that causation is still not a given in these circumstances. Nevertheless, hypotheses can still be supported by causal-comparative research, just not definitively.

Another confounding factor that must be taken into account when interpreting results is that there is no widely accepted operationalisation of LD. As such there is not a precise method of definitively calculating accurate identification of LD. Therefore, various methods of identification were utilised in the current study so that comparisons and generalisations between the various methods can be examined.

Although the research is conducted within the guidelines set by causal comparative research, the primary aim of this study is to pilot a screening battery of tests that will be suitable for use in the classroom. As such, it was vital that all the tests and
decisions made throughout the study were compatible with an applied research paradigm as well.

**Applied research**

As discussed in Chapter 3 there has been a long-standing clash between theoretical and applied research in the field of LD. This has resulted in issues with the application of scientifically-supported theories into the classroom, thus limiting their usefulness. One of the primary issues for universal screening tools highlighted by Glover and Albers (2007) is usability. Consequently, any screening tool developed must be easily administered and scored if eventual implementation is the aim.

Six issues under the umbrella of usability have been highlighted by Glover and Albers (2007). The first is that the cost of administration of the test should not be prohibitive and, second, that it should be administrable by a wide range of people. Third, the instrument should be considered acceptable by all relevant groups, for instance schools, teachers, parents and the schools’ governing bodies. Fourth, the required resources for the test, both materials and staff, must be accessible. Fifth, there needs to be appropriate accommodations stemming from the screening tool and finally, the information coming from the tool needs to be of use to all relevant parties.

To address the first point, cost, all the cognitive processing tests were newly developed with preliminary norms drawn up as part of the current study. Consequently, these tests can be made available free of charge to any school wishing to utilise them. The most important step for reducing the cost of LD testing is to remove the established intelligence tests as these represent a large proportion of the cost involved in current diagnosis. As discussed in Chapter 2, this can be done without adversely impacting the predictive validity of the screening test, as the place of intelligence testing in LD identification has long been controversial and much maligned (Vellutino et al., 2004). Instead, by concentrating on the cognitive processes that are thought to cause LD, it should be possible to develop easy-to-administer tests that can not only screen for LD, but also give teachers information on the all students in their class.

Furthermore, ensuring that the tests can be administered by a teacher, or even teacher’s aide, is an important part of the study. This is because, if these tests do not
need to be administered by a psychologist, the cost of testing will be greatly reduced. This contributes not only to the cost-related first requirement, but also meets the second criterion, which is that administration can be carried out by a broad range of people.

The third issue, acceptance by relevant groups, cannot really be assessed until the screening tool is released. However, the screening tools will be free of charge and will provide information on why a student may be struggling to read. This should result in positive reception of the tests by those who deal with them. The fact that the tests are simple and free, with the exception of the PAT-R that is administered and already paid for in many Australian schools, will result in meeting the fourth requirement, accessibility of resources.

Although the fifth issue, appropriate accommodations, will not be addressed in the current study, specific reports of each students strengths and weaknesses will inform remediation. For instance, students with phonological deficits respond better to phonologically-based remediation than poor readers who do not respond to this form of remediation (Hempenstall, 2008). These individual reports should also ensure that the sixth and final requirement, dissemination of useful information, is met.

Because the study aligns itself with an applied research paradigm, all decisions made throughout the study were made with these six points in mind, and particularly the first two, namely the cost of administration and increasing the range of people who can administer the tests. If this test is to be implemented in schools as a screening tool, it is important that all aspects of the tests’ administration, scoring and interpretation are suited to a classroom setting and not just a research one.

**Test Theory**

Although there is a wide range of research paradigms under which one can conduct research, there are fewer options available when it comes to psychometric theories. Whilst the RAN and digit span tests have been very well researched and validated in various forms, the irregular and nonword lists developed in the current study will need psychometric validation. Two approaches, Classical Test Theory (CTT) and Item Response Theory (IRT) were used for this. These theories will be discussed in the next section.
Classical test theory

The basis of Classical Test Theory (CTT) is that all test scores are the sum of a participant’s true score for the test plus error (Novick, 1966). By attempting to ascertain the level of error in a test, one can get an idea of the range in which a true score may lie on that test (Novick). One of the major issues inherent to CTT is that the true score (i.e., the actual score that a person would receive on a test if it were possible to control all error) is the true score on the test itself, rather than the true score for the tested construct (van der Linden & Hambleton, 1997). Furthermore, all variation between test takers is presumed to be systematic, as error variance is assumed to be either constant or randomly distributed throughout the participants (van der Linden, 1986).

One of the primary issues with the application of CTT with a screening tool is that the assumption of normality can be easily violated in a test designed to only separate those at one end of a spectrum. Ceiling effects can occur when a test, designed to pick up poor performance, is given to a normal sample (Alexander & Martin, 2000). However, it could also be argued that any screening test that is not skewed towards the target population, especially if a small minority is targeted, is not effectively doing its job.

The dangers of using tests with floor or ceiling effects at the same end of the spectrum where those who are of interest lie, have been investigated. An example of this would include many achievement tests in which the poorest performing students are not able to understand the easiest questions. The conclusion was that, when this occurs, lower-level testing is required in order to differentiate those students at the lower end of the spectrum (Catts, Petscher, Schatschneider, Sittner Bridges & Mendoza, 2009). Although two levels of tests do not fit in with the applied ethos of the current study, it is already clear that the population of interest is poor readers. Consequently, the tests developed were firmly targeted to that cohort. Therefore, one of the aims of the current study is to develop the irregular and nonword word lists with negative skew so that more accurate information can be gleaned on the poorer-performing students.

Screening tools that aim to detect specific issues in the wider community are often strongly skewed because those with normal or high functioning are not targeted by
the scale. For example the Psychological Consequences Questionnaire (Rijnsburger, Essink-Bot, van As, Cockburn & de Koning, 2006), the Self Harm Inventory (Latimer, Covic, Cumming & Tennant, 2009), The Depression Anxiety Stress Scales (Lovibond & Lovibond, 1995; Shea, Tennant & Pallant, 2009) and the Kessler measure of psychological distress (Kessler et al., 2002) are all skewed so that maximum information can be extracted from the scale on the people of interest. All these tests are popular and widely used, despite the assumption of normality in CTT that assumes that data will be normally distributed without skew. In order to avoid the issues inherent in having a test skewed by design within the paradigm of CTT, it is possible to check the psychometric properties of a test through IRT, which is much more robust to violations of normality.

**Item response theory**

Unlike CTT, Item Response Theory (IRT) has a focus on the interaction between each participant and each item, instead of on each participant and the entire scale. One of the primary advantages of IRT over CTT is flexibility. Although some of the assumptions of normality are strict, for instance unidimensionality, others, such as skew, are not as important in IRT (Edelen, Jaycox, McCaffrey & Marshall, 2006).

One analysis within the IRT paradigm is Rasch modelling. A Rasch model “assumes that the probability of a given respondent affirming an item is a logistic function of the relative distance between the item location and the respondent location on a linear scale” (Pallant & Tennant, 2007, p.3). In other words, the model is based on the relationship between the difficulty of each item and each participant’s level of proficiency on the measure. Both items and respondents are assessed on the same scale. Difficult items have logit scores akin to high-performing respondents and easy items have similar logit scores as poor-performing respondents. Therefore, if a respondent and an item have the same logit score, there is a 50% chance the respondent will get that item correct.

IRT has been successfully applied to screening tools in other areas (Fink et al., 2004; Shea et al., 2009) and the theory behind using CTT in conjunction with IRT has been addressed both theoretically and mathematically (Bechger, Maris, Verstralen & Beguin, 2003). Although IRT is thought to be superior to CTT in some areas of scale development, most advocates strongly suggest that IRT be used in conjunction
with CTT rather than in its place (e.g., Edelen & Reeve, 2007; Erhart et al., 2009). It has been noted that although the end result of a scale refined by IRT or by CTT will often be very similar, if not the same, the value lies in how these decisions were made, as well as in the relatively rare situations where differences may occur (Edelen & Reeve).

This section has reviewed the two test theories, CTT and IRT, which were used in the current study. The decision to use both of these theories in the validation of these tests was made, in part, in order to ensure that the irregular and nonword tests used within the screening battery were as valid and reliable as possible. There are also psychometric issues that are specific to the use of screening tools, these are discussed in the next section.

**Screening Tools**

It is important to acknowledge the differences between a diagnostic and screening tool as not only are the outcomes of the use of these batteries obviously different, but the requirements of the two are also very different. Although a definitive diagnosis cannot come from a screening tool, they are developed in order to identify those people that are at-risk for the targeted disorder. This is done in the hope that the disorder can be identified before it develops beyond the point of easy remediation (Morrison, 1992).

Although early identification with screening tools can have a high misclassification rate, more accurate tests are often too time-consuming and costly to be widely implemented (Compton et al., 2010). Fortunately, many of these issues can be avoided in the case of a screening tool, as it is possible to allow some leeway on Type II error in order to ensure that the Type I error is low. This is because the aim of a screening tool is to identify possible, not definite, issues in the targeted area. It is important however that this room for error is noted and accounted for, as false positives in screening tools have not always been appropriately acknowledged (Gredler, 1997).

Diagnosis on the basis of any form of testing should be done with extreme caution. Pena, Spaulding and Plante (2006) calculated that in order to statistically confident of classifying 90% of people in a hypothetical set of parameters, the test needs a reliability coefficient of .98 or more. However, reliability coefficients of that
magnitude are extremely rare. The authors of this study concluded that not including impaired children in the normative sample may help increase the reliability of a diagnostic test.

Even with the selection of reliable, fixed tests and the safety net of higher acceptable Type II error, screening tools can be inappropriate if the method of identifying issues is inappropriate to the problem being identified. For instance, some LD screening tools use actuarial methods (Benson & Newman, 2010) or equations developed through a Discriminant Function Analysis (DFA; Leonard et al., 2001; Merrell & Plante, 1997). However, it could be argued that tests such as these fail to take into account the finding that the importance of cognitive processes increases as reading ability decreases. For instance, while decreasing levels of phonological processing or naming speed will result in decreasing reading ability, the converse does not necessarily occur as phonological processing or naming speed levels increase (Kirby et al., 2010; Schatschneider et al., 2002). This could be interpreted as reflecting that only a certain level of proficiency in any given process is required for reading proficiency, and that increases in abilities above and beyond this level become irrelevant to reading as proficiency in the process rises. Consequently, in the current study, all identification were done on the basis of the existence of deficits in one or more cognitive processes, not on a cumulative score based on the relative proficiency of any given student on a range of tests.

One method of increasing reliability is to use a fixed, as compared to flexible, test battery (Russell, Russell & Hill, 2005). A fixed test battery, such as the one developed for the current study, uses the same tests to the same end for every student tested. A more flexible test battery is often used in a clinical setting because individual psychologists base their choice of test on client, settings, expected outcomes and availability (Olm-Madden, 2008). Bigler (2007) has argued that the benefit of this flexibility and autonomy is thought to outweigh any advantages of a fixed test battery. However, in the case of a screening tool that is not going to be interpreted by those administering it, a fixed test battery is required in order to ensure that appropriate conclusions are drawn from the tests.

With the method of identification and type of screening tool selected, and the issues inherent to this decision acknowledged and discussed, the next step is to select the
tests that were used within the battery. In the next section the research instruments used in the current study, both established and newly developed, are reviewed.

**Research Instruments**

Students were administered a combination of both established and new tests, with students in different grades receiving a different combination of the tests below. As the target sample in the current study were Grade 3 students, all tests were primarily developed or chosen for this sample, but as will be discussed, some accommodations needed to be made for the Grade 1 and 2 samples.

**Tests of aptitude and achievement**

Because of the focus on usability in the current study, group-administered tests were selected wherever verbal responses were not required. This makes administration in class time, by a teacher, more realistic and will have less of an impact on planned classes. Furthermore, by selecting tests that may have already been administered it will be possible to reduce the amount of time and effort that needs to go into the screening process. The three tests used in the current study to measure aptitude and achievement are all well established and group administered. These tests are the Progressive Achievement Test in Reading, the South Australian Spelling Test, and Raven’s Progressive Matrices, which assesses non-verbal intelligence.

**Progressive achievement test in reading**

The Progressive Achievement Test in Reading (PAT-R), 4th Edition (ACER, 2008), is a widely used and accepted Australian test of reading comprehension. The test is divided up into sections, such that each section consists of a passage to read and then a set of questions to answer about that passage. Students are given 40 minutes to complete the test as well as approximately 15 minutes administration time.

The group-administered test has a multiple-choice format and each level of the test can cater to three consecutive academic years at school. Students in the current study sat either the Grade 2 or Grade 3 PAT-R, as each test is available to three different grade levels. The choice between the Grade 2 and Grade 3 test was primarily dependent on whether the test had already been administered in the school. Each grade-level has its own norms for each test, and each possible score for each grade-level has a corresponding percentile score. These scores are comparable across grades (Australian Council for Educational Research, 2004). Therefore a
Grade 3 student who got a score in the 75th percentile on the Grade 3 PAT-R is equivalent to a Grade 3 student who got a score in the 75th percentile on the Grade 2 PAT-R, as long as both are compared to their grade appropriate norms within each test.

The PAT-R is widely recognised as psychometrically sound. The Grade 2 and Grade 3 tests had a Cronbach’s alpha coefficient of .90 and .88 respectively (Australian Council for Educational Research, 2004). Furthermore, validity has been supported by correlations between the PAT-R and other tests of reading comprehension of .65 to .86 (Australian Council for Educational Research). Questions in both of the tests address both implicit and explicit information in the reading passages. The Grade 2 PAT-R is similar to the Grade 3 PAT-R, however there are small differences. In the Grade 2 test there are five passages of text with four or five questions allocated to each passage. In the Grade 3 PAT-R test there are six passages with four or six questions allocated to each passage.

Although the PAT-R has tests catered towards Grade 1 to Year 10 students, it could not be administered to the Grade 1 sample in the current study as both the answer sheet format and the content of the questions were deemed inappropriate for such a young cohort by both the researcher and the Grade 1 teachers involved. This was not wholly unanticipated as the Grade 1 test manual recommends one-to-one or one to small group testing with guidance for this age group. This was not appropriate for the current study, as group-administered tests of reading comprehension were required. Failing that, a test that had been already administered, in order to not unduly add to teacher workloads was considered preferable. Therefore, a new measure of reading ability and comprehension was required for Grade 1 students. This measure is outlined in the next section.

**Test of reading comprehension for Grade 1 students**

As part of their normal classroom routine, Grade 1 students were tested individually on a monthly basis on a reading task from their homework readers. This is a common practice in Australian primary schools. All students are given a reading level that dictates the level of reader that they take home every night for that month. Monthly tests are used to ascertain whether or not a student should be upgraded to the next reading level. In these tests teachers literally mark each word that the
students read out loud so that reading accuracy is represented as a percentage. Reading comprehension is tested with a few questions about the passage of work the student has just read to the teacher.

These reading levels, taken at the time of individual testing, were used in place of the PAT-R for Grade 1 students. There are substantial differences between the two methods of reading ability testing, namely the verbal and written mediums in which the students are tested. However, there are inherent difficulties in testing the reading ability of such a young cohort that are difficult to avoid. Furthermore, the regular monthly testing, as well as the objective method with which the teachers gauged accuracy, indicate that the reading level should be a fairly reliable indicator of any given students ‘true score’ in reading.

It should also be noted that the primary goal behind giving a test of reading comprehension to Grade 1 and 2 students was so that a reading age comparison could be developed for the Grade 3 students. Therefore, if a Grade 3 student is given a Grade 1 reading age, based on the norms given by the PAT-R, then the reading levels will simply be used to ascertain which Grade 1 students are reading at an age-appropriate level and can thus be used as a basis for comparison. Consequently, precise reading scores were not required. Instead, a measure that gave us an approximate of the inter-quartile range of reading ability for that Grade was all that was required. This measure could provide this range, so it was used in place of the PAT-R for Grade 1 students.

South Australian spelling test

The South Australian Spelling Test (SAST; Westwood, 1979) consists of 70 spelling words with verbal presentation and written response. It was first developed in 1979 and the norms have been updated twice since then with the most recent update prepared in 2004 (Westwood, 2005). The test can be group-administered to a whole class and is suitable for students aged between 6 and 16 years of age. In many cases in the current study, the test had already been administered by the teacher at the beginning of the year, as it is used in many schools as a gauge of progress in spelling. The test has been widely used in Australian schools for decades and is psychometrically sound with an extremely high .96 test-retest reliability (Westwood, 2005).
The test is administered verbally, with the administrator reading each word out twice, as well as putting each one in a sentence. For instance: “Number sixteen is ship. A ship is on the sea. Write ship” (Westwood, 2005, p.77). Students write their answers on a numbered piece of paper. Although the test has a stop rule of ten incorrect answers in a row, this is difficult to implement in a group setting. Therefore, the entire test was read out, even to a relatively young Grade 3 cohort.

Raw scores from the SAST are usually converted into a spelling age of years and months. This can then be compared to a student’s chronological age and this comparison is then used as the basis of descriptive terms such as Normal, High and Very Low to describe a given student’s spelling ability. However, as the current study is based on comparisons of grade-level and reading age, the students spelling ages were calculated and this age was used as the final score instead. This resulted in an easier to interpret score than those based on discrepancies between spelling and chronological ages. This had the added advantage of removing any confounds inherent to having age-based norms as compared to the grade-based norms in the PAT-R and cognitive processing tests. A copy of the SAST can be found in Appendix A.

Raven’s progressive matrices

In order to assess aptitude as per the aptitude/achievement discrepancy paradigm, Raven’s Progressive Matrices (RPM) was selected. The RPM consists of 60 multiple-choice items in which students are asked to select the missing piece of a larger pattern. These 60 questions are divided into five sets of increasing difficulty. The RPM examines non-verbal reasoning in a manner that does not need any reading or language comprehension, thus providing an excellent juxtaposition to the spelling and reading tests for a classic aptitude/achievement discrepancy. The test has a multiple-choice format and can be group-administered to an entire class. The administrator explains the concept of the test by going through the first two questions with the group. Once the administrator is sure that students understand the requirements of the test, they are asked to work silently on the multiple-choice test.

Norms have been developed for an Australian population and are sorted by age groups split by six months. Therefore, students in the same class will have norms taken from up to three different age groups, although these differences are not large.
As per the administration manual, raw scores are transformed into norm scores based on these age groups. RPM norms are given in percentile groups of 5, 10, 25, 50, 75, 90 and 95th percentiles. This presented an issue, as most of the tests in the current study were normed on the basis of grade-level and not age. Furthermore, these percentile groups provided little discrimination for the majority of students, namely those between the 25th and 75th percentile.

Given the lack of separation in these norms, it was decided to utilise raw scores on the test. As the test was only administered to a single grade-level in one city, the comparisons made within groups would be based on a relatively homogenous cohort. Similar to the SAST, this approach had the added advantage of removing any confounds inherent in using age-based norms as compared to the grade-based norms in the PAT-R and cognitive processing tests.

Cognitive processing tests
Three tests were developed specifically for this study. These tests have three striking similarities. Firstly, they are all based on commonly-used tests in this area of research. Secondly, they are designed to tap into very basic cognitive processes, therefore they are kept simple, so that the focus of the test is direct. Finally, all three tests involve verbal responses from the student. This was a deliberate decision, made in order to avoid confounding variables such as fine motor control, writing and spelling abilities in students that could affect scores. The negative aspect of this decision is that the verbal responses lend themselves to individual administration. Therefore, efforts were made to ensure that the tests were kept short. Scoring and administration of all the cognitive processing tests was completed by the researcher.

Because these tests were developed by the researcher, scoring systems also needed to be developed. Although the contents of the tests are new, the ideas and conventions on which they are based are not, so traditional scoring methods were utilised. This was especially important for working within the applied research paradigm of this study. All tests needed to be easily scored so that the methods could be simply translated for classroom use.

The first two tests developed were word lists. The first, an irregular word list, is a measure of orthographic processing and the second, a nonword list, is a measure of phonological processing. Two RAN tests were developed in order to test naming
speed and two digit span tests were developed in order to measure verbal memory capacity. Each of these is discussed separately in the following sections.

**Irregular word reading list**

In order to measure orthographic processing, irregular word reading lists were developed and as part of the current study, were psychometrically refined. Irregular words are words that are not pronounced the way they are spelt. In other words, there is no strict spelling-to-sound correspondence. Examples of irregular words include *yacht* and *eye*. This list consisted of words that were frequently used by children in early primary years. This was done primarily to ensure that all the words on the list were words that Grade 3 students could have every opportunity to recognise. This removed much of the variability that could arise from any given participants exposure to any of the words on the list.

In order to ensure that the word list was appropriately targeted, a preliminary list was drawn up and administered to a sample of twelve children from the first school. From this it was gathered that the original test was too easy, so it was amended before administration to the entire sample.

The number of letters, syllables, phonemes and consonant clusters in each word were all monitored to ensure a wide range of combinations of these factors. Consonant clusters are the largest number of consecutive consonants in a word that are not broken up by a vowel. Similar to the other three factors, the number of consonant clusters is positively correlated with difficulty in reading a word (Martens & de Jong, 2006; Rispens & Parigger, 2010; Ziegler & Goswami, 2005; Ziegler et al., 2000). Efforts were made to ensure a wide range of words were used for the list, and a table demonstrating the properties of the words can be seen in Appendix B.

A wide range of words were originally short listed for the full list, this was then reduced to 37 words that covered a wide range of the previously mentioned factors. However, it is hoped that the list can be significantly shortened through the use of Rasch analysis. An actual size sample page of the irregular word list can be seen in Figure 1. Students were asked to read all the words presented to them out loud. They were allowed to pass words, however, if they started passing whole pages they were asked to check each word to make sure that there were not some words on that page that they could read.
done
blue
juice
cough
laugh
towel

Figure 1. Sample Page from the Irregular Word List
Because students often got the stem of the word right, but the prefix or plural wrong, a three level scoring system was developed with zero points for an incorrect response, two points for a correct response and one point for a partially correct response. For instance, a student who said *exhausting* instead of *exhausted*, or *handstand* instead of *handstands*, would be awarded one point. A list of acceptable one-point answers can be seen in Table 1. Please note that although there may be other variations that would have received a one-point response, these were the only acceptable responses received in the current study.

**Nonword reading list**

In order to measure phonological processing a nonword list was developed and refined as part of the current study. Nonwords are made-up words that are pronounceable but have no meaning. For instance, *blurk* is a nonword, *thdfaxc* is not. Some of these words had more than one correct way of pronouncing them based on letter sounds and precedent in commonly-used English words. For instance *kour* could be pronounced to rhyme with *four* or *hour*.

Table 1

*List of One-point Responses for the Full Irregular Word List*

<table>
<thead>
<tr>
<th>Word List Item</th>
<th>One-point Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>higher</td>
</tr>
<tr>
<td>know</td>
<td>known</td>
</tr>
<tr>
<td>juice</td>
<td>juicy</td>
</tr>
<tr>
<td>frightened</td>
<td>frighten</td>
</tr>
<tr>
<td></td>
<td>frightening</td>
</tr>
<tr>
<td>excited</td>
<td>excite</td>
</tr>
<tr>
<td></td>
<td>exciting</td>
</tr>
<tr>
<td>characters</td>
<td>character</td>
</tr>
<tr>
<td>jewellery</td>
<td>jeweller</td>
</tr>
<tr>
<td>somersault</td>
<td>somersaults</td>
</tr>
<tr>
<td>rhinoceros</td>
<td>rhinoceroses</td>
</tr>
<tr>
<td>certificates</td>
<td>certificate</td>
</tr>
<tr>
<td>exhausted</td>
<td>exhausting</td>
</tr>
<tr>
<td>cooperation</td>
<td>cooperating</td>
</tr>
</tbody>
</table>
Because there is little information available on the difficulty level required for any
given grade-level for made-up words, a preliminary list was drawn up and
administered to a sample of 12 children from the first school. As the test was too
simple it was amended before administration to the full sample.

The number of syllables, letters, phonemes and consonant clusters were taken into
account when developing the nonword list. Similar to irregular word reading, all of
these factors are positively correlated with the difficulty in reading a word (Pammer,
Lavis, Hansen & Cornelissen, 2004; Ziegler & Goswami, 2005; Ziegler et al., 2000).
All letters were used at least once in the test. Furthermore, all vowels and phonemes
were also used with the exception of /ʊə/, the vowel sound in tour, which is
gradually being phased out of Australian English (Cox & Palethorpe, 2007). The
representation of all these factors can be seen in Appendix C. Once again, it should
be noted at this point that although these factors were a major part in the selection of
the full nonword list, the refined word list was selected on psychometric grounds,
now that a theoretically diverse full list was selected. A sample page of the nonword
list, as it was given to participants, can be seen in Figure 2.

Students were asked to attempt to read all the words out loud. They were allowed to
pass words, however if they started passing whole pages they were asked to check
each word on that page to make sure that there were not some words on each page
that they could read. Because there was a one-point option that was put in place for
the irregular word list for incorrect suffixes, a similar system was put in place for the
nonword list. There were some small errors such as technically correct but
implausible pronunciations (e.g., pronouncing the hard g in reng) that were given
one point. Table 2 displays the words with a one-point option and the
pronunciations that would receive a one-point score. The effectiveness of this
scoring system will also be assessed with a Rasch analysis.

**Rapid automatic naming**

Denckla tested the ability to rapidly name colours in a Rapid Automatic Naming
(RAN) test in a group of LD boys (1972b). Students with unexpected reading failure
had a “lack of automaticity” that resulted in these students taking longer to name a
Figure 2. Sample Page from the Nonword List
Table 2  
*Correct and Partially Correct Responses to Nonwords in the Full Nonword List*

<table>
<thead>
<tr>
<th>Item</th>
<th>Pronunciation</th>
<th>Approximation</th>
<th>Two Points</th>
<th>Pronunciation</th>
<th>Approximation</th>
<th>One Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>ra</td>
<td>ra</td>
<td>ra</td>
<td>reI</td>
<td>ray</td>
<td></td>
<td></td>
</tr>
<tr>
<td>zay</td>
<td>za</td>
<td>zay</td>
<td>zə</td>
<td>za</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aip</td>
<td>aip</td>
<td>æp</td>
<td>ap</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>oot</td>
<td>oot (boot)</td>
<td>oot</td>
<td>ot</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>oji</td>
<td>oji</td>
<td>odi</td>
<td>oj-eye</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reng</td>
<td>reŋ</td>
<td>reŋ</td>
<td>renG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>crex</td>
<td>crex</td>
<td>krek</td>
<td>creck</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kory</td>
<td>kɔɾi</td>
<td>kɔɾi</td>
<td>korry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rilch</td>
<td>rɪltʃ</td>
<td>rɪltʃ</td>
<td>rilich</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>quelp</td>
<td>kwɛlp</td>
<td>kwɛlp</td>
<td>quilp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>olloy</td>
<td>olloy</td>
<td>ollo</td>
<td>olly</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>joota</td>
<td>dʒʊta</td>
<td>dʒʊta</td>
<td>jotta</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bupode</td>
<td>bəpəʊd</td>
<td>bəpəʊd</td>
<td>bupodeh</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>geasure</td>
<td>geʒ3</td>
<td>geesure</td>
<td>geeser</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gíʒ3</td>
<td>gis3</td>
<td>gis3</td>
<td>gisa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dʒeʒ3</td>
<td>dʒes3</td>
<td>dʒes3</td>
<td>jeeser</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dʒiʒ3</td>
<td>dʒis3</td>
<td>dʒis3</td>
<td>jisa</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

given list of colours. Consequently, the term ‘newly developed’ is used cautiously in the case of the RAN tests developed for the current study. The concept behind the test is well established; however new stimuli and tentative norms for a specific population were developed as part of this study.

In these RAN tests, students were given a laminated sheet of paper with five colours repeated in random order over 48 (6x8) blocks of colour. There were ten occurrences of three of the colours and nine occurrences of the other two. Students were asked to read the colours out loud, left to right, line by line, as fast as they
could. This was recorded so that it could be timed at a later date. The same thing was then repeated for letters, B, O, M, E and R that were once again repeated in random order for 48 letters. A copy of both of the cards that students were handed to read out can be seen in Figures 3 and 4. There has been some research to suggest that although non-symbol RAN tests such as colour naming are suitable for younger students that by the time students reach Grade 3, alphanumeric based RAN tests are better indicators of difficulties (van den Bos, Zijlstra & Iutje Spelberg, 2002). However, both tests were used in the analyses until this has been ascertained for this cohort. Scores in the RAN tests were obtained by measuring the time taken to read the colours or letters in whole seconds. Consequently, high performing students will receive lower scores.

**Digit span**

Digit span tests, both forward and backward, are often used as a measure of verbal memory (Helland & Asbjørnsen, 2004). As discussed in Chapter 3, the digit span forward is a widely-accepted measure of short term memory, but the role of backwards digit span is unclear with some studies utilising it as a measure of short term memory (Richardson, 2007) and some as a measure of working memory (Savage & Frederickson, 2005). What is more widely accepted is that both tests tap into a verbal memory construct that has relevance to studies on LD (Service, 2006).

For the current study, new numbers were generated and new norms developed, purely so that the norms could be made freely available. Numbers were randomly generated using Microsoft Excel and repeated digits were deleted from the list. The numbers were then placed into sets, three of each digit length. There were sets of two to nine digits for the forward span, and two to eight digits for the backwards span. These were then printed on scoring sheets for easy administration.

These sheets were filled out by the test administrator as the test consisted of verbal questions and responses. For instance, in the forward digit span a student may be asked to repeat the numbers 6, 8, 3. If they repeat them, in the same order, they are marked as getting that item correct. If a student got all three attempts at a given digit string length incorrect, then the testing was stopped. This process was then repeated for the backward digit span. In that test if a student was asked to repeat the numbers 9, 4, 2 the correct answer is 2, 4, 9. The longest digit string that the student was able
Figure 3. Copy of the Colour Rapid Automatic Naming Test
Figure 4. Copy of the Letter Rapid Automatic Naming Test
to recall for the forward and backward tests respectively was noted as their score. The digit span test has established good test-retest reliability (Flanagan & Kaufman, 2004; Gray, 2003; Karpicke & Pisoni, 2000), as such this will not be tested in the current study.

**Test administration over grade levels**
The Grade 3 cohort was administered all tests except for the Grade 1 specific reading measure. The primary motivation for testing Grade 1 and 2 students was to generate a reading age-matched cohort for the nonword reading for the poor readers in Grade 3. As such, there was no reason to administer either the SAST or the RPM to these students. However, the irregular word lists, RAN tests and digit span tests were given to Grade 1 and 2 students so that informative reading reports could be developed for all those who participated. The tests administered to students in Grades 1 to 3 are shown in Table 3.

In this section the tests that were administered in the current study have been reviewed. In the next section the sample that these tests were administered to will be outlined.

| Table 3 |
|-----------------|----------|----------|----------|
| **Tally of Tests Administered to Different Grade Levels** |
| Test              | Grade 3 | Grade 2 | Grade 1 |
| Word Lists        | ✓       | ✓       | ✓        |
| RAN               | ✓       | ✓       | ✓        |
| Digit Span        | ✓       | ✓       | ✓        |
| RPM               | ✓       |         |          |
| PAT-R             | ✓       | ✓       |          |
| Reading Levels    | ✓       |         | ✓        |
| SAST              | ✓       |         |          |

*Note. RAN = Rapid Automatic Naming; RPM = Raven’s Progressive Matrices; PAT-R = Progressive Achievement Test in Reading; SAST = South Australian Spelling Test.*
**Sampling Framework**

**Sampling method**

At both the school and student level, the sample was essentially one of convenience. Seventeen Catholic and State schools within Victoria were sent a letter inviting them to participate in the study and five agreed. At the student level, all Grade 3 students at each of the participating schools were invited to participate in the study and those that consented were selected for participation. The school from which the Grade 1 and 2 samples were sourced was also a matter of convenience. This school agreed to allow all three grade levels to be tested.

**Schools**

The sample was essentially one of convenience such that a wide range of schools was asked to participate in the study and, of these schools, five schools agreed to participate in the study. Four of these schools were State Schools in the outer eastern suburbs of Melbourne and one was a Catholic Primary School in the south-eastern suburbs. In each school, all Grade 3 students were invited to participate and at one of the schools all Grade 1 and 2 students were also asked to participate in the study.

**Students**

Two hundred and ninety students from Grades 1, 2 and 3 participated in the study. Three of these students were not actually suited to the sample on the basis of disability or having English as a second language, however in order to ensure they did not feel excluded they were still asked to participate in the testing. Within the remaining 287 students there were 172 Grade 3 students ($M = 8.83$ years, $SD = .42$ years) from all five participating schools. The number of Grade 3 participants, mean age and response rates from each school, in chronological order of testing, is shown in Table 4.

Students from School 4 and 5 were also asked to be tested again a month after the first round of testing, in order to establish test-retest reliability of the wordlists and Rapid Automatic Naming (RAN) tests. Of these 35 students, 86% (i.e., 30) of those asked agreed to participate in the second round of testing. Furthermore, 58 Grade 1 students ($M = 7.05$ years, $SD = 0.32$ years) and 57 Grade 2 students ($M = 8.09$ years, $SD = 0.32$ years) from School 3 also participated in the study. The response rates for the Grade 1 and 2 classes was 63% percent in both grade levels. Although the current study is focussed on providing a screening tool for Grade 3 students, small
Table 4

*Number of Grade 3 Participants, Mean Ages and Response Rates per School*

<table>
<thead>
<tr>
<th>School</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Response Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>42</td>
<td>8.63</td>
<td>0.30</td>
<td>73.68</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>8.97</td>
<td>0.33</td>
<td>58.14</td>
</tr>
<tr>
<td>3</td>
<td>70</td>
<td>8.66</td>
<td>0.33</td>
<td>72.61</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>8.92</td>
<td>0.32</td>
<td>28.57</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>9.48</td>
<td>0.29</td>
<td>31.84</td>
</tr>
</tbody>
</table>

N = 172

Grade 1 and 2 cohorts are required in order to provide reading age comparisons for phonological processing to the Grade 3 students.

*Procedure*

**Ethics**

Ethical clearance was initially obtained from Swinburne University Human Research Ethics Committee. Clearance was then gained from the Victorian Department of Education and Early Childhood Development and the Catholic Education Office in the Archdiocese of Melbourne, in order to conduct the study in both State and Catholic schools. Ethical approval forms from all three bodies can be found in Appendix D.

The principals of the participating schools all consented to having the study conducted at their school and it was at this point that students were sent home with information sheets and consent forms. Only those students whose parents had signed the form and agreed to their child’s participation took part in the study. A copy of the consent form and information sheet is in Appendix E. Once a school had agreed to participate in a study and consent forms were returned, the researcher consulted with the relevant classroom teachers in order to determine when both the group and individual tests could be administered.

**Administration**

The first test to be administered at the Grade 3 level was the RPM. This was decided for three reasons; firstly it gave the researcher a chance to introduce herself to all the students with the teacher present; secondly, this was the only group-administered test
that did not rely on reading and, as such, was considered the most ‘fun’. Consequently, this put the students at ease about the testing. Finally, and perhaps most importantly, it gave the researcher a chance to reassure students about the difficulty of the testing. The researcher was able to tell the students that, as the RPM tested both children and adults, it would become more difficult as they progressed through the test. Accordingly, students were told that there were no expectations for them to obtain a high number of correct answers towards the end of the test. This set up a pattern for the individual tests whereby the researcher could reassure the student that, just as in the RPM, there were some very difficult questions towards the end of each test. Consequently, they were less likely to feel any undue pressure. This appeared to be a very effective tool in alleviating any concerns that students may have had prior to individual testing.

From this point onwards, the order of administration of tests varied, depending on a number of factors. Firstly, some schools had already administered the PAT-R or the SAST before the researcher started testing and this obviously impacted on the order of administration. Secondly, the timing of the group testing was much more important to the classroom teachers than the timing of the individual testing. Therefore, in a given classroom there may have been some students who were administered the individual tests before the PAT-R and/or the SAST, and vice-versa for the others.

The administration order for the individual tests remained consistent throughout testing. The working memory test was administered first, the RAN test second and the word lists last. This order was selected because the digit span tests had a stop rule that ensured students would not be frustrated for future tests. The word lists were the obvious choice to put last, as some students who were poor readers could struggle on this test. The RAN test was the most entertaining individual test for students and was extremely short. As such it was put between the two longer tests in order to break up the testing.

Although the exact timing of administration was not recorded, anecdotally, individual testing took anywhere from 5 to 25 minutes depending on the student, with the majority of students taking between 10 and 15 minutes. The administration
of these tests and the instructions for these cognitive processing tests were uniform and are outlined in Appendix F.

**Grade 1 and 2 cognitive processing tests administration**

Students in Grade 1 and 2 were administered the tests in exactly the same manner except for the following three exceptions.

1. Slight variations were made to the script to ensure that the younger students knew that the tests were targeted to Grade 3 students. These variations are outlined with the rest of the script in Appendix F.
2. As previously noted, the tests that were administered varied according to the different grade levels.
3. As previously mentioned, the reading levels that were used as a measure of Grade 1 reading comprehension were administered by each students classroom teacher as a part of the schools own testing procedures.

Interestingly, when the Grade 1 and 2 students found out that the words in the test were of a Grade 3 standard and not meant for them, they relaxed. The researcher is of the opinion that once they realised this was not a standard that was expected of them, they felt no pressure and seemed to genuinely enjoy the testing. In this section the procedure for collecting the data has been reviewed. In the next section the methods for grouping students and analysing the data are outlined and discussed.

**Data Analysis**

**Statistical programs**

Data was analysed using three statistical analysis programs. First, the Predictive Analytics SoftWare (PASW; formally SPSS) version 17 (PASW, 2009) was used for most analyses. Second, the Rasch Unidimensional Model Management (RUMM) version 2020 (Andrich, Sheridan & Luo, 2004) was used for the Rasch analysis. Finally, a freeware Monte Carlo Parallel Analysis eigenvalue generator (Watkins, 2008) was utilised in conjunction with PASW for the exploratory factor analysis (EFA).

**Statistical Analyses**

A Rasch analysis in conjunction with exploratory factor and reliability analysis was utilised in order to ascertain which items from both word lists were most suitable for
the refined scales. As part of the scale refinement of both word lists, each analysis was performed iteratively after each item’s removal. The results from these analyses are reported in Chapter 5.

All data was screened to ensure that it met all the assumptions of the analyses used in the current study. Univariate outliers were identified through extreme $z$-scores and multivariate outliers by examining scatter plots and calculating Mahalanobis’s distances. All analyses were run with and without outliers to ensure that they are not affecting results, if not they were retained. In order to ensure that the data is suitable for EFA the Keiser-Meyer-Olkin (KMO) was used. KMO values over .9 indicate that the data is ‘marvellous’ for factor analysis and values over .8 indicate that the data is ‘meritorious’ (Kaiser, 1974). Box’s M was calculated; non significant alpha values indicated that the data was suitable for DFA.

Once the tests were finalised and assumptions checked, summary statistics, Cronbach’s alpha and correlation analyses were performed. To ascertain how reading ability and cognitive processing variables co-occur, a cluster analysis was utilised. In order to study LD identification, students were grouped on the basis of reading discrepancy using both a traditional and regression-adjusted discrepancy. DFA was utilised in order to ascertain if the cognitive processing variables could predict membership in the LD, LA or normal groups, dictated by the discrepancy groupings. Finally a tentative hypothesis of a new method of LD identification based on cognitive processing deficits was put forward. The results from these analyses can be found in Chapter 6.

It should be noted that many of the tests conducted in this study were conducted on multiple variables at the same time, either items within scales in Chapter 5, or full tests in Chapter 6. Therefore Bonferroni’s corrections were used whenever applicable. Levene’s test for homogeneity of variance was calculated before any significance testing. When the results indicated heterogeneity, degrees of freedom were adjusted accordingly.

**Hierarchical data**

The data collected for the current study was nested as there were students from different classes within different grades within different schools. One of the issues with using various schools or classes the primary sampling unit is that there tends to
be less variation within any given class of students than there would be between
them (O'Connell & McCoach, 2008). A variance components analysis was
conducted to ensure that membership in schools or classes is not unduly affecting
results.

There were many different groupings within the sample, particularly the Grade 3
cohort, which must be taken into account. For instance some of these schools had
already administered some of the tests used in the current study earlier in the year
and thus some participants sat them at different times of the year to others.
Furthermore, some groups had the tests administered by the teacher and others had
them administered by the researcher. Therefore t-tests will be utilised in order to
ensure that there was no systematic variance stemming from these procedural
differences.

**Missing data**

In total, 287 students participated in the current study. Of these, all 57 Grade 2 and
58 Grade 1 students completed all of the tests, so there was no missing data in either
of these cohorts. Of the 172 Grade 3 students who participated, six did not complete
the PAT-R, two did not complete the RPM and one did not complete the SAST.
Given the importance of these tests to the study and the small number of students
with missing data, these students were excluded from the sample for the analyses in
Chapter 6. However, as all 172 participants completed the word lists they were
included in the Rasch analysis in Chapter 5. There were no other missing data in the
study.

**Identifying students with learning disabilities**

Students were designated as LD, LA or RA as per the discrepancy method in order to
examine the differences between different methods of identification. Two
discrepancy methods of identification were utilised in the current study, traditional
and regression-adjusted discrepancy. These two methods of allocating the Grade 3
students are detailed below.

**Traditional discrepancy**

The traditional discrepancy method involves identifying students who have an
aptitude score, in this case the RPM, that is significantly higher than their reading
score, in this case the PAT-R. The logic behind this is basically that students should
be reading at a level commensurate with their aptitude and that LD could explain this unexpected discrepancy between aptitude and achievement.

Because both the RPM and the PAT-R scores, as outlined in Chapter 6, will have already been converted to $z$-scores, identifying the LD cohort was simply a matter of identifying students who had a $z$-score in the PAT-R one standard deviation below their score on the RPM. Although the arbitrary nature of a one standard deviation discrepancy is recognised, it is often used in previous research and as such will provide generalisable results (Schatschneider et al., 2002). Therefore, any student that performed more than one standard deviation below the mean (i.e., a $z$-score of -1), who did not meet the criterion of LD, was designated LA.

*Regression-adjusted discrepancy*

In order to ascertain the groupings of students via the regression-adjusted discrepancy, the same basic process for identifying students as per the traditional discrepancy method was repeated. As discussed in Chapter 2, this method of identifying students is similar in theory to the traditional discrepancy method. However, regression to the mean is accounted for by incorporating the correlation between the two tests. The difference score in a regression-adjusted discrepancy is representative of the difference between aptitude and achievement as predicted by aptitude (Evans, 1996), rather than the discrepancy between aptitude and achievement. Therefore, students were designated as discrepant if the criterion in Equation 1 is met:

$$ (r \times z_{\text{RPM}}) - z_{\text{PAT-R}} \geq 1 \quad (1) $$

where $z_{\text{RPM}}$ is any given student’s $z$-score on the RPM and $z_{\text{PAT-R}}$ is the same student’s $z$-score for the PAT-R and $r$ is the correlation between the two for the entire sample.

Once these groups of students were identified, one-way ANOVA will examine the cognitive processing differences between these three groups. Although there are multiple dependent variables in these analyses, MANOVA will not be calculated in the current study. The use of multiple ANOVAs as a follow up to a significant MANOVA is not recommended despite the regular use of this method (Enders, 2003). However, it is this follow up, in which the group differences on the individual dependent variables are analysed that is of interest in these analyses.
Furthermore, Grice and Iwasaki (2007) have warned against using MANOVA if the relationships between the variables are not going to be taken into account in the interpretation of the results. They suggest that if this is not going to be done, then separate univariate ANOVAs are more appropriate. As all analyses in the current study are geared towards identifying LD on the basis of dichotomous impaired/not impaired type deficits, rather than being based on multivariate linear equations, multiple ANOVAs were deemed more appropriate than MANOVA. This decision requires Bonferroni corrections to be applied to alpha levels for these analyses.

**Rasch analysis**

As discussed in Chapters 2 and 3, one of the aims of the current study is to develop word lists for use as screening tools. These lists need to be psychometrically sound. In order to test these psychometric properties using the best possible current knowledge, a Rasch analysis was utilised for the refinement of the word lists. While Rasch analysis is often considered to fit under the umbrella of IRT, there is a fundamental difference. In IRT it is expected that the model will fit the data, whereas in Rasch the expectation is that the data will fit the model (Andrich, 2004). However, in recent times there has been more agreement than disagreement between the two paradigms (Andrich).

Much of the analyses within Rasch are based on logit scores. These scores represent the ability of the person taking the test as measured by the items within that test. Therefore, someone with high proficiency in irregular word reading will have a higher logit score than someone with a low proficiency in irregular word reading. The interesting thing about these logit scores is that items are measured on the same scale as participants. For example, high scoring participants will receive similar logit scores as difficult questions. Consequently, items can be seen as separating participants on the logit scale, and one factor that can be taken into consideration when judging the value of an item is how many participants, or what participants, that item separates.

There are many different criteria that can be tested via a Rasch analysis. Those used in this analysis are thresholds, individual item fit, targeting, reliability and dimensionality. It is often recommended to utilise tests of CTT in conjunction with Rasch, hence both EFA and internal consistency measures were utilised as part of the psychometric validation of the word lists.
It has been found that in order to be 99 percent confident of having item calibrations stable within half a logit, a precise level of measurement, at least 150 participants are required (Linacre, 1994). Because there were 172 participants in the Rasch analysis, the data set was sufficient for fairly precise measurement.

** Thresholds 
It is important to ensure that each item in a scale is correlated with the total score from that same scale (Churchill, 1979). Traditionally, within CTT, an item-total correlation has been used to ensure that this happens. However, Rasch analysis goes one step further by analysing the thresholds of each individual item. This is done by looking at the overall logit score (which represents the total score on the scale) and examining the point at which each possible response to an item becomes the most likely (Erhart et al., 2009).

The first thing to check when examining thresholds is that the responses are in the right order. That is, do participants who endorse the one-point option on a given item have lower total scores than those who endorse the two-point option, and equally, higher total scores than those who endorse a zero point option? It is also important to ensure that each possible response for any given item is at one point the most likely response at an appropriate point on the logit scale. If this is not happening, especially on multiple items throughout the scale, then that item response is not providing sufficient new information to significantly contribute to the scale.

In RUMM, two functions are used to examine this issue. Firstly, the threshold map can provide an overview of the entire scale and highlight any items that are not appropriately ordered. Secondly, the category probability curves can give more insight into each item so that those items with disordered thresholds can be more closely examined.

** Individual item fit 
Item fit analysis can be used to indicate if any items are not sufficiently measuring the targeted construct. If you imagine a straight diagonal line with the difficulty of items on one axis and the ability of people getting that item right or wrong on the other, the item fit analysis will highlight items that stray too far from that line (Bond & Fox, 2001).
In terms of checking if the items as a whole fit, the standard deviation of the mean residual will be under 1.5 if the overall item fit is good (Pallant, personal communication, January 28, 2009). When analysing individual item fit there are two figures that need to be checked. Firstly, any item with a significant $\chi^2 p$ value does not have good item fit. Secondly, any item with an item fit residual exceeding an absolute value of 1.5 is also demonstrating poor fit. Although these cut-offs are to a certain extent arbitrary, they are assessed with the overall item fit of the whole scale taken into account.

**Person fit**

Due to the constant feedback loop between items and persons in Rasch analysis, the checks and balances for person fit are similar to those of item fit. Once again, a standard deviation of the mean fit residual that is less than a magnitude of 1.5 indicates overall person fit (Pallant, personal communication, January 28, 2009). Individual person fit is also indicated by a significant $\chi^2 p$ value or a person fit residual with a higher absolute value than 2.5. Furthermore, any persons who correctly answered all, or none, of the items in the test are considered extreme within a Rasch paradigm, as their true ability can not be represented in the model. This is because one cannot differentiate between two students who got all items correct, even if there is a significant difference in ability between the two, as there are no items on the test to split them.

Although it is sometimes recommended that any students with extreme scores are removed from an analysis (Ecosse et al., 2004), it was decided that they should be retained in the current study. This is because the anticipated negative skew may result in the removal of too many persons from the analysis. Furthermore, it can be assumed that an extreme score on a scale targeted to a certain population, such as a grade level, is representative of a score fairly similar to the extreme score (Wright, 1990). It is important to note that if the overall person fit statistics indicate problems, then steps were taken to rectify this, and that these steps may include removal of extreme persons. However, if the overall person fit does not indicate problem then these participants may not be removed.
Dimensionality
Although the knowledge of the dimensionality of a scale is always important to the understanding of the scale, unidimensionality is an especially important assumption when adding items together to get an overall score (Tennant, 2000). This was tested, as recommended by those who work within a Rasch paradigm, not just with the principal components analysis residuals developed through the Rasch model, but also with EFA as per CTT (Tennant & Pallant, 2006). It is vital that EFA is tested through CTT and not just internal consistency because unidimensionality is an assumption underlying Cronbach’s $\alpha$ (Cronbach & Shavelson, 2004).

In terms of what type of EFA were used for the current study, a principle axis factor analysis was chosen, firstly because it has been demonstrated that factor analysis is less prone to falsely inflated loadings than principal components analysis (Guedens, Sandra & Van den Broeck, 2004). Furthermore, this specific type of factor analysis has been used successfully in conjunction with Rasch analysis in the past (e.g., Strong & Kahler, 2007; Strong, Kahler, Ramsey & Brown, 2003).

In any EFA, there is a range of decisions to be made such as the criteria for retaining factors, whether to allow correlation between factors and the method of rotation. Because the primary objective of this analysis was to determine unidimensionality, the most important of these decisions was on how to decide the number of factors to retain. A Monte Carlo parallel analysis was utilised rather than a scree plot, which is quite subjective, or an ‘eigenvalues greater than one’ method, that often results in over-identification of factors (Hayton, Allen & Scarpello, 2004). A Monte Carlo parallel analysis generates a random data set with the same parameters as the existing data set. The eigenvalues from the actual and random data sets are then compared. The point at which the lines on the scree plots coincide is the point representing the number of factors that should be retained. Therefore, it is simply a measure of how many eigenvalues in the data set are larger than would be expected by chance.

Within the Rasch analysis unidimensionality was tested via the principal components analysis residuals produced in RUMM. A $t$-test was conducted on the residuals of the three items with the highest positive residuals against the residuals of the three items with the highest negative residuals (Smith Jr, 2002). If less than five percent
of simulations have significantly different scores on these two sets of items at an alpha level of .05, then unidimensionality is supported.

**Targeting and reliability**

Within the paradigm of CTT, Cronbach’s $\alpha$ was used to ascertain each test’s reliability. A value of .8 indicates good internal consistency, a value of .9, excellent internal consistency (George & Mallory, 2003). Although the Kuder-Richardson 20 is often used for dichotomous variables, Cronbach’s alpha was actually originally developed in order to cater for scales that were dichotomous and has been deemed suitable for dichotomous data (Cortina, 1993). Within IRT, the Person Separation Index (PSI) is a measure of internal consistency that unlike Cronbach’s $\alpha$, is not unduly affected by extreme scores (Schumacker & Smith Jr, 2007). The values of the PSI are similar to what could be expected from a Cronbach’s alpha, which was also assessed in this scale in order to utilise measures from both CTT and IRT.

Another method of assessing if a scale has been appropriately targeted is the use of a person-item map. This graph shows the difficulty of items lined up against the skill level of participants in order to see if the two are well matched. It is hoped in the current study that although the majority of participants have items with a difficulty corresponding to their skill level, there is a higher concentration of items targeted at those participants at the lower skill level.

**Cluster analysis**

Cluster analysis is primarily used as a tool for designating participants into groups or clusters. Participants within the generated clusters will have similar scores on the variables (Hartigan, 1964). Cluster analysis could be seen as similar to EFA; however instead of grouping variables, participants are grouped.

In the current study the aim of the cluster analysis is of an exploratory nature. This is important as it has been suggested that cluster analysis will provide poor answers to specific questions (Cormack, 1971). What is certainly not happening in this cluster analysis is an attempt to ascertain the relevance of the cognitive processing variables to LD. Instead, the primary objective of this cluster analysis is to ascertain how related these variables are to reading and spelling proficiency, before the LD based analyses are run.
In order to ascertain how students could be grouped within the Grade 3 data set, a two-step cluster analysis with log-likelihood distance measures was performed. This method was selected on the basis of its simplicity and flexibility with data that does not meet the stricter criterion of normality (Chiu, Fang, Chen, Wang & Jeris, 2001). This is important given the anticipated skew of the word lists. The Bayesian Information Criterion (BIC) has been shown to be an excellent method of automatically identifying an appropriate number of clusters to retain (Fraley & Raftery, 1998), so the automatically generated number of clusters in both the reading and spelling models were retained. Furthermore, the log-likelihood distance measure, suitable for both continuous and categorical data and less effected by limited ranges within a variable (Zheng, Wang & Hu, 2007), was selected due to the narrow range of the digit span measures.

The focus of the analysis is on those who are not reading or spelling well for a wide range of reasons, possibly including LD. Consequently, any post hoc analyses on the differences between the groups identified via the cluster analysis was based on finding out as much information as possible about the poorest reading and spelling students. This is because the cognitive processing variables relevant to high achievement, although important, are not relevant to this particular study.

**Discriminant function analysis**

The cluster analysis was performed in order to identify the cognitive processes common to poor readers. However with LD, LA and RA groups identified by two different methods, it is possible to see if these processes can predict membership in these groups as dictated by both discrepancy methods. The primary aim of the Discriminant Function Analysis (DFA) is to see if cognitive variables alone can predict the groupings of students based on their aptitude and reading achievement scores. Another point of interest in the analysis is that it may illuminate some of the cognitive processing differences between LD and LA students. This is because the DFA predicts membership in groups conceptually aligned with the concept of LD, using cognitive variables that are thought to be the cause of it. Both LD and LA groups consist of poor readers, but it is hoped some light can be shed on the question of what, if anything, separates the two groups.

It is common for DFA to be utilised in order to ascertain whether a set of variables can predict a grouping variable, such as the existence of a disorder (Betz, 1987). It
has been said that logistical regression should be chosen over DFA if there are normality issues with the data set (Press & Wilson, 1978). However, the five variables chosen, nonword reading, the two digit span tests and the two RAN tests were fairly normally-distributed and as such suitable for DFA. If there was more theoretical justification for selecting orthographic processing as a deficit inherent to LD then logistic regression would have been more appropriate, but without it, DFA is preferable.

**Identifying students with learning disabilities via cognitive deficits**

The cluster analysis was used to test how closely related the cognitive processing variables are related to reading and spelling. The DFA was used to see if these same processes can be used to predict the presence of LD. The final step was to set up a provisional model under which students with possible LD can be identified on the basis of these cognitive processes alone.

Given the finding that the correlation between reading and these cognitive processing variables can be stronger at lower levels than at higher levels (Schatschneider et al., 2002), students were identified on the basis of absolute deficits in cognitive processes rather than regression based equations. Furthermore given the finding that some of these processes can be used to compensate if other variables are deficient as per the double deficit model, more liberal cut-offs were put in place when there is more than one deficit present.

Much of the decision making in this section will depend on the results of the other analyses. For instance there is no need to use both naming speed tests if one seems to be more appropriate in the DFA or cluster analysis. Furthermore, based on previous literature, any deficits in naming speed or verbal memory were identified on the basis of grade-age comparisons, whereas phonological deficits were based on reading-age comparisons.

**Age-Comparisons**

In all norm-referenced testing there needs to be a normative sample to which a given participant’s score is compared. In LD research two primary comparison groups have been utilised, chronological age groups and reading age-matched groups.
Chronological age comparison
Chronological age comparisons are the most commonly-used age comparison, especially in education. This is not an age comparison as found in the RPM and SAST, in which norms are given in six month and one month age groups, respectively. However, many tests have norms that are developed at grade levels, including the PAT-R, with a similar effect.

The advantages of using grade-level age comparisons are of a pragmatic nature; teachers can assess students by their peers rather than their age groups, which require comparisons of age groups across different grade levels where learning experiences are a confounding factor. Because the SAST and RPM used chronological age rather than grade-level age comparisons, and the latter is more appropriate for the current study, raw scores in both tests were converted into z-scores. Because both tests were only administered to the Grade 3 cohort this means that, by default, the z-scores are based on a grade-level comparison. Deficits in naming speed and verbal memory, however, will be based on grade-age deficits as previous studies have demonstrated that LD students do not have reading-age deficits in these processes (de Jong, 1998; Griffiths & Snowling, 2001).

Reading-age comparisons
Reading-age comparisons can provide valuable information on qualitative differences in the reading ability of students. For instance compare these two hypothetical students: Both are in Grade 3 and read at a Grade 2 level. Neither of these students is able to read nonwords at a normal Grade 3 standard, one of them can read nonwords at a Grade 2 level, as predicted by their reading ability. However the other student does not even have the phonological processing abilities of a Grade 2 student at a normal reading level. This is an indication that the problems that the student is facing are not simply developmental and instead may be due to a processing deficit (Stanovich, 1988).

Because the relationship between nonword reading and reading comprehension is to some extent reciprocal rather than a direct causal relationship (Perfetti, Beck, Bell & Hughes, 1987), a chronological age comparison of nonword reading is less indicative of phonological awareness than a reading age comparison. Reading age comparisons are used when the aim is to ascertain how well a student is performing on a test given their proficiency in reading, rather than their chronological age (Snowling et
al., 1986). For an example of how reading age-matches can be calculated, see Share and Shalev (2004).

Although the reading-level matched design has many advantages, it is vital to ensure that the students are matched on appropriate tasks (Stanovich, 1988). Although the Grade 1 test of reading ability is different to the Grade 2 and 3 PAT-R, both are measures of reading achievement. Therefore, reading ages for Grade 3 students were utilised using the equivalency tables for the Grade 2 PAT-R, which provides comparable normative data for students in Grades 1, 2 and 3. Using this as a basis for cross referencing an average score, a Grade 2 reader at the 50th percentile was found to be equivalent to a Grade 3 reader at the 25th percentile. This is because a student in Grade 2 getting a raw score of 17 out of 25 on the Grade 2 PAT-R is in the 50th percentile, and a Grade 3 getting the same score is in the 25th percentile. Furthermore, using the same logic, a Grade 1 reader at the 50th percentile is equivalent to a Grade 3 reader in the 10th percentile.

Based on this approximation, a Grade 3 student who scores in the 25th percentile in the PAT-R is reading at the average Grade 2 level and a student who scores in the 10th percentile in Grade 3 is reading at an average Grade 1 level. However, it is not enough to simply designate cut-offs at the 25th and 10th percentile as this is not an accurate representation of some student’s reading level. For instance, a Grade 3 student who scored in the 26th percentile would surely be closer to reading at a Grade 2 rather than Grade 3 level. This is best demonstrated by the normal distribution curves in Figure 5.

In Figure 5 the assumption that has been made is that all three grade levels have similar distributions. As can be seen in Figure 5, there are three distribution curves representing the reading performance of the three grade levels, with Grade 3 percentiles on the x-axis. To keep this information in line with the rest of the study, these percentiles then need to be converted to z-scores, but on one overall scale rather than one scale per grade level. Therefore, the 50th percentile is equivalent to a z-score of 0, the 25th percentile is equivalent to a z-score of -0.67, and the 10th percentile is equivalent to a z score of -1.28. Using this as a baseline, the z-scores
Figure 5. Normal Distribution Curves of Grade 1, 2 and 3 Readers on a Grade 3 Percentile Axis

calculated per year level for Grade 1 and 2 can now be transformed to be relevant to a Grade 3 standard. For Grade 2 students this was done by subjecting the Grade 2 scores to the following transformation:

\[ z = z_2 - 0.67 \quad (2) \]

where \( z \) = the Grade 3 equivalent reading comprehension score and \( z_2 \) = the original \( z \) scores developed for reading comprehension for the Grade 2 sample. Following on from this the Grade 1 \( z \)-scores were adjusted in the following manner

\[ z = z_1 - 1.28 \quad (3) \]

where \( z \) = the grade 3 equivalent reading comprehension score and \( z_1 \) = the original \( z \) scores developed for reading comprehension for the Grade 1 sample. Now all three grades have comparable scores from the same scale and can be treated as one data set. It must be acknowledged that this is a fairly crude manner for developing equivalency scores, especially given that the Grade 1 reading comprehension test was qualitatively different from the Grade 2 and Grade 3 tests. However, given the exploratory nature of this study and the unavailability of appropriate Grade 1 tests, these tests were used as an equivalent to the Grade 2 and Grade 3 tests. Figure 6 is a graph of the reading comprehension scores at a Grade 3 level, for Grade 1 to 3 students.
Figure 6. Normal Distribution Curves of Grade 1, 2 and 3 Readers on a Grade 3 z-Score Axis with Cut-Points Designating Reading Age

As can be seen in Figure 6, all three of these curves have a standard deviation of one and the mid-point between the means is the point at which a given test score is best represented by a given grade-level changes from one grade to another. Therefore, as noted on the graph, the z-scores on these transposed scores that divide the standard of the Grade 3 readers as those reading at a Grade 1, 2 or 3 level are -0.98 and -0.34. This means that for a student to be designated as reading at a Grade 2 level they need to have a z-score between -0.98 to -0.34. Furthermore, a student needs a score below -0.98 to be given a Grade 1 reading age, or over -0.34 to be given a Grade 3 reading age.

With the reading age of students allocated, the next step was to choose the method by which reading age-based phonological deficits would be identified. Share and Shalev (2004) developed a standard of nonword reading required to meet a reading age comparison that was represented by:

\[ X > M - SE \] (4)

Where X is the score of any given student, M is the mean for the appropriate Grade-level and SE is the standard error of that mean. Any students who did not meet this requirement could be seen as having a reading age-based deficiency. However the primary problem with this method is that the standard error is dependent on sample size, and as such will vary significantly from group to group. Consequently, a different cut-off was required, one that would be more consistent.
primary advantages of using the standard error is that it is smaller than a standard deviation. One standard deviation below the mean for a reading age-based score will result in very low cut off scores, resulting in under-identification of deficits. Therefore Equation 5 was used to identify students with reading age-based phonological deficits

\[ X > M - (SD/2) \]  \hspace{1cm} (5)

where \( X \) is the score of any given student, \( M \) is the mean for the appropriate Grade-level and \( SD \) is the standard deviation of that mean. Those that do not reach the minimum score for their reading age are designated as deficient in phonological processing. This process is outlined in more detail in Chapter 6, once the relevant information to perform the calculations has been ascertained through the analyses in Chapters 5 and 6.

**Summary**

This chapter has detailed how the study was conducted and the research paradigms that guided the decisions made in the process. These research paradigms, causal-comparative research and applied research are fairly orthogonal; as such there were no clashes between the two paradigms. The psychometric issues specific to the use of screening tools were also reviewed. Furthermore the two test theories utilised for the current study, CTT and IRT can be used in a complementary fashion in order to obtain the most comprehensive information on the word lists that have been developed for this study.

The tests used in the current study, including both the established and newly developed tests were reviewed. The sampling framework and method of data collection were also evaluated. The method of identifying LD, LA and RA students with the traditional and regression-adjusted discrepancy methods was outlined. Furthermore the method of identifying grade-age and reading-age based deficits in cognitive processes was also summarised. Finally, the decision-making process, as pertinent to the data analysis was reviewed. The Rasch analysis is outlined in Chapter 5 and the results of all the other analyses are reported in Chapter 6.
Chapter 5

Overview

In Chapter 3 it was argued that there is a pressing need for psychometrically validated tests that measure the cognitive processing deficits that may cause Learning Disabilities (LD). The most important cognitive process in reading research is phonological processing (Shankweiler & Fowler, 2004). Nonword reading, as a test of phonological processing, has been proposed as the primary diagnostic test for reading disabilities (Stanovich, 1999). The role of orthographic processing in LD is less clear, inasmuch as its role as a cause or effect of LD is still unknown (Rack et al., 1992). However, the irregular word list can provide an interesting contrast to the nonword list as it consists of words that must be learned, as compared to the nonwords which are made up of novel stimuli. This usefulness has been recognised, as is evidenced by similar tests that utilised orthographic processing tests (Edwards & Hogben, 1999; Good, Kaminski & Dill, 2007). One of the more common and simple measures of orthographic processing is irregular word reading (Kirby et al., 2008).

The aim of this chapter is to refine and assess the psychometric properties of the irregular and nonword lists as per the tenets of Item Response Theory (IRT), using Rasch analysis. However, tests from Classical Test Theory (CTT) will also be used to assess the psychometric properties of the scale. Of the six tests developed in the current study, only the word lists will be validated through IRT given that the Rapid Automatic Naming (RAN) and digit span tests are single-item scales. Furthermore, both the RAN and the digit span tests, in the same formats as used in the current study, have been psychometrically validated in other studies (e.g., Denckla & Cutting, 1999; C. Reynolds, 1997).

This chapter is in three sections. The first section is a brief outline of how a Rasch analysis is conducted and what kinds of results are expected. The next two sections present results from Rasch Analyses, first for the irregular word list and second for the nonword list. In each of these two sections, a brief summary of the analysis on the full word list is presented, as well as information on the selection process for removing items. A complete Rasch analysis is also presented for the final refined word lists.
**Method**

Data from all 172 Grade 3 students in the sample outlined in Chapter 4 were utilised in the analyses reported in this chapter. Each of these students was individually tested in a written stimuli and verbal response format. Each student was asked to read a list of irregular and nonwords out loud. Their responses were recorded and scored at a later date. All 172 students completed both lists, so there were no missing data.

Scoring consisted of two-points for each word read out loud correctly and zero points for each incorrect or passed response. One-point was awarded for those students who gave the correct response for the root of the word but got the suffix or plural status of the word incorrect. For instance, *handstands* read as *handstand*, or *exhausted* read as *exhausting*, received one-point. The acceptable one-point responses were outlined in Chapter 4.

**Rasch Analysis**

In this section, the processes used to refine the scales are outlined, along with any guidelines that have been put in place as part of the analysis. As discussed in Chapter 4, common to many statistical analyses, there are somewhat arbitrary cut-offs used in Rasch analysis to assess the psychometric properties of a scale. First of all, overall model fit is assessed by a $\chi^2$ model fit analysis in which non-significant $p$ values indicate that there is model fit (Pallant & Tennant, 2007). The Person Separation Index (PSI) is a measure of how well the scale separates people with different skill levels. Expectations of PSI are similar to that of Cronbach’s alpha, with values of .7 considered acceptable but values over .8 preferred (Tennant & Conaghan, 2007).

Much of the analyses within Rasch modelling are based on logit scores. These scores represent the ability of the person taking the test as measured by the items within that test. For instance, someone with high proficiency in irregular word reading will have a higher logit score than someone with a low proficiency in word reading. These logit scores are similar to $z$-scores, and can be interpreted in the same way, although there are differences between how the two types of scores can be interpreted. The first main difference is that the mean of a $z$-score will always be zero but a mean logit score of zero indicates what the mean would be if the scale is
perfectly targeted. The second main difference is that it is not just scores that are on this scale. Items and participants are measured on the same scale in a Rasch analysis. Consequently, an item with a high positive score will be quite difficult and have a similar logit score to an item that differentiated high achieving persons on that scale. Because both items and participants are measured on the same scale, items can be seen as separating participants. For instance, the most difficult item should separate the highest performers in the test from the rest of the sample.

Another outcome from the measurement of participants and items on the same scale is that information about the targeting of the scale can be gained by the mean person logit score. This is because the mean item logit score is anchored at zero (Linacre, 2005), and therefore, the mean person logit score is an indication of how easy or difficult the item is. If the mean person logit score is negative, this is an indication that the scale is positively scored because, overall, the participants found the scale easy and thus had negative logit scores.

In Rasch analysis there are cut-off scores that are used to ascertain item and person fit. Overall item fit residual is once again expected to have a mean of approximately zero and most importantly a standard deviation under 1.5 (Pallant, personal communication, January 28, 2009). Similar again to person fit, individual item fit residuals are expected to have an absolute value of less than 2.5 as well as non-significant $\chi^2$ values (Pallant, Miller & Tennant, 2006). Differential Item Function (DIF) analysis measures how influenced any given item is by the existence of clusters within the sample. In this study these clusters are represented by gender and school. Any items that have different response patterns from different groups at the same proficiency level of the measured trait should be deleted from a scale (Pallant et al.).

Two analyses from CTT are used in conjunction with the Rasch analysis. The first CTT analysis, Exploratory Factor Analysis (EFA), will be used for this. Specifically, a principal axis factor analysis will be used in a conjunction with a Monte Carlo parallel analysis, to ascertain how many factors best fit into the scale. If only one-factor is found then this supports unidimensionality for the scale. With unidimensionality found, the second CTT analysis, Cronbach’s $\alpha$, will be used to measure internal consistency.
In addition to using EFA to check for unidimensionality as per CTT, there is also a method of checking this within IRT. The first step of this analysis is to check the principal components analysis residuals generated by the Rasch Unidimensional Model Management (RUMM) program to see which items have the highest loadings on the second component identified by the analysis. The person locations of the highest positive loading items are then compared to the person locations of the highest negative loading items in an individual group’s $t$-test to ensure that there is no difference between these groups. If the tests are statistically significant in less than five percent of simulations then unidimensionality is supported (Smith Jr, 2002).

**Rasch analysis results for the irregular word list**

All 172 participants from the Grade 3 cohort completed the word lists and only their responses were utilised for the Rasch analysis. Analyses were conducted on the 37 items on the irregular word list. The original scoring system involved three response options, namely, two-points for a correct answer, zero points for an incorrect answer or one-point for a partially correct answer.

In addition to examining the efficacy of this scoring system, the value of each item in the list was assessed so that the scale could be refined through item deletion. It should be noted that no decisions about item deletion or rescoring were made on the basis of any one test. Rather, the analyses from all the tests were considered as a whole before any decisions were made. Once the full analysis was examined, items were then removed one by one and the analyses repeated after each deletion. The results from the initial analysis are briefly presented below and are followed by the full analyses on the refined word list in order to demonstrate the psychometric properties of the scale.

**Full word list**

The simple descriptive statistics scores for each word in the full irregular word list are shown in order of difficulty in Table 5. The scoring system for each item is also shown in Table 5. Some items did not have a one-point option resulting in only two possible scores. Those words that did have the one-point option had three possible scores.
Table 5

*Means and Standard Deviations of the Original Irregular Word List*

<table>
<thead>
<tr>
<th>Word</th>
<th>M</th>
<th>SD</th>
<th>SO</th>
</tr>
</thead>
<tbody>
<tr>
<td>To</td>
<td>2.00</td>
<td>0.00</td>
<td>2</td>
</tr>
<tr>
<td>said</td>
<td>2.00</td>
<td>0.00</td>
<td>2</td>
</tr>
<tr>
<td>we</td>
<td>1.99</td>
<td>0.15</td>
<td>2</td>
</tr>
<tr>
<td>sea</td>
<td>1.99</td>
<td>0.15</td>
<td>2</td>
</tr>
<tr>
<td>blue</td>
<td>1.99</td>
<td>0.15</td>
<td>2</td>
</tr>
<tr>
<td>have</td>
<td>1.98</td>
<td>0.22</td>
<td>2</td>
</tr>
<tr>
<td>put</td>
<td>1.98</td>
<td>0.22</td>
<td>2</td>
</tr>
<tr>
<td>use</td>
<td>1.97</td>
<td>0.26</td>
<td>2</td>
</tr>
<tr>
<td>people</td>
<td>1.97</td>
<td>0.26</td>
<td>2</td>
</tr>
<tr>
<td>high</td>
<td>1.96</td>
<td>0.27</td>
<td>3</td>
</tr>
<tr>
<td>eight</td>
<td>1.95</td>
<td>0.30</td>
<td>2</td>
</tr>
<tr>
<td>who</td>
<td>1.95</td>
<td>0.30</td>
<td>2</td>
</tr>
<tr>
<td>zero</td>
<td>1.92</td>
<td>0.40</td>
<td>2</td>
</tr>
<tr>
<td>know</td>
<td>1.91</td>
<td>0.41</td>
<td>3</td>
</tr>
<tr>
<td>boat</td>
<td>1.91</td>
<td>0.42</td>
<td>2</td>
</tr>
<tr>
<td>done</td>
<td>1.91</td>
<td>0.42</td>
<td>2</td>
</tr>
<tr>
<td>four</td>
<td>1.90</td>
<td>0.45</td>
<td>2</td>
</tr>
<tr>
<td>juice</td>
<td>1.90</td>
<td>0.42</td>
<td>3</td>
</tr>
<tr>
<td>laugh</td>
<td>1.87</td>
<td>0.49</td>
<td>2</td>
</tr>
<tr>
<td>emu</td>
<td>1.84</td>
<td>0.55</td>
<td>2</td>
</tr>
<tr>
<td>of</td>
<td>1.82</td>
<td>0.57</td>
<td>2</td>
</tr>
<tr>
<td>frightened</td>
<td>1.77</td>
<td>0.61</td>
<td>3</td>
</tr>
<tr>
<td>towel</td>
<td>1.71</td>
<td>0.71</td>
<td>2</td>
</tr>
<tr>
<td>stomach</td>
<td>1.70</td>
<td>0.72</td>
<td>2</td>
</tr>
<tr>
<td>want</td>
<td>1.69</td>
<td>0.73</td>
<td>2</td>
</tr>
<tr>
<td>excited</td>
<td>1.66</td>
<td>0.73</td>
<td>3</td>
</tr>
<tr>
<td>especially</td>
<td>1.63</td>
<td>0.78</td>
<td>2</td>
</tr>
<tr>
<td>characters</td>
<td>1.59</td>
<td>0.80</td>
<td>3</td>
</tr>
<tr>
<td>jewellery</td>
<td>1.51</td>
<td>0.86</td>
<td>3</td>
</tr>
<tr>
<td>somersault</td>
<td>1.48</td>
<td>0.88</td>
<td>3</td>
</tr>
<tr>
<td>rhinoceros</td>
<td>1.45</td>
<td>0.89</td>
<td>3</td>
</tr>
<tr>
<td>tortoise</td>
<td>1.30</td>
<td>0.96</td>
<td>2</td>
</tr>
<tr>
<td>certificates</td>
<td>1.23</td>
<td>0.96</td>
<td>3</td>
</tr>
<tr>
<td>cough</td>
<td>1.21</td>
<td>0.98</td>
<td>2</td>
</tr>
<tr>
<td>exhausted</td>
<td>1.05</td>
<td>1.00</td>
<td>3</td>
</tr>
<tr>
<td>sew</td>
<td>0.79</td>
<td>0.98</td>
<td>2</td>
</tr>
<tr>
<td>cooperation</td>
<td>0.77</td>
<td>0.97</td>
<td>3</td>
</tr>
</tbody>
</table>

N = 172. *Note.* SO = Number of Scoring Options
Two words, namely, to and said, were correctly read by all participants and were thus deleted from the pool of items. Furthermore, the words blue and sea had a correlation coefficient of 1 so sea was deleted from the item pool in order to avoid singularity. The mean total of the adjusted 34 item list was 57.21 (SD= 10.95) out of a possible maximum score of 68 and 13% of students attained this maximum score. This latter statistic indicated issues with a ceiling effect.

In order to test for unidimensionality, a principal axis factor analysis in conjunction with a Monte Carlo parallel analysis was conducted. Although the data was suitable for EFA (KMO = .83), unidimensionality for the entire scale was not supported. A comparison of the eigenvalues generated from the EFA and the eigenvalues generated from the parallel analysis indicated that a two-factor model was best suited to the data. This can be seen clearly in Figure 7 in which a scree plot of the eigenvalues from each analysis is shown. It is not until the third factor that the eigenvalues from the irregular word list become lower than what would be expected by chance. This indicates that a two-factor model is most appropriate for this list, and therefore unidimensionality is not supported.
Based on the CTT analyses alone, the full irregular word list has issues with both unidimensionality and normality. However, more information was needed before any decisions about item deletion were made so a Rasch analysis was performed on the 34-item scale. A highly significant model chi-square, $\chi^2 (68) = 163.05, p < .001$, indicated model misfit, despite an excellent PSI of .90.

The threshold map for the 34 remaining items can be seen in Figure 8. For items with ordered thresholds (e.g., we, of, who etc), the point on the logit scale that the most likely endorsed response switches from one scoring option to another can be seen in the threshold map. For example, students who score zero points on an item should have lower overall scores than those students who scored one-point on the same item. Similarly, those students who scored two-points on the same items should have the highest overall scores.

When an item has ordered thresholds, a bar graph can be generated demonstrating at what points on the overall logit scale the most likely outcome for each item becomes a zero, one or two-point response. However, when an item does not have ordered thresholds and the likelihood of a higher response on a given item does not increase as overall score increases, no such bar graph can be generated. As can be seen in Figure 8 none of the items with a three-level scoring scale had an ordered threshold. Not only did all the items with the three-level scoring system have disordered thresholds, but all of the items with the two-level scoring scale had ordered thresholds. This indicates that the problem is with the scoring system and not the items.

In order to further investigate the issues with the three-level scoring system, the category probability curves for individual items within the list were generated. These graphs display the likelihood of endorsing a given response on a particular item at various levels of total scores. Similar to most of the figures generated within a Rasch analysis, the x axis on this graph displays the logit scores based on the whole analysis. The y axis displays the probability of the item response in question being endorsed at that point on the logit scale. In an item that fits well into the scale, the higher the overall logit score, the greater the likelihood that the maximum score will be received on that particular item. Figure 9 displays the category probability curve for a typical item with a disordered threshold from the irregular word list.
Figure 8. Threshold Map for the Original Irregular Word List

$N = 172$

** = Disordered Threshold.
Figure 9. Category Probability Curve for the item frightened

Figure 9 displays the issue with the items with disordered thresholds. The one-point option is not more likely to be endorsed by students with lower scores than those who selected the zero point option, nor by students with higher scores than those who selected the two-point options. Instead the one-point option is never the most likely endorsed option at any point on the logit scale and, as such, it is not providing enough information to be of use in the scale. In other words, the partial credit one-point response is not discriminating between participants. Therefore, the scoring system was amended so that students received one-point for a correct answer and 0 points for an incorrect or partially correct answer. For instance, a student reading the word frightened as frightening will receive zero points on that item.

This amended scoring system is reflected in the category probability curve for the re-scored item, frightened, as can be seen in Figure 10. The point on the overall logit scale when a participant is equally likely to endorse either the zero or one-point option is shown in Figure 10 by the point where the two lines intersect. It is worth noting that this point is not at zero on the logit scale. This is because this was one of the more relatively difficult items within the scale. For an easier item, this intersection point will occur at a negative point on the logit scale.
With the scoring system adjusted the Rasch analysis still indicated that there was model misfit, $\chi^2(68)=180.15$, $p<.001$, however the PSI was still high at .89. There were no issues with either overall item or overall person fit residuals ($SD = 0.91$ and .40 respectively). Closer analysis at the item level revealed that there were issues with a few of the individual items which were taken into account when deleting items for the final scale.

**Item removal**

The full, re-scored, original irregular word list had four issues that needed to be kept in mind while refining this list. Items were selected for removal on the basis of item misfit, dimensionality, DIF, or redundancy. After each item removal, all analyses were re-calculated to ensure that the most appropriate items were removed. In Table 6, the reason for each item’s removal and a brief snapshot of the summary statistics for the scale after each item was removed is displayed.

There were a range of reasons listed in Table 6 that would result in item removal. Items were removed on the basis of individual item misfit if the individual $\chi^2$ value was significant after a Bonferroni correction. By removing these items, the overall $\chi^2$ for the scale improved. Items were assessed for Differential Item Functioning (DIF) by either gender or school. For instance, the item *sew* was removed on the basis of DIF by gender, because males and females with the same overall logit score
Table 6
Reasons for Item Removal and Summary Statistics of the Rasch Analysis after each Item Removal for the Irregular Word List

<table>
<thead>
<tr>
<th>Item</th>
<th>Reason for Removal</th>
<th>Statistics After Item Removal</th>
<th>$\chi^2$</th>
<th>$p$</th>
<th>PSI</th>
<th>EFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>four</td>
<td>Individual item misfit, ($\chi^2(2) = 58.74, p &lt; .001$)</td>
<td></td>
<td>180.15</td>
<td>&lt; .001</td>
<td>.89</td>
<td>2</td>
</tr>
<tr>
<td>we</td>
<td>Individual item misfit, ($\chi^2(2) = 42.19, p &lt; .001$)</td>
<td></td>
<td>132.31</td>
<td>&lt; .001</td>
<td>.89</td>
<td>2</td>
</tr>
<tr>
<td>of</td>
<td>Individual item misfit, ($\chi^2(2) = 21.21, p &lt; .001$)</td>
<td></td>
<td>91.371</td>
<td>.014</td>
<td>.90</td>
<td>2</td>
</tr>
<tr>
<td>excited</td>
<td>DIF: class interval-by-school, $F(2,6) = 4.90, p &lt; .001$</td>
<td></td>
<td>69.578</td>
<td>.238</td>
<td>.90</td>
<td>2</td>
</tr>
<tr>
<td>put</td>
<td>Unidimensionality: PC residuals $t$-test (put, eight, people v who sew use, significant in 40% of simulations) put had highest absolute loading on first PC residual (.55)</td>
<td></td>
<td>64.40</td>
<td>.263</td>
<td>.89</td>
<td>2</td>
</tr>
<tr>
<td>blue</td>
<td>Extreme item (was lower on the logit score than the lowest scoring participant)</td>
<td></td>
<td>62.87</td>
<td>.246</td>
<td>.89</td>
<td>2</td>
</tr>
<tr>
<td>eight</td>
<td>Unidimensionality: EFA 2 factor solution; loading of .89 on second factor</td>
<td></td>
<td>60.70</td>
<td>.247</td>
<td>.89</td>
<td>2</td>
</tr>
<tr>
<td>people</td>
<td>Redundant item (have, use, people, high and who separating lowest four participants)</td>
<td></td>
<td>60.34</td>
<td>.200</td>
<td>.89</td>
<td>2</td>
</tr>
<tr>
<td>have</td>
<td>Redundant item (have, use, high and who separating lowest four participants)</td>
<td></td>
<td>55.70</td>
<td>.269</td>
<td>.89</td>
<td>2</td>
</tr>
<tr>
<td>high</td>
<td>Redundant item (high, use and who separating lowest four participants)</td>
<td></td>
<td>55.15</td>
<td>.223</td>
<td>.88</td>
<td>2</td>
</tr>
<tr>
<td>laugh</td>
<td>Unidimensionality: EFA 2 factor solution; loading of .88 on second factor</td>
<td></td>
<td>54.43</td>
<td>.184</td>
<td>.88</td>
<td>2</td>
</tr>
<tr>
<td>use</td>
<td>Redundant item (use and who separating three participants)</td>
<td></td>
<td>49.29</td>
<td>.270</td>
<td>.88</td>
<td>2</td>
</tr>
<tr>
<td>know</td>
<td>Redundant item (know, boat, juice and done not differentiating any participants)</td>
<td></td>
<td>41.71</td>
<td>.484</td>
<td>.89</td>
<td>2</td>
</tr>
</tbody>
</table>

Continued on page 153.
Table 6
Continued from page 152.

<table>
<thead>
<tr>
<th>boat</th>
<th>Redundant item (boat, juice and done not differentiating any participants)</th>
<th>40</th>
<th>.47</th>
<th>.89</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>done</td>
<td>Redundant Item (juice and done not differentiating any participants)</td>
<td>37.42</td>
<td>0.496</td>
<td>0.89</td>
<td>2</td>
</tr>
<tr>
<td>sew</td>
<td>DIF by Gender F (1,2) = 10.25, p= 0.002</td>
<td>42.41</td>
<td>0.214</td>
<td>0.89</td>
<td>1</td>
</tr>
<tr>
<td>juice</td>
<td>Redundant Item (juice separating only two participants)</td>
<td>41.26</td>
<td>0.183</td>
<td>0.88</td>
<td>1</td>
</tr>
<tr>
<td>want</td>
<td>Redundant Item (want and especially not separating any participants)</td>
<td>31.42</td>
<td>0.496</td>
<td>0.88</td>
<td>1</td>
</tr>
<tr>
<td>frightened</td>
<td>Redundant Item (frightened and towel not separating any participants)</td>
<td>30.9</td>
<td>0.42</td>
<td>0.88</td>
<td>1</td>
</tr>
</tbody>
</table>

N = 172. Note. PSI = Person Separation Index; DIF = Differential Item Functioning

*Number of Factors as dictated by a Exploratory Factor Analysis in conjunction with a Monte Carlo Parallel analysis.

gave significantly different responses for this item, indicating that the item was unduly influenced by gender.

Items were removed on the basis of dimensionality for one of the two reasons. The first was that the item loaded poorly on the first factor in the principal factors analysis and highly on the second factor, as was the case with item eight. Furthermore, it was also one of the highest loading items in a t-test comparing the principal components residuals within the Rasch model. As stated earlier, this means that the residuals from the Rasch model are analysed and those items with the highest positive residual loadings are compared to those with the highest negative loadings through t-test simulations. If these t-tests are significant in more than 5% of simulations, then unidimensionality is not supported and the highest loading items need to be considered for removal.

Some items were removed because they were poorly targeted to the sample and thus extreme. For instance, the item blue was actually lower on the logit scale than the poorest performing participant and could therefore be considered redundant.
Similarly, some items were considered redundant if a large number of items were separating only a small number of participants. For instance, the item *have*, was removed, because it was one of four items separating only four participants. With 19 items removed the remaining 15-item scale had good model fit and was unidimensional. The next section outlines the full Rasch analysis of the final 15-item scale.

**Final irregular word list**

The final word list consisted of 15 items with a possible score range of zero to 15. The mean score was 10.53 (SD = 3.45). Rasch analysis of the scale found that there was good model fit, $\chi^2(30) = 30.9, p = .42$, for this refined scale. The means and standard deviations on the re-scored and refined irregular word list, in order of difficulty, are shown in Table 7.

### Table 7

**Means and Standard Deviations of the Final Irregular Word List**

<table>
<thead>
<tr>
<th>Item</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>who</td>
<td>0.98</td>
<td>0.15</td>
</tr>
<tr>
<td>zero</td>
<td>0.96</td>
<td>0.20</td>
</tr>
<tr>
<td>emu</td>
<td>0.92</td>
<td>0.27</td>
</tr>
<tr>
<td>stomach</td>
<td>0.85</td>
<td>0.36</td>
</tr>
<tr>
<td>towel</td>
<td>0.85</td>
<td>0.35</td>
</tr>
<tr>
<td>especially</td>
<td>0.81</td>
<td>0.39</td>
</tr>
<tr>
<td>characters</td>
<td>0.78</td>
<td>0.41</td>
</tr>
<tr>
<td>jewellery</td>
<td>0.75</td>
<td>0.43</td>
</tr>
<tr>
<td>somersault</td>
<td>0.74</td>
<td>0.44</td>
</tr>
<tr>
<td>rhinoceros</td>
<td>0.72</td>
<td>0.45</td>
</tr>
<tr>
<td>tortoise</td>
<td>0.65</td>
<td>0.48</td>
</tr>
<tr>
<td>certificates</td>
<td>0.60</td>
<td>0.49</td>
</tr>
<tr>
<td>cough</td>
<td>0.60</td>
<td>0.49</td>
</tr>
<tr>
<td>exhausted</td>
<td>0.52</td>
<td>0.50</td>
</tr>
<tr>
<td>cooperation</td>
<td>0.38</td>
<td>0.49</td>
</tr>
</tbody>
</table>

$N = 172$. 
Thresholds

A threshold map for the refined irregular word list was also generated. As can be seen in Figure 11, there were no disordered thresholds within the re-scored scale. The threshold map demonstrates the point on the logit scale where it becomes more likely that a correct answer will be given instead of an incorrect one (the point where the bar changes from blue to red). Consequently, this also means that the range of difficulties of the items is also on display in the threshold map. Figure 6 shows that there is quite a good spread of difficulty throughout the scale. In the refined list, who is the easiest item and cooperation is the most difficult.

Individual item analysis

The item fit analysis was examined in order to ensure that all items fit well to the scale as well as the scale having good item fit overall. The overall item fit residual standard deviation of 0.74 was clearly lower than the upper limit of 1.5 and indicated that overall item fit was good. This was further supported by the individual item fit analysis displayed in Table 8. Not only is the overall item fit for the final irregular word list good, but there were no individual items with high residuals or χ² values to suggest poor item fit.

![Threshold Map for the Final Irregular Word List](image_url)

*Figure 11. Threshold Map for the Final Irregular Word List*
Table 8.

*Item Fit Residuals for the Final Irregular Word List*

<table>
<thead>
<tr>
<th>Item</th>
<th>Fit Residual</th>
<th>$\chi^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>who</td>
<td>-0.62</td>
<td>0.20</td>
<td>.903</td>
</tr>
<tr>
<td>emu</td>
<td>-0.69</td>
<td>0.76</td>
<td>.685</td>
</tr>
<tr>
<td>zero</td>
<td>-0.27</td>
<td>1.14</td>
<td>.566</td>
</tr>
<tr>
<td>cough</td>
<td>-0.74</td>
<td>0.67</td>
<td>.717</td>
</tr>
<tr>
<td>towel</td>
<td>-0.99</td>
<td>2.10</td>
<td>.351</td>
</tr>
<tr>
<td>stomach</td>
<td>0.24</td>
<td>0.20</td>
<td>.906</td>
</tr>
<tr>
<td>tortoise</td>
<td>0.61</td>
<td>1.52</td>
<td>.467</td>
</tr>
<tr>
<td>jewellery</td>
<td>-1.87</td>
<td>2.84</td>
<td>.242</td>
</tr>
<tr>
<td>rhinoceros</td>
<td>1.22</td>
<td>2.26</td>
<td>.323</td>
</tr>
<tr>
<td>certificates</td>
<td>-0.47</td>
<td>4.79</td>
<td>.091</td>
</tr>
<tr>
<td>especially</td>
<td>-1.93</td>
<td>2.51</td>
<td>.285</td>
</tr>
<tr>
<td>exhausted</td>
<td>-0.84</td>
<td>6.61</td>
<td>.037</td>
</tr>
<tr>
<td>characters</td>
<td>-0.72</td>
<td>0.96</td>
<td>.619</td>
</tr>
<tr>
<td>somersault</td>
<td>-0.90</td>
<td>1.29</td>
<td>.525</td>
</tr>
<tr>
<td>cooperation</td>
<td>-0.19</td>
<td>3.06</td>
<td>.217</td>
</tr>
</tbody>
</table>

$N = 172$

Although none of the items had significant item misfit, one item, *exhausted*, would have been significant before a Bonferroni correction. However, the low fit residual of -0.84, well below 2.5, combined with the strong overall model fit indicated that the item was appropriate for the scale. None of the other items had significant fit residuals or $p$ values before or after a Bonferroni correction. Overall, the items in the scale fit well together in the model. To ensure that none of the items were unduly influenced by either gender or school, a DIF analysis was also conducted. None of the items were unduly influenced by either gender or school after a Bonferroni correction was applied.

*Person fit*

Similar to item fit, person fit is measured on two fronts, namely overall person fit and individual person fit. The overall person fit residual standard deviation should be under a 1.5 if there is overall person fit in the scale. The overall person fit...
residual standard deviation for the refined regular word list of 0.45 indicated that the person fit for this analysis was good. Individual person fit residuals should also be under 2.5 for all participants. There were no individuals with person fit residuals over 2.5, indicating that all participants fit the model well. However, 37 students answered all items correctly and one student answered them all incorrectly. These participants were considered extreme in the Rasch analysis. Given the applied ethos in the study, combined with the expectation that high performing students would do very well on this scale, these students were retained in the analysis.

**Targeting and reliability**

The Cronbach’s alpha for the revised scale was good at $\alpha = .88$. The skew in the refined list, displayed in the histogram in Figure 12, was a concern, but it did indicate that the scale was targeted at the population of interest, namely poor readers. Despite the ceiling effect, clearly demonstrated in Figure 12, it can be argued that this scale is doing exactly what is required. Furthermore, the PSI of .88 indicates that this skewness has not affected the test’s propensity to appropriately target the sample. The mean person location for the finalised word list of 1.92, although high,

![Figure 12. Histogram for Total Scores for the Final Irregular Word List](image)  
$N = 172$
is not unreasonable given the aim of the study. The reason for this high mean person location is illustrated in Figure 13, where the items on the bottom portion of the figure, are separating all students quite well, except for the highest performing students. Because of the skew, this scale differentiates between the low performing students quite effectively. Furthermore, the number of students (16%) getting all questions right is not so high that poor irregular word readers would not be separated from their higher performing peers.

Tests of local independence
Dimensionality was tested via a principal axis factor analysis in conjunction with a Monte Carlo parallel analysis, as well as a $t$-test of items identified through the principal component residuals. The KMO measure of sampling adequacy indicated that the data was marvellous for EFA (KMO = .91). Although there were two eigenvalues in the analysis with values over one, a Monte Carlo PCA parallel analysis indicated that a one-factor structure was more appropriate, supporting unidimensionality. A scree plot of the principal factors analysis and randomly generated eigenvalues from the parallel analysis can be seen in Figure 14.

The scree plot shows that the randomly generated eigenvalues are higher than the EFA generated eigenvalues for the second factor, indicating that a one-factor

![Figure 13. Person-Item Interaction for the Final Irregular Word List](image)
structure is best suited to the data. To further the initial EFA analysis for unidimensionality, the principal components residuals were tested through the RUMM program. Accordingly $t$-tests were conducted in order to compare the results of those items with the highest positive residuals (i.e., cooperation, certificates and somersaults), to those items with the highest negative residuals (i.e., tortoise, zero and stomach). These $t$-tests found significant differences in less than 5% of simulations and hence unidimensionality was supported.

**Irregular word list Rasch analysis summary**

The irregular word list was refined from 37 items to 15 items within the guidelines of both CTT and IRT. Analyses were conducted to assess the item thresholds, item fit, person fit, targeting, reliability and dimensionality and on the whole, the list was found to be psychometrically sound. Although there was a ceiling effect in this scale, the scale was not so skewed that the majority of students did not have items separating them, except for the highest performing students who are not of interest in this area of research. With the irregular word list refined the same process was completed on the nonword list. This analysis is detailed in the next section.
**Rasch analysis results for the nonword list**
The process for refining the nonword list was the same as it was for the irregular word list. All analyses were first conducted on the original 36 item list. No decisions about item deletion or rescoring were made on the basis of any one test. Instead, the analyses from all the tests were considered as a whole. In determining item removals, items were removed one by one with analyses repeated after each deletion. The results for the initial analyses are displayed briefly in the next section and are followed by the full analysis on the refined word list in order to demonstrate the psychometric properties of the scale.

**Full word list**
The means, standard deviations and the number of response scoring options for each of the items in the original nonword list are shown in Table 9. Items that did not have a one-point option had only two possible scores. Those that did have the one-point option had three possible scores. Similar to the irregular word list, all analyses were based on the Grade 3 cohort only (N = 172).

Table 9 shows that all of the words provided some item variance as there were correct and incorrect responses for each item within the sample. The mean of the full 36-item list was 50.75 (SD = 16.81), out of a possible 72, indicating that there was some skew. However, the small percentage of participants getting all 36 items correct (5%), suggests that there was not a problem with a ceiling effect. This finding was more important for the nonword list than the irregular word list, as deficits in phonological processing, as measured by nonword reading, will be gauged with reading-age matches rather than chronological-age matches. Therefore, it was important that the scale differentiated between students who read nonwords in an average range, as well as those with limited phonological processing proficiency.

In order to test for unidimensionality, a principal axis factor analysis in conjunction with a Monte Carlo Parallel analysis was conducted. The analysis was deemed appropriate by a KMO of .83, and as can be seen from the scree plot of the parallel analysis in Figure 15, a one-factor structure is best suited to the full word list. This is because there is only one eigenvalue higher than what would be dictated by chance. The full list of nonwords also had very high reliability (α = .94).
Table 9

*Means and Standard Deviations within the Full Nonword List*

<table>
<thead>
<tr>
<th>Item</th>
<th>$M$</th>
<th>$SD$</th>
<th>SO</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>1.75</td>
<td>0.66</td>
<td>2</td>
</tr>
<tr>
<td>ab</td>
<td>1.92</td>
<td>0.37</td>
<td>2</td>
</tr>
<tr>
<td>ek</td>
<td>1.55</td>
<td>0.58</td>
<td>3</td>
</tr>
<tr>
<td>ra</td>
<td>1.81</td>
<td>0.54</td>
<td>3</td>
</tr>
<tr>
<td>zay</td>
<td>1.78</td>
<td>0.59</td>
<td>3</td>
</tr>
<tr>
<td>aip</td>
<td>1.48</td>
<td>0.77</td>
<td>3</td>
</tr>
<tr>
<td>vun</td>
<td>1.73</td>
<td>0.65</td>
<td>2</td>
</tr>
<tr>
<td>huk</td>
<td>1.69</td>
<td>0.71</td>
<td>2</td>
</tr>
<tr>
<td>yex</td>
<td>1.54</td>
<td>0.84</td>
<td>2</td>
</tr>
<tr>
<td>foy</td>
<td>1.55</td>
<td>0.83</td>
<td>2</td>
</tr>
<tr>
<td>oot</td>
<td>1.74</td>
<td>0.63</td>
<td>3</td>
</tr>
<tr>
<td>oji</td>
<td>1.33</td>
<td>0.85</td>
<td>3</td>
</tr>
<tr>
<td>wopa</td>
<td>1.47</td>
<td>0.74</td>
<td>3</td>
</tr>
<tr>
<td>nied</td>
<td>1.22</td>
<td>0.80</td>
<td>3</td>
</tr>
<tr>
<td>jite</td>
<td>1.24</td>
<td>0.81</td>
<td>3</td>
</tr>
<tr>
<td>reng</td>
<td>1.41</td>
<td>0.82</td>
<td>3</td>
</tr>
<tr>
<td>nawk</td>
<td>1.58</td>
<td>0.77</td>
<td>2</td>
</tr>
<tr>
<td>kore</td>
<td>1.47</td>
<td>0.85</td>
<td>2</td>
</tr>
<tr>
<td>crex</td>
<td>1.63</td>
<td>0.74</td>
<td>3</td>
</tr>
<tr>
<td>joid</td>
<td>1.53</td>
<td>0.74</td>
<td>2</td>
</tr>
<tr>
<td>eero</td>
<td>1.44</td>
<td>0.73</td>
<td>3</td>
</tr>
<tr>
<td>kory</td>
<td>1.42</td>
<td>0.83</td>
<td>3</td>
</tr>
<tr>
<td>zatch</td>
<td>1.59</td>
<td>0.79</td>
<td>2</td>
</tr>
<tr>
<td>thunf</td>
<td>1.55</td>
<td>0.83</td>
<td>2</td>
</tr>
<tr>
<td>rilch</td>
<td>1.38</td>
<td>0.86</td>
<td>3</td>
</tr>
<tr>
<td>quelp</td>
<td>1.28</td>
<td>0.86</td>
<td>3</td>
</tr>
<tr>
<td>olloy</td>
<td>1.01</td>
<td>0.72</td>
<td>3</td>
</tr>
<tr>
<td>poslip</td>
<td>1.51</td>
<td>0.85</td>
<td>2</td>
</tr>
<tr>
<td>habay</td>
<td>1.61</td>
<td>0.71</td>
<td>2</td>
</tr>
<tr>
<td>joota</td>
<td>1.56</td>
<td>0.74</td>
<td>3</td>
</tr>
<tr>
<td>frimpler</td>
<td>1.37</td>
<td>0.89</td>
<td>2</td>
</tr>
<tr>
<td>bubode</td>
<td>0.89</td>
<td>0.87</td>
<td>3</td>
</tr>
<tr>
<td>shumba</td>
<td>1.53</td>
<td>0.79</td>
<td>2</td>
</tr>
<tr>
<td>geasure</td>
<td>0.91</td>
<td>0.94</td>
<td>3</td>
</tr>
<tr>
<td>lantople</td>
<td>1.20</td>
<td>0.91</td>
<td>2</td>
</tr>
<tr>
<td>strimpex</td>
<td>1.33</td>
<td>0.91</td>
<td>2</td>
</tr>
</tbody>
</table>

$N = 172$. *Note.* SO = Number of Scoring Options.
To further investigate the psychometric properties of the scale, a Rasch analysis was conducted on the full 36 items. Model fit for the full scale was poor, $\chi^2(72) = 168, p < .001$, although the PSI was very high at .91. Similar to the irregular word list, the three-level scoring system was an issue, albeit slightly more successful in this list as only nine out of fourteen of the items with the three-level scoring system had disordered thresholds. The threshold map for the original nonword list can be seen in Figure 16.

Examination of this issue revealed that the disordered nonword items with the three-level scoring scale had exactly the same issues as the corresponding items in the irregular word list. Ordered thresholds could be found in five of the items with one-point options, *ab, wopa, nied, jite* and *eero*. In each of these items, the most probable response item selected within a given logit score range is the one-point option. This is not the case in the nine items with disordered thresholds. The category probability curve for the item *eero* can be seen in Figure 17.

*Figure 15. Scree Plot Comparing Eigenvalues from the Principal Axis Factor Analysis of the Full Nonword List and the Eigenvalues Randomly Generated in the Parallel Analysis*
Figure 16. Threshold Map for the Full Nonword List

Note. $N = 172.$

** = disordered threshold
The one-point option for *eero*, seen in Figure 17, is the most likely response for a very small range of the overall person logit scores. As such, little will be lost by removing this scoring option, even for the five items in which the three-level scoring system is appropriate. Given that more than half of the items with this scoring system had disordered thresholds and the lack of success with the same system in the irregular word list, this scoring method was again discarded and a simple correct or incorrect scoring system was adopted instead. Consequently, all items were re-scored to give one-point for a correct answer or zero points for an incorrect or partially correct answer.

**Item removal**

With the scoring system for the scale adjusted, the original nonword list was refined in a manner similar to that of the irregular word list. Even though skew was not the same issue that it was in the irregular word list, it was still important to take skew into consideration when selecting items for deletion. Dimensionality and model fit are also important considerations. Therefore, items were removed on the basis of misfit, dimensionality, skew, thresholds and redundancy. Because it is always possible that items that do not fit well in the unrefined list, will fit into a more refined list, items were removed one by one and all analyses repeated after each deletion. The reasons for each deletion are shown in Table 10, resulting in a final list of 20 items for the nonword list.

*Figure 17. Category Probability Curve for the Item eero*
Table 10

Reasons for Item Removal and Summary Statistics of the Rasch Analysis after each Item Removal for the Nonword List

<table>
<thead>
<tr>
<th>Word Removed</th>
<th>Reason for Removal</th>
<th>$\chi^2$</th>
<th>$p$</th>
<th>PSI</th>
<th>Factors$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original re-scored list</td>
<td></td>
<td>122.62</td>
<td>&lt;.001</td>
<td>.89</td>
<td>1</td>
</tr>
<tr>
<td>eero</td>
<td>Poor item fit ($\chi^2 (2) = 13.78$, $p = .001$)</td>
<td>107.21</td>
<td>.003</td>
<td>.90</td>
<td>1</td>
</tr>
<tr>
<td>nied</td>
<td>Poor item fit ($\chi^2 (2) = 14.04$, $p = .001$)</td>
<td>102.91</td>
<td>.004</td>
<td>.90</td>
<td>1</td>
</tr>
<tr>
<td>huk</td>
<td>Poor item fit ($\chi^2 (2) = 12.87$, $p = .002$)</td>
<td>89.71</td>
<td>.028</td>
<td>.90</td>
<td>1</td>
</tr>
<tr>
<td>ek</td>
<td>Poor item fit ($\chi^2 (2) = 14.10$, $p = .001$)</td>
<td>66.63</td>
<td>.387</td>
<td>.90</td>
<td>1</td>
</tr>
<tr>
<td>wopa</td>
<td>Redundant item (kory, strimpex, rilch, reng and wopa separating 6 participants) and poor item fit ($\chi^2 (2) = 8.48$, $p = .014$)</td>
<td>53.85</td>
<td>.760</td>
<td>.90</td>
<td>1</td>
</tr>
<tr>
<td>rilch</td>
<td>Redundant item (strimpex, rilch, and reng separating 8 participants)</td>
<td>54.53</td>
<td>.675</td>
<td>.89</td>
<td>1</td>
</tr>
<tr>
<td>kory</td>
<td>Poor item fit ($\chi^2 (2) = 6.04$, $p = .048$)</td>
<td>46.26</td>
<td>.867</td>
<td>.89</td>
<td>1</td>
</tr>
<tr>
<td>crex</td>
<td>Poor item fit ($\chi^2 (2) = 5.38$, $p = .068$)</td>
<td>41.39</td>
<td>.928</td>
<td>.89</td>
<td>1</td>
</tr>
<tr>
<td>thunf</td>
<td>Redundant item (zatch and thunf, yex, foy separating 2 participants)</td>
<td>34.32</td>
<td>.983</td>
<td>.88</td>
<td>1</td>
</tr>
<tr>
<td>shumba</td>
<td>Redundant item (nawk, habay, poslip and shumba separating 4 participants)</td>
<td>33.36</td>
<td>.979</td>
<td>.88</td>
<td>1</td>
</tr>
<tr>
<td>nawk</td>
<td>Redundant item (zatch, yex, nawk and foy separating 4 participants)</td>
<td>37.79</td>
<td>.898</td>
<td>.87</td>
<td>1</td>
</tr>
<tr>
<td>ma</td>
<td>Redundant item (ma and zay separating 4 participants)</td>
<td>42.10</td>
<td>.712</td>
<td>.87</td>
<td>1</td>
</tr>
<tr>
<td>oot</td>
<td>Redundant item (oot and vun separating 6 participants)</td>
<td>39.26</td>
<td>.748</td>
<td>.87</td>
<td>1</td>
</tr>
<tr>
<td>joota</td>
<td>Poor item fit ($\chi^2 (2) = 5.78$, $p = .06$)</td>
<td>29.72</td>
<td>.951</td>
<td>.87</td>
<td>1</td>
</tr>
<tr>
<td>kore</td>
<td>Redundant item (kore and poslip separating 3 participants)</td>
<td>26.49</td>
<td>.970</td>
<td>.86</td>
<td>1</td>
</tr>
<tr>
<td>frimpler</td>
<td>Redundant item (aip and frimpler separating no participants)</td>
<td>35.49</td>
<td>.674</td>
<td>.86</td>
<td>1</td>
</tr>
</tbody>
</table>

$N = 172$. Note. PSI = Person Separation Index

$^a$ Number of Factors as dictated by a Principal Factors Analysis in conjunction with a Monte Carlo Parallel analysis.
Items were deleted on the basis of poor item fit if individual $\chi^2$ scores indicated that items were not fitting well in to the scale. Because this particular list fit together very well there were some items removed on the basis of poor item fit, even if they were not significant after a Bonferroni correction. Therefore, their removal was not necessary but it did improve the item fit of the scale overall. In the case of redundant items, there were ranges on the logit scale where there were a similar number of items and participants, indicating that not all of these items were required to separate the skill levels of students. No items in this list obtained significant results from a DIF analysis, indicating that all items were a similar level of difficulty to all students per level of proficiency, regardless of school or gender. Because there was only one-factor in the full list, no items were removed in order to achieve unidimensionality for the scale.

These item deletions resulted in a final nonword list of 20 items. This list was longer than the irregular list as it was important that it could separate the poorest readers, and also the poorest nonword readers who were reading at an age-appropriate level. Therefore, the nonword list needed to separate students who were in the low average range as well as those who were at the very low range of nonword reading.

**Final nonword list**

The final word list consisted of 20 items with a scoring range of zero to 20. The mean score for the scale was 13.09 (SD = 4.97). Rasch analysis on this final 20 item list indicated very good model fit, $\chi^2(40) = 35.49, p = .674$. The re-scored means and standard deviations of the items, in order of difficulty, are shown in Table 11.

**Thresholds**

The threshold map for the refined nonword list is shown in Figure 18. There is a wide range of difficulty and no disordered thresholds within the re-scored scale for the twenty remaining items. As noted earlier, the threshold map can also give an idea of item difficulty, the higher on the logit score that the most probable response on a given item is zero, the more difficult the item. The item *olloy* was the most difficult for the sample and *ab* was the easiest.
Table 11
*Means and Standard Deviations of the Final Nonword List*

<table>
<thead>
<tr>
<th>Item</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ab</td>
<td>0.97</td>
<td>0.18</td>
</tr>
<tr>
<td>ra</td>
<td>0.88</td>
<td>0.33</td>
</tr>
<tr>
<td>zay</td>
<td>0.87</td>
<td>0.34</td>
</tr>
<tr>
<td>vun</td>
<td>0.84</td>
<td>0.37</td>
</tr>
<tr>
<td>yex</td>
<td>0.77</td>
<td>0.42</td>
</tr>
<tr>
<td>zatch</td>
<td>0.77</td>
<td>0.42</td>
</tr>
<tr>
<td>foy</td>
<td>0.76</td>
<td>0.43</td>
</tr>
<tr>
<td>habay</td>
<td>0.74</td>
<td>0.44</td>
</tr>
<tr>
<td>poslip</td>
<td>0.73</td>
<td>0.44</td>
</tr>
<tr>
<td>joid</td>
<td>0.68</td>
<td>0.47</td>
</tr>
<tr>
<td>aip</td>
<td>0.65</td>
<td>0.48</td>
</tr>
<tr>
<td>strimpex</td>
<td>0.63</td>
<td>0.49</td>
</tr>
<tr>
<td>reng</td>
<td>0.62</td>
<td>0.49</td>
</tr>
<tr>
<td>oji</td>
<td>0.58</td>
<td>0.50</td>
</tr>
<tr>
<td>quelp</td>
<td>0.55</td>
<td>0.50</td>
</tr>
<tr>
<td>lantople</td>
<td>0.52</td>
<td>0.50</td>
</tr>
<tr>
<td>geasure</td>
<td>0.48</td>
<td>0.50</td>
</tr>
<tr>
<td>jite</td>
<td>0.47</td>
<td>0.50</td>
</tr>
<tr>
<td>bupode</td>
<td>0.33</td>
<td>0.47</td>
</tr>
<tr>
<td>olloy</td>
<td>0.26</td>
<td>0.44</td>
</tr>
</tbody>
</table>

_N = 172._

*Individual item analysis*

Item fit analysis was examined in order to ensure that the 20 remaining items fitted well to the scale. An overall item fit residual standard deviation of 0.59, lower than the prescribed magnitude limit of 1.5, indicated good overall item fit within the scale. Individual item fit was also analysed, with the results detailed in Table 12. As noted earlier, it is important that the \( \chi^2 \) values are not significant and that fit residuals are not higher than a magnitude of 2.5.
As can be seen in Table 12, none of the items had significant item fit issues. Closer analysis at the item level indicated that none of the remaining items had problems with misfit as measured by fit residuals or chi-square probability. It is worth noting that some of these items, such as strimpex, had a significant $\chi^2$ value in the full list but had good model fit once inappropriate items had been removed. This highlights the importance of removing items iteratively.

To further the analysis at the individual item level, DIF analysis was conducted in order to ensure that none of the items were unduly influenced by either gender or school. None of the items showed significant variation as a function of either of these variables after Bonferroni corrections were applied. Therefore, it can be assumed that students of different gender or schools of the same proficiency level were answering the items in similar ways.
Table 12

*Item Fit Analysis for the Final Nonword List*

<table>
<thead>
<tr>
<th>Item</th>
<th>Fit Residual</th>
<th>$\chi^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ab</td>
<td>-0.85</td>
<td>1.00</td>
<td>.607</td>
</tr>
<tr>
<td>ra</td>
<td>-0.50</td>
<td>0.72</td>
<td>.698</td>
</tr>
<tr>
<td>zay</td>
<td>-0.51</td>
<td>1.68</td>
<td>.432</td>
</tr>
<tr>
<td>aip</td>
<td>1.41</td>
<td>1.18</td>
<td>.554</td>
</tr>
<tr>
<td>vun</td>
<td>0.52</td>
<td>1.45</td>
<td>.485</td>
</tr>
<tr>
<td>yex</td>
<td>-0.71</td>
<td>0.56</td>
<td>.758</td>
</tr>
<tr>
<td>foy</td>
<td>-0.11</td>
<td>0.87</td>
<td>.647</td>
</tr>
<tr>
<td>oji</td>
<td>-0.10</td>
<td>3.14</td>
<td>.208</td>
</tr>
<tr>
<td>jite</td>
<td>-0.22</td>
<td>2.72</td>
<td>.256</td>
</tr>
<tr>
<td>reng</td>
<td>-0.32</td>
<td>0.56</td>
<td>.754</td>
</tr>
<tr>
<td>joid</td>
<td>-1.04</td>
<td>1.41</td>
<td>.495</td>
</tr>
<tr>
<td>zatch</td>
<td>-1.01</td>
<td>3.49</td>
<td>.175</td>
</tr>
<tr>
<td>quelp</td>
<td>-0.12</td>
<td>1.65</td>
<td>.437</td>
</tr>
<tr>
<td>olloy</td>
<td>0.23</td>
<td>0.93</td>
<td>.628</td>
</tr>
<tr>
<td>poslip</td>
<td>0.39</td>
<td>0.31</td>
<td>.859</td>
</tr>
<tr>
<td>habay</td>
<td>0.56</td>
<td>0.48</td>
<td>.788</td>
</tr>
<tr>
<td>bupode</td>
<td>-0.33</td>
<td>2.61</td>
<td>.271</td>
</tr>
<tr>
<td>geasure</td>
<td>0.00</td>
<td>0.29</td>
<td>.865</td>
</tr>
<tr>
<td>lantople</td>
<td>-0.23</td>
<td>1.58</td>
<td>.454</td>
</tr>
<tr>
<td>strimpex</td>
<td>-0.34</td>
<td>0.89</td>
<td>.640</td>
</tr>
</tbody>
</table>

*Note.* $N = 172.$

*Person fit*

So as to ensure that all of the participants fit well in the Rasch model, the person fit statistics were analysed. An overall person fit residual standard deviation of 0.59, lower than the limit of 1.5, indicated good person fit within the scale. Closer analysis showed that although there were no participants with misfit, six participants got all of the items in the scale correct and as such were considered extreme within the Rasch model. As noted earlier, these extreme participants were anticipated, and in keeping with the applied ethos of the study, were retained.
**Targeting and reliability**

The refined nonword scale has high reliability ($\alpha = .89$) and the PSI is good at .85. In order to investigate how appropriately targeted the scale is to the population, a histogram was generated and can be seen in Figure 19. The nonword scale has much less skew than irregular word list. The mean person value location of 1.05, where zero represents no skew, indicated that the scale was well targeted given the reason for its development. This was further supported by the small number of students correctly reading all items, only 6%.

The targeting of the scale can be seen more clearly in the person-item interaction graph displayed in Figure 20. Once again there were few items targeted at the highest performers and the majority of items were targeted towards low to average performers.

*Figure 19. Histogram of Total Scores for the Final Nonword List*

$N = 172$
Tests of local independence

In order to test unidimensionality, the scale was first deemed appropriate for a principal axis factor analysis (KMO = .87) and the eigenvalues from that analysis in conjunction with a Monte Carlo Parallel Analysis indicated that the factor structure was still unidimensional. The scree plot from this parallel analysis can be seen in Figure 21. Only one of the nonword eigenvalues is higher than would be dictated by chance, thus supporting unidimensionality for the scale.

To further examine dimensionality, the principal components residuals from the Rasch model were examined and the items with the highest positive residuals (i.e., vun, reng, and olloy) were compared to the items with the highest negative residuals (i.e., aip, jobd, and poslip) through t-test simulations. None of these were significant in any simulations, and hence unidimensionality was further supported.

Nonword list Rasch analysis summary

This scale was refined from a 36-item list to a 20-item list, with items removed on the basis of poor item fit or redundancy. This list of nonwords was psychometrically sound by both the tenets of CTT and IRT. Although there was still skew, there was no ceiling effect and the reliability of the scale is tentatively supported.
Furthermore, the 20-item nonword list was unidimensional and not unduly affected by gender or school. The scale has good separation between participants of different proficiency levels and could be used in an applied setting to assess the phonological processing abilities of Grade Three students.

**Overall Summary**

This chapter detailed the results of the Rasch analysis, performed on the irregular and nonword list. Both of these lists, despite the heavy skew in the irregular word list, were psychometrically validated as per the tenets of IRT. However, because of the skew in the irregular word list, any analyses involving this variable will need to be interpreted with caution. In Chapter 6, the utility of cognitive processing tests, including the two lists examined in this chapter, to separate LD and non LD students will be examined.

---

*Figure 21. Scree Plot Comparing Eigenvalues from the Principal Axis Factor Analysis of the Full Nonword List and the Eigenvalues Randomly Generated in the Parallel Analysis*
Chapter 6

Overview
In Chapter 5, the irregular and nonword lists generated by the researcher were refined and preliminary support for their psychometric validation was provided. The aim of this chapter is to explore how the cognitive processes of phonological processing, naming speed and verbal memory relate to reading ability and, more specifically, Learning Disabilities (LD). Furthermore, the role of orthographic processing will also be examined, as will the manner in which these processes relate to spelling deficits. It is hoped that through this analysis the guidelines for a screening battery designed to identify possible LD will be developed.

The results in this chapter will be presented in five sections. In the first section, the descriptive statistics of all tests at all grade-levels will be examined. As well as the nonword and irregular word lists, Rapid Automatic Naming (RAN) tests were developed to measure naming speed, and digit span tests were developed to measure verbal memory. Reading and spelling ability were tested through the Progressive Achievement Test in Reading (PAT-R) and the South Australian Spelling Test (SAST) respectively. Non-verbal intelligence was tested with Raven’s Progressive Matrices (RPM). The homogeneity of the Grade 3 sample will also be tested through independent samples \( t \)-tests and variance components analysis. Finally, the reliability of the cognitive processing tests will be examined, as will the correlation matrix for all of the variables, in order to begin to explore the relationships between the variables.

The second section in this chapter establishes the links between reading and cognitive processing and demonstrates that those who are low in these processes are also, generally speaking, poor readers. Although a distinction needs to be made between poor readers and students with LD, it is still important to establish which tests are best related to reading at any given age group. Consequently, demonstrating that these cognitive processes are correlated with reading ability is an important first step in this area of research, but is only that, a step. The method of doing this in the current study will be the use of cluster analysis, used in order to ascertain if groups of students with different levels of reading have corresponding abilities in the cognitive processes selected. Furthermore the relationship between
these cognitive processes and spelling will also be examined, so the same process will be repeated with the SAST in the place of the PAT-R.

The third section in this chapter examines what cognitive processes are aligned with the presence of LD as per the discrepancy method of identification. The sample will be split into those designated LD, the Low Achievement (LA) group who have poor reading skills without a significant discrepancy and the rest of the cohort who are designated as the Regular Achievement (RA) group. In order to generalise these results to those of previous studies, this analysis will be done with both traditional discrepancy methods and regression-adjusted discrepancy methods. The cognitive processes of these three groups, as dictated by the two discrepancy methods will also be examined using one-way Analysis of Variance (ANOVA).

The results of the Discriminant Function Analyses (DFA) will be reported in the fourth section. In the first DFA, the traditional discrepancy-defined LD, LA and RA groupings will be predicted by nonword reading, RAN and digit span tests. In the second DFA, the LD, LA and RA groupings as per the regression-adjusted discrepancy method will be predicted by the same five tests. Using these DFAs, groups, as predicted by the five cognitive processing tests, will be determined. With the groups allocated as per these processes, one-way ANOVAs will be conducted in order to ascertain if the differences between the groups are consistent, as compared to when the allocation is dictated by discrepancy alone.

In the final section of this chapter, students who would be flagged as possibly having LD via the multiple deficits model will be identified. As discussed in Chapter 3, this battery of tests is a screening rather than diagnostic tool. As such, the identification process will be fairly liberal in order to reduce the presence of false negatives within the screening tool. Furthermore, as previously discussed, the method of identification will be based on absolute deficits, not regression based equations, as it has been shown that the correlation between LD and these processes is weakened as ability increases (Neuhaus et al., 2006). Descriptive statistics for these groups as dictated by a cognitive processing deficit framework will then be generated and examined.
Method

In order to collect the data for this section of the study, 163 Grade 3 students were recruited from five different schools and 58 Grade 1 and 57 Grade 2 students were also recruited from one of these schools. The Grade 3 cohort completed three group-administered tests, namely the Raven’s Progressive Matrices (RPM), the Progressive Achievement Testing in Reading (PAT-R) and the South Australian Spelling Test (SAST). These students also completed six individually administered tests, namely irregular and nonword reading, letter and colour RAN, and digit span forwards and backwards. The Grade 1 and 2 students completed the same six individual tests and a test of reading ability. In the case of the Grade 2 students, the reading test was also the PAT-R whereas in the case of the Grade 1 students, it was a teacher administered measure of reading proficiency, outlined in Chapter 4.

As discussed in Chapter 4, the Grade 3 sample size in Chapter 5 ($N = 172$), is reduced in the analyses in this chapter ($N = 163$). This is due to the fact that all participants completed the word lists, the only variables analysed in Chapter 5. However, nine participants missed one of the RPM, PAT-R or SAST tests and, as such, could not be included in the analyses in this chapter. There is no other missing data within the entire sample, so no missing value analyses were required. Grade 1 and 2 data will also be utilised in this chapter, but only as a basis for developing reading-age norms.

Summary Statistics

In this section the descriptive statistics of all the tests over all grade-levels will be displayed. The reliability of the newly-generated word lists will also be examined, however as the digit span and RAN tests were single-item measures, internal consistency was only calculated for the word lists. Furthermore, variance components analysis and $t$-tests will be utilised to demonstrate that the Grade 3 data are suitable to be analysed as independent data and a correlation matrix will be used to examine the relationships between the variables tested.

Descriptive statistics

The descriptive statistics of all the measures completed by each grade-level are shown in Table 13. Note that there are minor variations in the summary statistics between the final word lists in Chapter 5 and the statistics presented here, due to the
Table 13

**Means, Standard Deviations, Skew and Kurtosis for all Grades, on all Tests**

<table>
<thead>
<tr>
<th>Test</th>
<th>Grade</th>
<th>M</th>
<th>SD</th>
<th>Skew(^1)</th>
<th>Kurtosis(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irregular</td>
<td>1</td>
<td>3.28</td>
<td>3.82</td>
<td>1.54</td>
<td>1.82</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>8.79</td>
<td>4.45</td>
<td>-0.25</td>
<td>-1.22</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>10.42</td>
<td>3.51</td>
<td>-0.95</td>
<td>0.09</td>
</tr>
<tr>
<td>Nonword</td>
<td>1</td>
<td>7.53</td>
<td>5.08</td>
<td>0.42</td>
<td>-1.09</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11.51</td>
<td>4.86</td>
<td>-0.59</td>
<td>-0.44</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>12.98</td>
<td>5.00</td>
<td>-0.62</td>
<td>-0.65</td>
</tr>
<tr>
<td>CRAN</td>
<td>1</td>
<td>43.84</td>
<td>7.57</td>
<td>0.40</td>
<td>-0.55</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>42.40</td>
<td>8.74</td>
<td>0.52</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>38.09</td>
<td>7.60</td>
<td>0.53</td>
<td>0.33</td>
</tr>
<tr>
<td>LRAN</td>
<td>1</td>
<td>34.00</td>
<td>6.54</td>
<td>0.72</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>28.44</td>
<td>5.42</td>
<td>0.68</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>26.99</td>
<td>6.08</td>
<td>0.53</td>
<td>-0.11</td>
</tr>
<tr>
<td>DSF</td>
<td>1</td>
<td>4.55</td>
<td>0.60</td>
<td>0.56</td>
<td>-0.57</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5.18</td>
<td>0.95</td>
<td>0.16</td>
<td>-1.05</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5.34</td>
<td>0.91</td>
<td>0.17</td>
<td>-0.28</td>
</tr>
<tr>
<td>DSB</td>
<td>1</td>
<td>3.26</td>
<td>0.64</td>
<td>0.57</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.49</td>
<td>0.78</td>
<td>1.19</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3.71</td>
<td>0.90</td>
<td>0.93</td>
<td>1.24</td>
</tr>
<tr>
<td>RPM</td>
<td>3</td>
<td>36.86</td>
<td>8.52</td>
<td>-0.07</td>
<td>0.73</td>
</tr>
<tr>
<td>RL</td>
<td>1</td>
<td>15.19</td>
<td>5.40</td>
<td>0.20</td>
<td>-0.65</td>
</tr>
<tr>
<td>PATR</td>
<td>2</td>
<td>54.85</td>
<td>24.45</td>
<td>-0.26</td>
<td>-0.93</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>53.00</td>
<td>26.32</td>
<td>-0.20</td>
<td>-0.96</td>
</tr>
<tr>
<td>SAST</td>
<td>3</td>
<td>9.10</td>
<td>1.64</td>
<td>-0.04</td>
<td>0.20</td>
</tr>
</tbody>
</table>

*Note.  N = 278 (Grade 3 = 163, Grade 2 = 57; Grade 1 = 58). Irregular = Irregular word list; Nonword = Nonword list; CRAN = Rapid Automatic Naming Colour; LRAN = Rapid Automatic Naming Letter; DSF = Digit Span Forward; DSB = Digit Span Backward; RPM = Raven’s Progressive Matrices; RL = Reading Level; PAT-R = Progressive Achievement Test in Reading; SAST = South Australian Spelling Test.*

\(^1\) Skewness standard error: Grade 3 = 0.19, Grade 2 = 0.32, Grade 1 = 0.31; \(^2\) Kurtosis standard error: Grade 3 = 0.38, Grade 2 = 0.62, Grade 1 = 0.62.
removal of nine participants with missing data. Please note that the reading level test for Grade 1 students was in place of the PAT-R which was deemed inappropriate for such a young cohort. There were consistently improving performances on all tests over the three grade-levels, shown in Table 13. This includes both RAN tests, which were timed. Consequently, lower scores indicate better performance. The only Grade 1 and 2 data that is used in this particular study is that reading comprehension and nonword reading, in order to develop reading-age comparisons for the Grade 3 cohort. However, these students were tested on all variables for recruitment and reporting purposes. As discussed in Chapter 4, the reading test used for the Grade 2 and 3 cohorts was not suitable for the Grade 1 students, as such a measure of reading proficiency already in place was utilised.

In order to be able to measure the relative importance of the different variables, all scores were converted to z-scores for subsequent analyses. Z-scores were calculated separately for each year level so that a student gaining the mean score for their year level will receive a z-score of zero, regardless of what the mean was for the entire sample. Furthermore, reading ability will now be treated as one variable for ease of presentation. For Grade 2 and 3 this score will be the z-score ascertained by the PAT-R, but for Grade 1 this score will be the z-score derived from the reading levels as discussed in Chapter 4. For ease of interpretation, the two RAN tests z-scores were multiplied by -1 so that high scores indicate strength in that test.

Internal consistency
The reliability of tests used for diagnostic purposes is extremely important. Even in a screening tool where the increased Type II error can provide some leeway, it is vital that all measures are reliable. However, the reliability of the RAN and digit span cannot be calculated as they do not comprise multiple questions. The word lists, however, were analysed with Cronbach’s alpha as a measure of internal consistency. Please note that there may be small variations from the same analysis in Chapter 5 due to the adjusted sample size.

The internal consistency for each of the word lists was high with a Cronbach’s alpha range of .88 and .93. The internal consistency of the lists at the targeted Grade 3 level is high with the irregular word list at .88 and nonword list at .89. The word lists were more reliable in the smaller samples from the younger grades. The
irregular word list had a Cronbach’s $\alpha$ of .92 and .91 in Grade 1 and 2 respectively. The nonword list had a Cronbach’s $\alpha$ of .90 and .86 in Grade 1 and 2 respectively.

**Test Re-Test Reliability**
Not only is internal consistency important in test development, it is also necessary that students taking tests will get similar scores at different sittings of the same test. Given the established test-retest reliability of the digit span, RPM, PAT-R and SAST, only the test-retest reliabilities of the two word lists and the two RAN tests were measured. Thirty students were retested on these four tests one month after the original testing. The test-retest correlations for the irregular word list, $r(29) = .88$, $p < .001$, nonword list, $r(29) = .86$, $p < .001$, colour RAN, $r(29) = .87$, $p < .001$, and letter RAN, $r(29) = .83$, $p < .001$, were all above .80. This indicated that the tests were stable and that they were suitable for use in an applied setting.

**Homogeneity of the Grade 3 sample**
Although the Grade 1 and 2 samples completed their tests in a consistent manner, there were some variations in administration throughout the Grade 3 cohort. In some schools, the PAT-R and/or SAST were administered by teachers in some classes earlier in the year. In order to ensure that the timing of the tests did not make a significant difference in the analyses independent $t$-tests were conducted to examine the mean differences between groups of students with differences in administration. The mean PAT-R score from students tested in first semester ($M = 50.98$, $SD = 23.99$) was not significantly different to the scores of students tested in second semester ($M = 53.66$, $SD = 27.09$), $t(161) = -0.56$, $p > .05$. In some schools the teacher had already administered these tests, and of those schools that had not administered these tests, some teachers preferred to administer the test themselves. There was no significant difference in PAT-R scores from tests administered by the teachers ($M = 51.80$, $SD = 24.86$) and the researcher ($M = 55.29$, $SD = 29.01$), $t(161) = -0.80$, $p > .05$. The SAST was only administered by the researcher in one school in the second semester ($M = 8.58$, $SD = 1.72$). Again, there was also no significant difference in the mean scores, $t (161) = 1.61$, $p > .05$, when compared to SAST mean scores from students that had the test administered by their teachers in first semester ($M = 9.20$, $SD = 1.72$).
Finally, a variance components analysis was conducted in SPSS with maximum likelihood estimation to ensure that membership in either school or classroom was not unduly influencing any of the tests administered. In this analysis, all of the tests administered to the Grade 3 cohort were the dependent variables and class membership was the independent variable. With the exception of the PAT-R, 0 to 1.7% of the variance in each test could be explained by class membership. However, 5.4% of the variance in the PAT-R could be explained by class membership. Given that only one of the variables was marginally above the widely used 5% cut-off, the Grade 3 sample was tested as a single data set.

**Assumptions of normality**

Data was examined to ensure that the assumptions of normality made as part of the analyses conducted in the current study were met. The word lists were positively skewed for the Grade 1 cohort, as would be expected on a test targeted to a Grade 3 sample. However, given that only the average scores on the nonword list for those of average reading-age will be used, transformations were not performed.

Furthermore, there was an issue with negative skew in both of the word lists within the Grade 3 sample. Once again this was anticipated given the aim of targeting poor readers in the current study. As discussed in Chapter 5, there was a ceiling effect in the irregular word list as a result of targeting poor-performing students. This will result in a reduced correlation with other variables and therefore the cluster analysis needs to be interpreted with caution. If orthographic processing had theoretical justification as a predictor of LD, then logistic regression rather than a DFA would have been used. However, as irregular word reading is not used in this analysis, the DFA is justified.

The linearity of the variables via scatter plots were examined. Scatter plots of all variables were paired with the PAT-R or the SAST. All of these were acceptable with the exception of irregular word reading, which had more variance at the higher-performing level than for low-performing students in the PAT-R and SAST. This could be a reflection of either the ceiling effect in the irregular word list or provide support for the idea that some tests are only accurate predictors of reading for poor performers. Regardless, it is important that all tests involving the irregular word lists are interpreted carefully and Levene’s test of homogeneity of variance will also be examined on all variables before ANOVAs are conducted on the data.
Checking for univariate outliers was facilitated by the fact that all tests had already been converted to $z$-scores. One score on the digit span backwards was higher than an absolute value of 3.29 ($p < .001$). All analyses were run with this value removed and included. As there was minimal difference in overall results, it was decided that the outlier was not unduly affecting results so the participant was retained. Mahalanobi's distances were calculated in order to check for multivariate outliers. Once again all analyses were conducted with and without the two outliers found and because the outliers were not influential, they were retained.

**Correlation matrix of all Grade 3 variables**

There were high levels of correlations between most of the tested variables, this can be seen in Table 14. With the exception of the correlation between letter RAN and the RPM, all correlations between variables were significant at an alpha level of .05. Despite heteroscedasticity, irregular word reading had a higher correlation with both the PAT-R and SAST than nonword reading. Letter RAN had higher correlations with both the PAT-R and SAST than colour RAN. Digit span forward was the better measure of the two verbal memory measures in terms of correlation with the PAT-R and SAST.

The correlations between the PAT-R and both digit span tests were fairly low. The relation between digit span backwards and reading comprehension was especially low at .19. Furthermore, the correlation between nonword reading and RAN was also not excessively high. This is important because, as discussed in Chapter 2, one of the primary criticisms of the double deficit theory is that the two tests are actually assessing the same construct (Schatzschneider et al., 2002) or that the high correlation between the two can confound results (Vellutino et al., 2004). The correlation between colour RAN and nonword reading was quite low at .28 and the correlation between Letter RAN and nonword reading was moderate at .50. Neither was high enough to support the theory that nonword reading and RAN were actually testing the same processes. The only non-significant correlation in the matrix was between letter RAN and the RPM. Given that one is a measure of non-verbal intelligence and the other a measure of letter naming automaticity, this is not entirely surprising or concerning.
Table 14
Correlations for Grade 3 Participants on all Reading and Cognitive Processing Tests

<table>
<thead>
<tr>
<th></th>
<th>Irregular</th>
<th>Nonword</th>
<th>CRAN</th>
<th>LRAN</th>
<th>DSF</th>
<th>DSB</th>
<th>RPM</th>
<th>PAT-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irregular</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonword</td>
<td>.74***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRAN</td>
<td>.30***</td>
<td>.28***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRAN</td>
<td>.44***</td>
<td>.50***</td>
<td>.54***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSF</td>
<td>.29***</td>
<td>.29***</td>
<td>.16*</td>
<td>.30***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSB</td>
<td>.25**</td>
<td>.28***</td>
<td>.23**</td>
<td>.24**</td>
<td>.37***</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPM</td>
<td>.30***</td>
<td>.24**</td>
<td>.19*</td>
<td>0.10</td>
<td>.19*</td>
<td>.20*</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>PAT-R</td>
<td>.69***</td>
<td>.53***</td>
<td>.25**</td>
<td>.42***</td>
<td>.30***</td>
<td>.19*</td>
<td>.42***</td>
<td>1</td>
</tr>
<tr>
<td>SAST</td>
<td>.76***</td>
<td>.72***</td>
<td>.35***</td>
<td>.53***</td>
<td>.29***</td>
<td>.22**</td>
<td>.21**</td>
<td>.62***</td>
</tr>
</tbody>
</table>

N = 163. Note. Irregular = Irregular word reading; Nonword = Nonword reading; CRAN = Rapid Automatic Naming Colour; LRAN = Rapid Automatic Naming Letter; DSF = Digit Span Forward, DSB = Digit Span Backward; RPM = Raven’s Progressive Matrices; PAT-R = Progressive Achievement Test in Reading; SAST = South Australian Spelling Test.

*** = p < .001; ** = p < .01; * = p < .05.

Of particular interest was the correlation between the RPM and the PAT-R and SAST. Although the correlation between the PAT-R and the RPM was lower than would be expected at .42, it was twice as high as the correlation between the SAST and the RPM, which was low at .21. These two correlations are important as the correlation between aptitude (as measured by the RPM) and achievement (as measured by the PAT-R for reading and the SAST for spelling) is the basis of the discrepancy method of LD identification. When calculating the regression-adjusted discrepancy cut-offs, these correlations indicate the gradient of the line used to separate LD from non-LD students. The correlation between the PAT-R and the RPM is only moderate, but acceptable for use in a discrepancy-based analysis. However, the correlation between spelling and aptitude is low by any standard and, as such, may not provide a suitable basis for comparison. This low correlation of .21 between the aptitude and spelling measures indicates that the two are relatively independent of each other. Consequently, a significant discrepancy between the RPM and SAST for any given student could be indicative of orthogonal tests rather
than a neurological disorder. Therefore, using these two tests as the basis of spelling-based LD would be inappropriate.

**Descriptive statistics summary**

In this section the descriptive statistics of all variables at all grade-levels were displayed and discussed. In order to assist in an interpretation of the results, all variables were converted into $z$-scores, calculated separately for each year level. Both internal consistency for the two word lists and the test-retest reliability for the word lists and RAN tests were found to be acceptable. Furthermore, independent $t$-tests demonstrated that minor differences in the administration of the tests did not unduly affect results. The low percentage of variance accounted for by class membership in the variance components analysis indicated that the Grade 3 data set could be treated as independent data.

Finally, the correlation matrix of all variables for the Grade 3 cohort indicated that although all the variables were closely related, there was no evidence of multicollinearity between the variables. The only cause for concern was the very low correlation between the SAST and the RPM, indicating that a spelling-based aptitude achievement measure would not be appropriate, at least in the current study. This is because the correlation between the two measures is so low that there is no basis to the assumption that a discrepancy between the two tests would be meaningful.

**Cluster Analysis**

This section summarises the findings from two separate cluster analyses. The first, a cluster analysis with all of the variables tested in the study except for the SAST, will ascertain if there are generic types of readers with specific skill levels in different cognitive processes. The second cluster analysis uses all of the variables tested in the current study, except for the PAT-R, so that the cognitive processes can be assessed in regards to their relevance to spelling proficiency. Once each cluster is identified, an independent groups ANOVA is used in order to ascertain if there are significant differences in cognitive processing measures between the groups.

The cluster analysis, performed on the entire Grade 3 sample, was the method chosen as it was best suited to accommodate the ceiling effect in the irregular word reading (Chiu et al., 2001). The Bayesian Information Criterion (BIC) has been shown to be an excellent method of automatically identifying the most appropriate
number of clusters to retain (Fraley & Raftery, 1998). Consequently, the number of clusters identified by the BIC in both the reading and spelling models were retained. Furthermore, the log-likelihood distance measure, suitable for both continuous and categorical data and less effected by limited range within a variable (Zheng, Wang & Hu, 2007), was selected for calculating the closest cluster due to the narrow range of the digit span measures.

**Reading cluster analysis**

The aim of this particular analysis was to ascertain if the cognitive processing tests, along with measures of reading and aptitude, could result in clusters of students that reflect the importance of cognitive processes in reading. The reading cluster analysis utilised irregular word reading, nonword reading, colour and letter RAN and digit span forwards and backwards. The first step in the analysis utilised the BIC to ascertain how many clusters are appropriate for the data. The BIC for the-three-cluster group was lowest at 896.16. The next lowest BIC value was for the two-factor model at 903.44. The four-factor model BIC value was 922.90. Therefore, the three-factor model was retained.

The three clusters identified by the analysis could be described as representing poor, average and good readers. The poor readers were the smallest group comprising 18% of the sample ($n=30$). The average readers and good readers made up larger groups with 47% ($n=77$) and 34% ($n=56$) of the sample respectively. To get a clearer idea of how the groups performed in the various tests, graphs displaying the 95% confidence interval for the means of the three groups are shown in Figure 25.

As can be seen in Figure 25, there were noticeable differences between the low and average groups on all tests except for the RPM, digit span backwards and colour RAN. Cognitive processing measures discriminated between good, average and poor readers better than non-verbal intelligence as measured by the RPM. Furthermore, the digit span forward and the letter RAN seem to be better tests of verbal memory and naming speed for discriminating readers, than the digit span backward and colour RAN respectively.
Figure 22. 95% Confidence Interval of the Mean of all Variables for Groups of Poor, Average and Good Readers as Identified by a Cluster Analysis

N = 163.
Reading cluster analysis: Comparison of groups

To test the consistency of the groups identified by the cluster analysis test in separating the groups in each variable, one-way ANOVAs were performed on all variables. Group membership was based on automatic allocation into clusters. Because the area of interest in the current study is that of poor-performing readers, it was decided that, instead of planned contrasts, one single post hoc test would be utilised on the variables with significant differences in the ANOVA, comparing the low and average readers. Because there were eight tests within the ANOVAs, the Bonferroni-corrected alpha level is .006. All of the analyses were significant at \( p < .006 \) when tested over the three levels as shown in Table 15.

As can be seen in Table 15, the poor readers scored below average on all tests, with especially low scores on the irregular word reading, nonword reading and the PAT-R. The average readers scored close to average on all of the tests with the RPM as the lowest score and the digit span forwards score as the highest (and the only test

Table 15

Descriptive Statistics and Results from the One-Way ANOVA Comparing Poor, Average and Good Readers

<table>
<thead>
<tr>
<th>Test</th>
<th>Poor Readers</th>
<th>Average Readers</th>
<th>Good Readers</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>(2,160)</td>
</tr>
<tr>
<td>n = 30</td>
<td>n = 77</td>
<td>n = 56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irregular</td>
<td>-1.38 (0.84)</td>
<td>-0.12 (0.63)</td>
<td>0.90 (0.39)</td>
<td>150.68*</td>
</tr>
<tr>
<td>Nonword</td>
<td>-1.41 (0.62)</td>
<td>-0.11 (0.62)</td>
<td>0.90 (0.54)</td>
<td>142.35*</td>
</tr>
<tr>
<td>CRAN</td>
<td>-0.61 (0.91)</td>
<td>-0.23 (0.97)</td>
<td>0.65 (0.71)</td>
<td>25.21*</td>
</tr>
<tr>
<td>LRAN</td>
<td>-0.96 (0.99)</td>
<td>-0.17 (0.79)</td>
<td>0.75 (0.68)</td>
<td>48.15*</td>
</tr>
<tr>
<td>DSF</td>
<td>-0.86 (0.75)</td>
<td>0.08 (0.99)</td>
<td>0.35 (0.88)</td>
<td>17.78*</td>
</tr>
<tr>
<td>DSB</td>
<td>-0.56 (0.68)</td>
<td>-0.13 (0.89)</td>
<td>0.48 (1.09)</td>
<td>13.89*</td>
</tr>
<tr>
<td>RPM</td>
<td>-0.23 (0.98)</td>
<td>-0.29 (1.00)</td>
<td>0.53 (0.78)</td>
<td>13.81*</td>
</tr>
<tr>
<td>PAT-R</td>
<td>-1.24 (0.48)</td>
<td>-0.06 (0.84)</td>
<td>0.75 (0.65)</td>
<td>74.04*</td>
</tr>
</tbody>
</table>

\( N = 163. \) Note. Irregular = Irregular word reading; Nonword = Nonword reading; CRAN = Rapid Automatic Naming Colour; LRAN = Rapid Automatic Naming Letter; DSF = Digit Span Forward, DSB = Digit Span Backward; PAT-R = Progressive Achievement Test in Reading; RPM = Raven’s Progressive Matrices.

* Significant to a Bonferroni’s corrected significance value of \( p < .001 \).
with an above average mean for the whole group). The good readers scored, on average, above average on all of the tests. Once again, the highest scores were on the irregular word lists, nonword lists and the PAT-R as well as the letter RAN.

Independent groups \( t \)-tests were used to compare the low and average reading groups. This post hoc test was selected as it was the most important split in group membership given the area of study. The results of these \( t \)-tests can be seen in Table 16. In order to test for homogeneity of variance, a Levene’s test for homogeneity of variance was performed on all variables and degrees of freedom were adjusted where homogeneity of variance could not be assumed.

The means of the word lists, letter RAN, digit span forwards and the PAT-R, as shown in Table 16, were significantly higher in the average reading cluster than in the poor reading cluster. There was, however, no significant difference between poor and average readers in the digit span backwards, colour RAN or the RPM.

Table 16

Post-Hoc \( t \)-Test Results Comparing Poor and Average Cluster Analysis-Defined Reading Groups

<table>
<thead>
<tr>
<th>Test</th>
<th>( t )</th>
<th>( df )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irregular</td>
<td>8.50*</td>
<td>105</td>
</tr>
<tr>
<td>Nonword</td>
<td>9.77*</td>
<td>105</td>
</tr>
<tr>
<td>CRAN</td>
<td>1.81</td>
<td>105</td>
</tr>
<tr>
<td>LRAN(^a)</td>
<td>3.89*</td>
<td>44.08</td>
</tr>
<tr>
<td>DSF</td>
<td>4.70*</td>
<td>105</td>
</tr>
<tr>
<td>DSB(^a)</td>
<td>2.37</td>
<td>69.07</td>
</tr>
<tr>
<td>RPM</td>
<td>-0.25</td>
<td>105</td>
</tr>
<tr>
<td>PAT-R(^a)</td>
<td>9.12*</td>
<td>90.63</td>
</tr>
</tbody>
</table>

\( N = 163. \) \textit{Note.} Irregular = Irregular word reading; Nonword = Nonword reading; CRAN = Rapid Automatic Naming Colour; LRAN = Rapid Automatic Naming Letter; DSF = Digit Span Forward, DSB = Digit Span Backward; RPM = Raven’s Progressive Matrices; PAT-R = Progressive Achievement Test in Reading

\(^a\) Homogeneity of variance could not be assumed in this test, \( df \) and \( p \) values adjusted accordingly.

* Significant to a Bonferroni’s corrected significance value of \( p < .006. \)
Spelling cluster analysis

Given the paucity of research on the relevance of cognitive variables to spelling (Vostal et al., 2008), another cluster analysis was performed in which the spelling measure replaced the reading measure in the analysis. This provided an opportunity to see how relevant the research into reading and cognitive variables is to spelling. Once again, a two-step hierarchical cluster analysis using log-likelihood distance measures were utilised. The results from the BIC analysis indicated that, unlike the reading cluster analysis, a two-cluster structure was most appropriate with a BIC of 885.38 for this model.

In this cluster analysis, the groups could be best summarised as poor and good spellers. The groups were fairly evenly split. Eighty-seven poor spellers (53%) and 76 good spellers (47%) were identified. Although the poor and good spellers will have been split by this analysis, there will also be average spellers in each group, therefore less information can be gained on the cognitive profiles of these students. Although a two group split is not as informative as three groups, especially compared to a scenario where a smaller group of poor spellers is identified, t-tests were used to check which variables were important in identifying the two groups.

Spelling cluster analysis: Comparison of groups

In Table 17 the means and standard deviations on all tests of the two groups can be found, as well as the t-test results comparing the two groups. Good spellers performed significantly better than poor spellers in all tests, including the RPM. This could be taken as an indication that the groups are summarising a more general level of achievement than spelling proficiency in particular, especially when the low correlation between spelling and the RPM is taken into account.

Because a smaller group of struggling students has not been identified, the cognitive processes that are low in particularly poor spellers cannot be examined in this analysis. This means that these cognitive variables fail to provide information on what leads to an inability to spell.
Table 17
Means, Standard Deviations and t-Test Results Comparing Low and Good Spellers on all Tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Poor Spellers M(SD), n = 87</th>
<th>Good Spellers M (SD), n = 76</th>
<th>t</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irregulara</td>
<td>-0.67 (0.82)</td>
<td>0.77 (0.52)</td>
<td>13.68*</td>
<td>147.31</td>
</tr>
<tr>
<td>Nonworda</td>
<td>-0.65 (0.82)</td>
<td>0.74 (0.58)</td>
<td>12.59*</td>
<td>154.36</td>
</tr>
<tr>
<td>CRANa</td>
<td>-0.44 (1.01)</td>
<td>0.50 (0.71)</td>
<td>6.94*</td>
<td>154.29</td>
</tr>
<tr>
<td>LRAN</td>
<td>-0.47 (0.94)</td>
<td>0.53 (0.77)</td>
<td>7.35*</td>
<td>161</td>
</tr>
<tr>
<td>DSF</td>
<td>-0.29 (0.99)</td>
<td>0.33 (0.90)</td>
<td>4.17*</td>
<td>161</td>
</tr>
<tr>
<td>DSB</td>
<td>-0.23 (0.83)</td>
<td>0.27 (1.11)</td>
<td>3.30*</td>
<td>161</td>
</tr>
<tr>
<td>RPMa</td>
<td>-0.39 (1.04)</td>
<td>0.44 (0.73)</td>
<td>6.03*</td>
<td>153.96</td>
</tr>
<tr>
<td>SAST</td>
<td>-0.62 (0.78)</td>
<td>0.71 (0.71)</td>
<td>11.19*</td>
<td>161</td>
</tr>
</tbody>
</table>

Note: N = 163. Irregular = Irregular word reading; Nonword = Nonword reading; CRAN = Rapid Automatic Naming Colour; LRAN = Rapid Automatic Naming Letter; DSF = Digit Span Forward, DSB = Digit Span Backward; RPM = Raven’s Progressive Matrices; SAST = South Australian Spelling Test

*aHomogeneity of variance could not be assumed in this test, df and p values adjusted accordingly.
*significant to a Bonferroni corrected value of p < .006

Summary of findings from the cluster analyses

The reading cluster analysis resulted in a three-cluster model that identified a small cohort of poor readers as well as average and good readers. Non-verbal intelligence was not significantly different between these groups, nor was digit span backward or colour RAN. All other tests had significant differences in means between the average and poor clusters. In contrast, the spelling cluster analysis identified a two-cluster model, with both clusters containing a similar number of participants that could be considered good or poor spellers. All tests had significantly different means between poor and good spellers, as would be expected for an analysis that only differentiated between two levels of spelling ability.

These results indicated that, although these cognitive processing variables of phonological processing, naming speed and verbal memory are well suited to measures of reading, they are less appropriate for a study on spelling. These tests seem to identify the processes that good spellers are good at, but not the processes that poor spellers are poor at. This was not wholly unexpected given the paucity of research on spelling disabilities as compared to reading disabilities.
When the results from the spelling-based cluster analysis are considered in conjunction with the low correlation between the SAST and the RPM, it becomes clear that any analysis based on spelling-based LD utilising the discrepancy method, is not tenable in this study. From this point onwards, the focus of the current study will be on reading-based LD. The next section will move forward from examining the cognitive variables associated with poor reading and instead focus on the cognitive processes that cause LD.

**Identifying students with learning disabilities with discrepancy-based methods**

In this section, the cognitive profiles of students identified via the traditional or regression-adjusted discrepancy methods will be analysed. As detailed in Chapter 4, there are two ways that discrepancy-defined groups can be identified. The first, traditional discrepancy, identifies LD students on the basis of a difference of one standard deviation or more between aptitude and reading achievement (C. Reynolds, 1984). LA students identified through traditional discrepancy are those who have a reading score of one standard deviation below the mean on reading and have not been designated as LD. All students not identified as LD or LA were designated as RA. The second method of identification, regression-adjusted discrepancy, is the same as traditional discrepancy except that the magnitude of the discrepancy between aptitude and achievement is adjusted to take into account regression to the mean. Consequently, students with higher IQ scores need a larger discrepancy between aptitude and achievement in order to qualify as LD than those with lower IQ scores (Fletcher et al., 1989). The process for identifying LA and RA students is the same in both methods, however the groups will not contain exactly the same students due to the difference in how the LD group is identified.

Once these groups have been allocated, the first analysis in this section is the ANOVA comparing the three groups, as dictated by traditional discrepancy, on all variables tested in the study. The second analysis is the ANOVA comparing the groups as dictated by regression-adjusted discrepancy. Both of these analyses are conducted in order to provide some context on how the groupings, as identified in previous literature, correspond to the groupings used later in the study. Two planned contrasts will be performed on all significantly different variables. The first of these contrasts will compare the RA group to the LD and LA groups combined, while the second will compare the LD and LA groups.
Traditional discrepancy

The first step in analysing how well students can be allocated into traditional discrepancy groups by cognitive processing variables is to ascertain which students will fall into these traditional discrepancy-defined groups. As discussed in Chapter 4, students who have a \( z \)-score on the PAT-R one standard deviation below their \( z \)-score on the RPM were designated LD. Any students who are more than one standard deviation below the mean in reading who do not meet this criterion, are designated LA. Using this method of identification, 27 (17\%) students were flagged as belonging in the LD group. The numbers for LA identification are a little lower than LD with 19 (12\%) students. The remaining 117 students (72\%) were allocated into the RA group.

The descriptive statistics for the traditional discrepancy based groups are shown in Table 18. Furthermore, in order to ascertain if the cognitive processing differences between the traditional discrepancy-defined groups were significant, a series one-way ANOVAs were conducted. The results from the one-way ANOVAs are also shown in Table 18.

<table>
<thead>
<tr>
<th>Test</th>
<th>RA (n = 117)</th>
<th>LD (n = 27)</th>
<th>LA (n = 19)</th>
<th>( F ) (2,160)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irregular</td>
<td>0.32 (0.72)</td>
<td>-0.74 (1.21)</td>
<td>-0.90 (1.09)</td>
<td>28.19*</td>
</tr>
<tr>
<td>Nonword</td>
<td>0.27 (0.83)</td>
<td>-0.69 (1.20)</td>
<td>-0.65 (0.95)</td>
<td>17.55*</td>
</tr>
<tr>
<td>CRAN</td>
<td>0.08 (0.95)</td>
<td>-0.32 (1.24)</td>
<td>-0.09 (0.87)</td>
<td>1.89</td>
</tr>
<tr>
<td>LRAN</td>
<td>0.20 (0.94)</td>
<td>-0.054 (0.97)</td>
<td>-0.49 (1.02)</td>
<td>9.62*</td>
</tr>
<tr>
<td>DSF</td>
<td>0.14 (0.97)</td>
<td>-0.13 (0.98)</td>
<td>-0.67 (0.96)</td>
<td>6.00*</td>
</tr>
<tr>
<td>DSB</td>
<td>0.07 (0.97)</td>
<td>0.08 (1.12)</td>
<td>-0.55 (0.87)</td>
<td>3.33</td>
</tr>
<tr>
<td>RPM</td>
<td>0.46 (0.71)</td>
<td>-0.93 (0.66)</td>
<td>-1.51 (0.30)</td>
<td>28.30*</td>
</tr>
<tr>
<td>PAT-R</td>
<td>0.04 (0.93)</td>
<td>0.69 (0.67)</td>
<td>-1.24 (0.66)</td>
<td>101.96*</td>
</tr>
</tbody>
</table>

\( N = 163. \) *Note.* Irregular = Irregular word reading; Nonword = Nonword reading; CRAN = Rapid Automatic Naming Colour; LRAN = Rapid Automatic Naming Letter; DSF = Digit Span Forward, DSB = Digit Span Backward; RPM = Raven’s Progressive Matrices; PAT-R = Progressive Achievement Test in Reading.

* Significant to a Bonferroni’s corrected \( p \) value of .008.
All of the variables tested in the current study had significantly different means between traditional discrepancy-defined groups except for the colour RAN and digit span backwards tests. There was, however, a non-significant trend for the digit span backwards test. In order to investigate these significant differences further, planned contrasts were used, the first comparing the RA group to the LD and LA groups and the second comparing the LD and LA groups. The results of the two planned contrasts can be seen in Table 19.

The RA group significantly outperformed the LD and LA groups, on average, on all six measures. On the two processing variables thought to be most intrinsically linked to LD, phonological processing and naming speed, as well as orthographic processing and verbal memory, there were no significant differences between the LD and LA students. Therefore, as per the third research question, the same deficits that are thought to be responsible for the LD students’ neurological disorder (Wolf & Bowers, 1999) are present in students who are not thought to have that same disorder, at least when grouped via the traditional discrepancy method. Furthermore, these cognitive deficits cannot be attributed to aptitude as the LD group had higher IQ scores, on average, than the regular achievement group.

Table 19

<table>
<thead>
<tr>
<th>Test</th>
<th>First Contrast(^a)</th>
<th>Second Contrast(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(t)</td>
<td>(df)</td>
</tr>
<tr>
<td>Irregular</td>
<td>-6.22*</td>
<td>54.28</td>
</tr>
<tr>
<td>Nonword</td>
<td>-5.31*</td>
<td>64.46</td>
</tr>
<tr>
<td>LRAN</td>
<td>-4.30*</td>
<td>160</td>
</tr>
<tr>
<td>DSF</td>
<td>-3.17*</td>
<td>160</td>
</tr>
<tr>
<td>RPM</td>
<td>-2.12</td>
<td>160</td>
</tr>
<tr>
<td>PAT-R</td>
<td>-17.22</td>
<td>105.87</td>
</tr>
</tbody>
</table>

Note. Irregular = Irregular word reading; Nonword = Nonword reading; LRAN = Rapid Automatic Naming Letter; DSF = Digit Span Forward; RPM = Raven’s Progressive Matrices; PAT-R = Progressive Achievement Test in Reading

\(^a\) First Contrast compares the RA group to the combined LD and LA groups. \(^b\) Second Contrast compares the LD group to the LA group. \(^c\) Levene’s test for homogeneity of variance was significant, degrees of freedom adjusted accordingly

* = significant to a Bonferroni’s corrected value of .013
Regression-adjusted discrepancy

So that the results of the current study can be compared to studies using regression-adjusted discrepancy as well as traditional discrepancy, students were also grouped as per the tenets of regression-adjusted discrepancy. The first step of this is to allocate students into the LD, LA or RA groups via this method. As discussed in Chapter 4, students who meet the following criterion, shown in Equation 6, are allocated as LD as per the tenets of the regression-adjusted discrepancy method.

\[(r \times z_{RPM}) - z_{PAT-R} \geq 1.\]  
\(6\)

where \(z_{RPM}\) is any given student’s \(z\)-score on the RPM, \(z_{PAT-R}\) is the same student’s \(z\)-score on the PAT-R and \(r\) is the correlation between the two scores. The correlation between the RPM and the PAT-R, representing aptitude and achievement, is .42 (see Table 14). Consequently, Equation 7 was used to identify LD students via the regression-adjusted discrepancy method.

\[(.42 \times z_{RPM}) - z_{PAT-R} \geq 1.\]  
\(7\)

where \(z_{RPM}\) is any given student’s \(z\)-score on the RPM and \(z_{PAT-R}\) is the same student’s score on the PAT-R. If students had a reading level one standard deviation or more below the mean on the PAT-R, but were not in the LD group, they were designated LA. All students not designated LD or LA were placed in the RA group. Scatter plots displaying samples’ RPM and PAT-R scores as identified via both discrepancy methods are shown in Figure 23, so that the two methods can be compared.

Many students remain in the same group irrespective of which method of identification has been used. The primary difference lies in the LD identification: high IQ students are more likely to be considered LD via the traditional method and more likely to be considered RA via the regression-adjusted method. Furthermore, Low IQ students are more likely to be considered LA via the traditional method and LD via the regression-adjusted method. Although the actual number of LD students identified via the regression-adjusted method remained steady at 28 (17%), the
Figure 23. Scatter Plot Displaying Aptitude and Achievement Scores of Grade 3 Students, Allocated into Groups via the Traditional and Regression-Adjusted Discrepancy Method

$N = 163$.  
Note: RA = Regular Achievement; LD = Learning Disability; LA = Low Achievement; PAT-R = Progressive Achievement Test in Reading; RPM = Raven’s Progressive Matrices.

A = Traditional discrepancy method of identification; B = Regression-adjusted discrepancy method of identification.
The number of LA students identified was reduced to 10 (6%, compared to 12% in the traditional discrepancy method) due to the change in the allocation methods. Furthermore, there were a few more students, 125 (77% compared to 72% in the traditional discrepancy method) identified as RA than there were for the traditional discrepancy scenario.

The means and standard deviations of the three groups on all tests are shown in Table 20. Additionally, the one-way ANOVA results comparing the three groups are also shown. Significant differences in means between the three groups in all tests except for the digit span backwards and colour RAN tests were found. In order to further investigate the significant differences, planned contrasts were performed on the remaining six variables.

Planned contrasts were calculated so that the significant differences could be further investigated. As can be seen in Table 21, the RA group significantly outperformed the LD and LA groups combined on all four measures, as would be expected. The second contrast compared the LD and LA groups. The only test that had significant mean differences between LD and LA groups was the irregular word list, in which

Table 20

Means, Standard Deviations and Corresponding ANOVA Results of the Three Regression-Adjusted Discrepancy-Defined Groups on Measures of Reading and Cognitive Processing

<table>
<thead>
<tr>
<th></th>
<th>RA (N = 125)</th>
<th>LD (N = 28)</th>
<th>LA (N = 10)</th>
<th>F (148,14)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td></td>
</tr>
<tr>
<td>Irregular</td>
<td>0.32 (0.73)</td>
<td>-1.26 (0.98)</td>
<td>-0.43 (1.11)</td>
<td>45.87*</td>
</tr>
<tr>
<td>Nonword</td>
<td>0.25 (0.83)</td>
<td>-1.02 (1.07)</td>
<td>-0.28 (0.93)</td>
<td>24.13*</td>
</tr>
<tr>
<td>CRAN</td>
<td>0.12 (0.97)</td>
<td>-0.43 (1.12)</td>
<td>-0.25 (0.74)</td>
<td>3.82</td>
</tr>
<tr>
<td>LRAN</td>
<td>0.18 (0.94)</td>
<td>-0.69 (1.01)</td>
<td>-0.40 (0.84)</td>
<td>10.94*</td>
</tr>
<tr>
<td>DSF</td>
<td>0.18 (0.97)</td>
<td>-0.70 (0.89)</td>
<td>-0.27 (0.82)</td>
<td>10.11*</td>
</tr>
<tr>
<td>DSB</td>
<td>0.11 (1.03)</td>
<td>-0.43 (0.68)</td>
<td>-0.23 (1.08)</td>
<td>3.72</td>
</tr>
</tbody>
</table>

N = 163. Note. RA = Regular Achievement; LD = Learning Disability; LA = Low Achievement; Irregular = Irregular word reading; Nonword = Nonword reading; CRAN = Rapid Automatic Naming Colour; LRAN = Rapid Automatic Naming Letter; DSF = Digit Span Forward, DSB = Digit Span Backward.

* significant to a Bonferroni’s corrected value of 0.008.
Table 21

Planned Contrasts Comparing Regression-Adjusted Discrepancy-Defined Groups on Irregular Word Reading, Nonword Reading, Digit Span Forward and Letter RAN

<table>
<thead>
<tr>
<th>Test</th>
<th>First Contrast&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Second Contrast&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( t )</td>
<td>( df )</td>
</tr>
<tr>
<td>Irregular</td>
<td>-7.09&lt;sup&gt;*&lt;/sup&gt;</td>
<td>160</td>
</tr>
<tr>
<td>Nonword</td>
<td>-4.96&lt;sup&gt;*&lt;/sup&gt;</td>
<td>160</td>
</tr>
<tr>
<td>LRAN</td>
<td>-3.79&lt;sup&gt;*&lt;/sup&gt;</td>
<td>160</td>
</tr>
<tr>
<td>DSF</td>
<td>-3.40&lt;sup&gt;*&lt;/sup&gt;</td>
<td>160</td>
</tr>
</tbody>
</table>

\( N = 163. \) Note. Irregular = Irregular word reading; Nonword = Nonword reading; LRAN = Rapid Automatic Naming Letter; DSF = Digit Span Forward.

<sup>a</sup> First Contrast compares the RA group to the combined LD and LA groups.

<sup>b</sup> Second Contrast compares the LD group to the LA group.

<sup>*</sup> = significant to a Bonferroni’s corrected value of .013.

the LA students outperformed the LD students. This finding was unexpected, as it was thought that if LD students were to be outperformed by LA students, it would be on more traditionally accepted cognitive processes such as phonological processing or naming speed. Once again there are no significant differences in phonological processing or naming speed, thought to be deficient in the LD group, between the LD and LA groups.

Overall, the lower cognitive processing scores in the regression-adjusted discrepancy-defined LD groups indicate that, given the importance of these variables to LD, regression-adjusted discrepancy groups may be more accurate than traditional discrepancy groups at identifying LD students. In order to further investigate correlation between cognitive processing variables and discrepancy-defined group membership, membership in these groups was predicted by the cognitive processing tests using DFA. This is outlined in the next section.

**Discriminant Function Analysis**

Two separate DFAs were calculated in order to ascertain if phonological processing, naming speed and verbal memory could predict membership in the groups dictated by both the traditional and regression-adjusted discrepancy methods. As LD, LA and RA groups are not equally represented, in either our sample or the general
population, the compute from group sizes option was selected as compared to the ‘all
groups equal’ option. Box’s M was used to assess the suitability of the data to a
DFA and Fishers classic function coefficients were utilised for assessing the function
loadings. As discussed in Chapter 3, it was decided that the irregular word reading
variable would not be included as one of the cognitive processing variables. This was
decided as the role of orthographic processing as a cause or effect in reading
comprehension is still unclear (Byrne et al., 1992; Castles et al., 1999; Price &
Mechelli, 2005).

The first analysis reported in this section is the DFA using the traditional
discrepancy groupings for the outcome variable. The predictor variables are
nonword reading, colour RAN, letter RAN, digit span forward and digit span
backward. With new groups developed on the basis of the DFA, one-way ANOVAs
will be performed on these new groups. This will be done in order to ascertain what
the differences are between groups if allocation into groups is ascertained by
cognitive processing test scores alone.

For the final part of this section, the same analyses will be performed with the
regression-adjusted discrepancy groupings taking the place of the traditional
discrepancy groupings. Therefore, a DFA with the five cognitive processing
predicting variables predicting LD, LA and RA groups as found by regression-
adjusted discrepancy will be performed. With the new groups developed as part of
the DFA, one-way ANOVAs will be performed on these new groups in the same
manner as they were done for the traditional discrepancy predicted DFA groupings.

**Traditional discrepancy discriminant function analysis**

A DFA was performed where the target variable was the groups dictated by LD, LA
or RA, utilising the traditional discrepancy method to ascertain if cognitive processes
were a good predictor of those designated LD, or LA, via the discrepancy method.
There were 117 students in the RA group, 27 in the LD group and 19 in the LA
group for this analysis. Despite concerns about the size of the LA group, it was still
sufficient for this analysis as the size of the smallest cell needs to be larger than the
number of predictor variables (Tabachnik & Fidell, 2001), in this case five.

Results from Box’s M indicated that the data was suitable for a DFA, $F(30,9) = 1.44$,
$p > .005$. Two discriminant functions were calculated, although the second function
on its own, was not significant, $\chi^2 (4) = 7.99, p > .05$. Importantly, the two functions combined were, $\chi^2 (10) = 43.36, p < .001$. The standardised loading matrix of the discriminant functions is shown in Table 22. The first discriminant function, responsible for 84% of the variance in the model, differentiated between the RA group and the combined LD and LA groups, and the second function, responsible for the remaining 16% of the variance in the model, differentiated between the LD and LA groups. Table 22 shows that the nonword test is a very strong differentiator between the RA group and the combined LD and LA groups in the first function. Letter RAN was also a good source of differentiation in the first function. The second function, separating the LD and LA groups, was more heavily centred on verbal memory. The subsequent classification of students into the RA, LD and LA groups is shown in Table 23.

In this DFA, 77% of cases were correctly classified. As outlined in Chapter 4, the probability of correct classification into groups by chance is affected by the proportion of students in each group. In the original traditional discrepancy groupings 72% of students were RA, 17% LD and 12% LA. Therefore the percentage of students that would be correctly allocated by chance alone is shown in Equation 8.

$$\% \text{ Allocated by chance} = \frac{[(pA*nA) + (pB*nB) + (pC*nC)]}{N}$$

$$= \frac{[(.72*117) + (.16*27) + (.12*19)]}{163}$$

$$= \frac{90.71}{163}$$

$$= 56\%$$

where $p$ = the percentage of students in group A, B or C = the number of students in each group and $N$ = the total number of participants in the sample. With 56% expected to be allocated by chance alone, the 77% allocated in this DFA is good, although not ideal. The stability of this classification was checked using the cross-validation procedure, in which 76% of the cases were correctly classified, and therefore indicating a fairly stable classification. The primary concern in this particular DFA was the low number of LD and LA students correctly identified in this analysis. Only eight of the 27 LD and five of the 19 LA students were correctly classified into their groups.
Table 22

| Discriminant Functions Predicting Membership in Traditional Discrepancy Groups |
|---------------------------------|-----------|-----------|
|                                 | Function 1 | Function 2 |
| Nonword                         | .90       | -.11      |
| LRAN                            | .69       | -.11      |
| CRAN                            | .28       | -.28      |
| DSF                             | .47       | .62       |
| DSB                             | .24       | .73       |


**Traditional discrepancy discriminant function analysis: Comparison of groups**

In order to ascertain the differences between the groups identified through the DFA a one-way ANOVA was utilised. The means and standard deviations for the three groups, along with the results from the ANOVA are shown in Table 24.

Significant differences in means were found between the LD, LA and RA groups on all tests except in the backwards digit span, colour RAN and the RPM. The RA group had the highest mean for all of the variables except for non-significant differences in backwards digit span and RPM. The LA group had the lowest mean in

Table 23.

| Classification of Students into Traditional Discrepancy-Defined Groups through a Discriminant Function Analysis |
|---------------------------------|-----------|-----------|-----------|-----------|
|                                 | RA        | LD        | LA        | Total     |
| TD Groups                       |           |           |           |           |
| RA                              | 113       | 3         | 1         | 117       |
| LD                              | 18        | 8         | 1         | 27        |
| LA                              | 14        | 0         | 5         | 19        |
| Total                           | 145       | 11        | 7         | 163       |

N = 163. Note. TD = Traditional Discrepancy; DFA = Discriminant Function Analysis; LD = Learning Disability, LA = Low Achievement.
Table 24

*Means, Standard Deviations and Results from the One-Way ANOVA Comparing Traditional Discrepancy Discriminant Function Analysis Defined Groups*

<table>
<thead>
<tr>
<th></th>
<th>RA (N = 145)</th>
<th>LD (N = 11)</th>
<th>LA (N = 7)</th>
<th>F (160,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M(SD)</td>
<td>M(SD)</td>
<td>M(SD)</td>
<td></td>
</tr>
<tr>
<td>Irregular</td>
<td>0.19 (0.82)</td>
<td>-1.35 (0.64)</td>
<td>-1.86 (1.40)</td>
<td>35.25*</td>
</tr>
<tr>
<td>Nonword</td>
<td>0.22 (0.81)</td>
<td>-1.68 (0.66)</td>
<td>-1.82 (0.50)</td>
<td>48.23*</td>
</tr>
<tr>
<td>CRAN</td>
<td>0.06 (0.99)</td>
<td>-0.68 (1.07)</td>
<td>-0.08 (0.65)</td>
<td>2.88</td>
</tr>
<tr>
<td>LRAN</td>
<td>0.17 (0.87)</td>
<td>-1.32 (1.05)</td>
<td>-1.39 (0.94)</td>
<td>23.36*</td>
</tr>
<tr>
<td>DSF</td>
<td>0.12 (0.96)</td>
<td>-0.48 (0.77)</td>
<td>-1.64 (0.42)</td>
<td>13.47*</td>
</tr>
<tr>
<td>DSB</td>
<td>0.04 (1.00)</td>
<td>0.12 (0.83)</td>
<td>-1.10 (0.54)</td>
<td>4.65</td>
</tr>
<tr>
<td>RPM</td>
<td>0.04 (0.99)</td>
<td>0.11 (0.93)</td>
<td>-0.92 (1.05)</td>
<td>3.23</td>
</tr>
<tr>
<td>PAT-R</td>
<td>0.14 (0.92)</td>
<td>-1.12 (0.64)</td>
<td>-1.23 (1.21)</td>
<td>16.23*</td>
</tr>
</tbody>
</table>

*Note. N = 163. Irregular = Irregular word reading; Nonword = Nonword reading; CRAN = Rapid Automatic Naming Colour; LRAN = Rapid Automatic Naming Letter; DSF = Digit Span Forward, DSB = Digit Span Backward; RPM = Raven’s Progressive Matrices; PAT-R = Progressive Achievement Test in Reading.

* = significant to a Bonferroni’s corrected value of .006.

all of the variables except for non-significant differences in colour RAN, in which the LD group had the lowest mean.

In order to further investigate the significant differences without compromising Type I error, planned contrasts were utilised. The first contrast compared the RA group to the LD and LA groups and the second contrast compared the LD and LA groups. The results from the two planned contrast are shown in Table 25. There were significant differences between the RA readers and the combined LD and LA group with the RA readers scoring better on the irregular word reading, nonword reading, digit span forward, letter RAN and the PAT-R. Interestingly, the digit span forward was the only test with a significant difference between the LD and LA groups, with the LD group performing significantly better than the LA group on this test. This suggests that the LD and LA groups have similar deficits in phonological processing and naming speed. However, in addition to these deficits, the LA group also has deficits in verbal memory.
Table 25

Results from the Planned Contrast Comparing Regular Achieving Students to Students with Learning Disabilities and Low Achieving Students in Traditional Discrepancy Discriminant Function Analysis Defined Groups

<table>
<thead>
<tr>
<th>Test</th>
<th>First Contrast(^a)</th>
<th>Second Contrast(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(t)</td>
<td>(df)</td>
</tr>
<tr>
<td>Irregular(^c)</td>
<td>6.20*</td>
<td>8.56</td>
</tr>
<tr>
<td>Nonword</td>
<td>9.67*</td>
<td>160</td>
</tr>
<tr>
<td>LRAN</td>
<td>6.72*</td>
<td>160</td>
</tr>
<tr>
<td>DSF(^c)</td>
<td>7.28*</td>
<td>27.05</td>
</tr>
<tr>
<td>PATR</td>
<td>5.62*</td>
<td>160</td>
</tr>
</tbody>
</table>

Note. \(N = 163\). Irregular = Irregular word reading; Nonword = Nonword reading; CRAN = Rapid Automatic Naming Colour; LRAN = Rapid Automatic Naming Letter; DSF = Digit Span Forward; PATR = Progressive Achievement Test in Reading

\(^a\) First Contrast compares the RA group to the combined LD and LA groups. \(^b\) Second Contrast compares the LD group to the LA group. \(^c\) Levene’s test for homogeneity of variance was significant, degrees of freedom adjusted accordingly

* Significant to Bonferroni corrected value of .010

It is possible that by using these processes to define membership in the groups, the confounding factor of alternative explanations for a discrepancy between aptitude and achievement has been reduced. As such, the actual difference between LD and LA students, as they are currently operationalised, may be short term memory deficits. In order to further investigate this idea, the same DFA was performed on the regression-adjusted discrepancy-defined groups.

**Regression-adjusted discriminant function analysis**

In this section the same analyses that were performed in the traditional discrepancy DFA section will be repeated, except that this time, the groups will have been determined by the regression-adjusted discrepancy method. As previously mentioned, there were 125 students in the RA group, 28 in the LD group and 10 students in the LA group. Consistent with the previous DFA, the smallest cell size was larger than the number of predictor variables (Tabachnik & Fidell, 2001), which were once again nonwords, colour and letter RAN, and digit span forward and backward.
Box’s M indicated that the data was suitable for a DFA, $F(30,2) = 1.82, p > .05$. Two
discriminant functions were calculated, the two functions combined were significant,
$\chi^2 (10) = 49.34, p < .001$, however the second function was not, $\chi^2 (4) = 0.55, p > .05$. The loading matrix of the discriminant functions is shown in Table 26. The
first function that discriminated between the RA and combined LD and LA groups,
was responsible for 99% of the variance in the model, and had high to moderate
loadings for the nonword, DSF and LRAN tests. The second function, responsible
for only 1% of the variance in the model, was based on the two RAN tests and
attempted to differentiate between the LD and LA groups.

As can be seen in Table 26 the scores on nonwords, DSF and LRAN were the best
variables for distinguishing between RA students and those that are either LD or LA.
This is particularly interesting as it could call into question this idea of low
achievement. If the cognitive variables that predict LD are used, no LA group is
found. The second factor that distinguished between LD and LA students was less
clear, and a reflection of the small amount of variance this factor accounted for by
this factor. The classification of students into the LD, LA and RA groups is shown
in Table 27.

This DFA was not able to predict membership into the LA group using the cognitive
variables provided. Despite this, 82% of students were correctly classified. Using
the method outlined in Equation 8, it would be expected that 63% of these students
would have been allocated correctly by chance. The stability of this classification

<table>
<thead>
<tr>
<th>Table 26</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discriminant Functions Predicting Membership in Regression-Adjusted Discrepancy Groups</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Nonword</td>
</tr>
<tr>
<td>CRAN</td>
</tr>
<tr>
<td>LRAN</td>
</tr>
<tr>
<td>DSF</td>
</tr>
<tr>
<td>DSB</td>
</tr>
</tbody>
</table>

Table 27

Classification of Students into Regression-Adjusted Discrepancy-Defined Groups.

<table>
<thead>
<tr>
<th>Original Groups</th>
<th>New Groups</th>
<th>RA</th>
<th>LD</th>
<th>LA</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>121</td>
<td>4</td>
<td>0</td>
<td></td>
<td>125</td>
</tr>
<tr>
<td>LD</td>
<td>15</td>
<td>13</td>
<td>0</td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>LA</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>144</td>
<td>19</td>
<td>0</td>
<td></td>
<td>163</td>
</tr>
</tbody>
</table>

N = 163. Note. RA = Regular Achievement; LD = Learning Disability; LA = Low Achievement

was checked using the cross-validation procedure. The full 100% of the cases were classified in the same manner with the corresponding 82% rate of correct classification.

Regression-adjusted discriminant function analysis: Comparison of groups

Because only two groups were identified through the DFA predicting membership in the regression-adjusted groups, independent t-tests were conducted in order to ascertain the differences between these groups. The means, standard deviations and the results of the t-tests on all of the variables for the two groups identified by the DFA are shown in Table 28.

Significant differences in means of all tests except for CRAN and RPM were found between groups. In all of the significant tests, the RA group scored significantly higher on average than the LD group. The fact there were significant differences is not surprising as all of these tests, except for the irregular word list, were utilised in making up the groups. What is of interest here is that the RPM, one of two factors utilised in order to discriminate between the groups in the outcome variable, was not significantly different for the two groups.

It should be noted at this point that the lack of an LA group in this analysis resulted in low-performing RA readers. This is consistent with the idea that there are different reasons for students having difficulty with reading other than LD. Therefore, it is possible that the regression-adjusted discrepancy is closer to the construct of LD, as it would be more represented by cognitive deficits than the traditional discrepancy method.
Table 28

Means, Standard Deviations and Results from the One-Way ANOVA comparing the Discriminant Function Analysis Defined Groups

<table>
<thead>
<tr>
<th></th>
<th>RA M(SD)</th>
<th>LD M(SD)</th>
<th>t</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 144</td>
<td></td>
<td>n = 19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irregular</td>
<td>0.18 (0.85)</td>
<td>-1.40 (0.98)</td>
<td>7.51*</td>
<td>161</td>
</tr>
<tr>
<td>Nonword</td>
<td>0.22 (0.81)</td>
<td>-1.69 (0.58)</td>
<td>9.96*</td>
<td>161</td>
</tr>
<tr>
<td>CRAN</td>
<td>0.07 (1.00)</td>
<td>-0.56 (0.87)</td>
<td>2.63</td>
<td>161</td>
</tr>
<tr>
<td>LRAN</td>
<td>0.15 (0.89)</td>
<td>-1.16 (1.04)</td>
<td>5.93*</td>
<td>161</td>
</tr>
<tr>
<td>DSF&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.17 (0.92)</td>
<td>-1.25 (0.59)</td>
<td>9.08*</td>
<td>31.05</td>
</tr>
<tr>
<td>DSB&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.08 (1.00)</td>
<td>-0.61 (0.76)</td>
<td>3.54*</td>
<td>26.95</td>
</tr>
<tr>
<td>RPM</td>
<td>0.05 (0.98)</td>
<td>-0.34 (1.09)</td>
<td>1.62</td>
<td>161</td>
</tr>
<tr>
<td>PATR</td>
<td>0.15 (0.92)</td>
<td>-1.16 (0.78)</td>
<td>5.92*</td>
<td>161</td>
</tr>
</tbody>
</table>

N = 163. Note. Irregular = Irregular word reading; Nonword = Nonword reading; CRAN = Rapid Automatic Naming Colour; LRAN = Rapid Automatic Naming Letter; DSF = Digit Span Forward, DSB = Digit Span Backward; RPM = Raven’s Progressive Matrices; PATR = Progressive Achievement Test in Reading.

<sup>a</sup>Levene’s test for homogeneity of variance was significant so the adjusted degrees of freedom were utilised.

* significant to a Bonferroni’s correct p value of .006

Summary of findings from the discriminant function analyses

The aim of this section was to ascertain how well cognitive processing variables predicted group membership in LD, LA or RA groups, as indicated by the traditional and regression-adjusted discrepancy analyses.

In the first DFA, five cognitive processing tests, namely the nonword reading, digit span forward, digit span backward, colour RAN and letter RAN, were used to predict membership in RA, LD and LA groups as dictated by the traditional discrepancy method of identification. This analysis was fairly successful with 77% of participants correctly classified in their groups. This suggests that there is some merit to the idea that these groupings are presenting, to some extent, cognitive processing deficits. When the new LD, LA and RA groups (as predicted by the DFA) were compared through a series of one-way ANOVA, it was found that once
again the digit span backwards, colour RAN and RPM, provided no differentiation between the three groups overall.

Furthermore, in a planned contrast analysis of the scores on the other four tests, irregular word reading, nonword reading, digit span forwards and letter RAN were significantly better, on average in the RA group than the LD and LA groups combined. The most interesting finding to come out of this analysis, however, was that only one test, the digit span forward, had a significant difference in the planned contrast between the traditional discrepancy LD and LA groups as predicted by the DFA. The mean of the digit span forward was higher in the LD group than it was in the LA group. This finding is of note given the assumption that LD students have neurological deficits that the LA group does not have. The importance of this finding will be discussed in Chapter 7.

The results from the regression-adjusted discrepancy DFA were slightly different. In this analysis the same five cognitive processing tests, were used to predict membership in the regression-adjusted discrepancy-defined groups and in this analysis, only the LD and RA groups had predicted members. The t-tests comparing these two groups found that the RA group significantly outperformed the LD group on all measures, except on the RPM and colour RAN. The implications of this finding will also be discussed in Chapter 7.

With the analyses in the previous section providing support for the idea that these cognitive processes can predict membership into discrepancy-defined groups, the next step was to devise a method of allocation that did not depend on a discrepancy-based method. This method is outlined in the next section.

**Identifying Students with Learning Disabilities through Cognitive Processing Deficits**

Given that cognitive processing variables can predict the discrepancy-dictated groupings, it was decided to identify groups of students based on these cognitive deficits alone. Due to the finding in the current study that letter RAN was a better predictor of reading in a Grade 3 cohort than colour RAN, the letter RAN was utilised as the naming speed deficit. Furthermore, as the digit span forward is better than the digit span backwards at differentiating between poor and LD readers and as
evidenced in the DFAs and the cluster analysis, the digit span forward was utilised as the verbal memory measure.

**Identifying reading-age-based phonological deficits**

Phonological deficits can be found in reading level-matched designs (Snowling, 1980) but not naming speed deficits (Griffiths & Snowling, 2001) or verbal memory deficits (De Jong, 1998). Therefore, two different procedures for identifying cognitive processing deficits in the sample were used. Phonological deficits are best ascertained through the use of reading-age, rather than chronological age-based norms so that the confounding reciprocal relationship between phonological processing can, at least in part, be controlled. Reading age norms were utilised so that the students with low phonological processing for their reading ability, rather than their chronological age, are identified.

In Chapter 4, a method for adjusting Grade 1 and 2 reading scores so that they could be compared to a Grade 3 standard was outlined. To reiterate, this was made possible by the fact that the PAT-R has tests that are suited to three grade-levels. This meant that each raw score had a Grade 1, 2 or 3 percentile score that could be cross-referenced. Consequently, the corresponding norms for Grades 1 to 3 provided an opportunity to calculate a rough approximation of what score would be required on the PAT-R in Grade 3 to indicate reading at a Grade 1 or 2 level. Any student with a z-score under -0.98 was designated as reading at a Grade 1 level and any student with a PAT-R z-score between -0.97 and -0.34 was designated as reading at a Grade 2 level. Any student with a z-score of -0.33 or above was designated as reading at a Grade 3 level (see Chapter 4).

Using this method, 33 Grade 3 students were identified as reading at a Grade 1 level and 25 Grade 3 students were identified as reading at a Grade 2 level. The remaining 105 students were identified as reading at or above a Grade 3 level. The next step was to calculate the three cut-off scores, one per reading-age, for nonword reading that would indicate a phonological processing deficit. As outlined in Chapter 4, the criterion that needed to be met to be considered as having a nonword deficit was presented. Students whose nonword reading score was half a standard deviation below the mean nonword reading score for their appropriate reading age were considered to have a phonological deficit.
To ensure that the students in the Grade 1 and 2 samples for reading-age comparisons of nonword reading were in fact reading at an appropriate level for their age, only those students who were in the inter-quartile range for their grade on reading were used in this analysis to calculate the average reading-age comparison. To elaborate, the average nonword reading score for each grade was calculated using only those students that fell in the inter-quartile range for reading ability. The $z$-scores on the nonword reading list required for Grade 1, 2 and 3 level readers to be designated as deficient (or not) are shown in Table 29. Please note that, unlike in the rest of the chapter, the $z$-scores for nonword reading shown in Table 29 are based on $z$-scores calculated for the entire sample, not just the Grade 3 cohort.

Using this method of identification, eight of the 33 Grade 3 students reading at Grade 1 level were identified as having a phonological deficit, seven of the 25 reading at a Grade 2 level were identified as having a phonological deficit and 10 of the 105 students reading at the grade appropriate level were identified as having a phonological deficit. Therefore a total of 25 Grade 3 students in the sample were identified as having a phonological processing deficit.

**Identifying chronological age matched deficits**

Unlike phonological processing, previous research indicates that verbal memory and naming speed deficits should be based on a chronologically-aged matched basis rather than a reading-aged one. Students were considered deficient in verbal memory or naming speed if they fell in the bottom 10% within their grade on either digit span forward or letter RAN. It has been suggested that some of these deficits

<table>
<thead>
<tr>
<th>Reading Age</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Cut Off</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 1</td>
<td>32</td>
<td>-0.74</td>
<td>1.00</td>
<td>-1.24</td>
</tr>
<tr>
<td>Grade 2</td>
<td>38</td>
<td>0.02</td>
<td>0.85</td>
<td>-0.83</td>
</tr>
<tr>
<td>Grade 3</td>
<td>95</td>
<td>0.37</td>
<td>0.77</td>
<td>-0.60</td>
</tr>
</tbody>
</table>

*Note. N = 165. The N for these groups represents the number of students who fall in the inter-quartile range for reading in their respective grade-levels, not the number of participants.*
have a cumulative effect on reading abilities. For instance, if a student has one deficit, they may be able to compensate if their other cognitive processes are efficient. However, if a student has a deficiency in two or more of these processes, they are less able to compensate for these shortcomings (Wood & Grigorenko, 2001).

As a result, a $z$-score of less than -1.28 (or the bottom 10%) was utilised as the required deficit if a given student was only deficient in one of the processing areas. However, if a student was deficient in more than one processing area then a score one standard deviation below the mean, or a $z$-score of -1 (the bottom 16%), was utilised as a cut-off. Consequently, a student with a digit span forward $z$-score of -1.1 would not be identified as deficient in this area unless that student was also identified as deficient in naming speed or phonological awareness. These cut-offs were selected on the basis that they would identify an appropriate percentage of students suitable for a screening tool for a disorder that is thought to exist in 5% of the population.

**Cognitive processing deficit groups**

By applying the criteria outlined in the previous section to the entire Grade 3 sample, seven groups of students were identified as shown in Table 30. Of the Grade 3 sample, 34% were identified as having at least one cognitive processing deficit, as would be expected with such liberal screening cut-offs in place. The means and standard deviations of the seven groups on the PAT-R, SAST and RPM are also shown in Table 30.

The triple deficit group is the poorest performing group on all three tests, and the three double deficit groups all performed more poorly than the single deficit groups, as was predicted. Addressing the second research question, both mean PAT-R and SAST scores seemed to steadily decrease as the number of deficits increased. This finding does need to be interpreted with extreme caution due to the small sample sizes of the double and triple deficit groups.
Table 30

Means and Standard Deviations on the Raven’s Progressive Matrices, Progressive Achievement Test in Reading and South Australian Spelling Test Scores within Multiple Deficit Groupings

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>%</th>
<th>PAT-R</th>
<th>SAST</th>
<th>RPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Deficit</td>
<td>106</td>
<td>65.03</td>
<td>0.19</td>
<td>0.35</td>
<td>0.01</td>
</tr>
<tr>
<td>VM</td>
<td>15</td>
<td>9.20</td>
<td>-0.01</td>
<td>-0.27</td>
<td>-0.15</td>
</tr>
<tr>
<td>RAN</td>
<td>13</td>
<td>7.98</td>
<td>-0.06</td>
<td>-0.26</td>
<td>0.49</td>
</tr>
<tr>
<td>PP</td>
<td>13</td>
<td>7.98</td>
<td>-0.12</td>
<td>-0.71</td>
<td>-0.16</td>
</tr>
<tr>
<td>VM &amp; RAN</td>
<td>4</td>
<td>2.45</td>
<td>-1.21</td>
<td>-0.91</td>
<td>-0.57</td>
</tr>
<tr>
<td>PP &amp; VM</td>
<td>6</td>
<td>3.68</td>
<td>-0.71</td>
<td>-1.05</td>
<td>0.11</td>
</tr>
<tr>
<td>PP &amp; RAN</td>
<td>3</td>
<td>1.84</td>
<td>-0.84</td>
<td>-1.66</td>
<td>0.96</td>
</tr>
<tr>
<td>Triple Deficit</td>
<td>3</td>
<td>1.84</td>
<td>-1.89</td>
<td>-1.95</td>
<td>-1.47</td>
</tr>
</tbody>
</table>

\(N = 163\). Note. PAT-R = Progressive Achievement Test in Reading; SAST = South Australian Spelling Test; RPM = Rapid automatic naming deficit group; VM = Verbal Memory deficit group; RAN = Rapid automatic naming deficit group; PP = Phonological Processing deficit group; VM & RAN = Verbal Memory and Rapid Automatic Naming deficits group; PP & VM = Phonological Processing and Verbal Memory deficits group; PP & RAN = Phonological Processing and Rapid Automatic Naming deficits group; Triple Deficit = Triple Deficit group.

However, this was not the case for the RPM. The traditional double discrepancy group, with phonological processing and naming speed deficits had the highest RPM score of any of the groups, including the no-deficit group. This could help explain why the double deficit theory fits so well in to a discrepancy based model. If these deficits have a detrimental affect on reading but do not have the same effect on aptitude, then they would fit the discrepancy method of identification well. However given the small sample size of this traditional double deficit group, this can only be considered a tentative explanation.

Those students with naming speed or verbal memory deficits on their own did not have, on average, lower scores on the PAT-R or the SAST. However, phonological deficits only, resulted in slightly lower scores on the PAT-R and quite low scores on the SAST. Furthermore, research has shown that phonological deficits are a predictor of future, rather than current, reading difficulty (Byrne et al., 1992). There is less in the literature on future outcomes of naming speed or verbal memory deficits.
Because of this, students were designated LD if they had multiple or phonological deficits. This was decided as verbal memory or naming speed deficits alone did not seem to be indicative of reading difficulty, at least without longitudinal confirmation. By utilising this criterion, 18% of students would be flagged as possibly having LD. Those identified via the cognitive deficits screening tool were designated LD, while those who were one standard deviation or more below the mean on reading, but not designated LD, were designated LA. The rest of the sample was designated RA.

Twenty-two students were designated LA, providing support for the third hypothesis that some of the students not identified as LD will still be poor readers. The means and standard deviations on all tests for these groups are shown in Table 31. As noted in Chapter 3, the heterogeneous nature of the LD group, especially when chosen on the basis of ranging cognitive deficits, dictates that significance testing is not appropriate. Furthermore the groups, as separated into cognitive deficit specific groups, are too small for significance testing. Consequently, descriptive statistics will be examined in this section.

As can be seen in Table 31, the LA group has lower PAT-R scores than the LD group. This is not entirely surprising given the fact that low reading ability is the sole criterion of the LA group, and membership in the LD group is prescribed by performance on other tasks. Irregular and nonword reading, letter RAN, DSF and the SAST is lowest in the LD group and colour RAN, DSB, RPM and the PAT-R is lowest in the LA cohort. This is to be expected as three of these variables were used to ascertain group membership. One of the most interesting outcomes from this analysis is the fact that the SAST is lower in the LD cohort than in the LA cohort. This provides a contrast with the range of PAT-R scores, which is lower in the LA cohort.

The mean irregular word reading score for the LA cohort was higher than it was for the LD cohort. However, further analysis indicated that of the 22 LA students, 73% did have below average irregular word reading scores, and 55% had scores more than one standard deviation below the mean. This provided very limited support for the fourth hypothesis that orthographic processing deficits could provide an alternate explanation for poor reading when there is no evidence of LD.
Table 31

Scores on all Tests for Students with Regular Achievement, Learning Disabilities and Low Achievement as Identified via the Multiple Deficit Model

<table>
<thead>
<tr>
<th>Test</th>
<th>RA (n = 112)</th>
<th>LD (n = 29)</th>
<th>LA (n = 22)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irregular</td>
<td>0.42 (0.64)</td>
<td>-1.12 (1.07)</td>
<td>-0.68 (0.91)</td>
</tr>
<tr>
<td>Nonword</td>
<td>0.49 (0.58)</td>
<td>-1.53 (0.53)</td>
<td>-0.46 (0.91)</td>
</tr>
<tr>
<td>CRAN</td>
<td>0.14 (0.93)</td>
<td>-0.29 (1.07)</td>
<td>-0.34 (1.11)</td>
</tr>
<tr>
<td>LRAN</td>
<td>0.24 (0.91)</td>
<td>-0.71 (1.07)</td>
<td>-0.30 (0.82)</td>
</tr>
<tr>
<td>DSB</td>
<td>0.12 (1.04)</td>
<td>-0.21 (0.96)</td>
<td>-0.33 (0.74)</td>
</tr>
<tr>
<td>DSF</td>
<td>0.16 (0.91)</td>
<td>-0.42 (1.30)</td>
<td>-0.28 (0.76)</td>
</tr>
<tr>
<td>RPM</td>
<td>0.17 (0.95)</td>
<td>-0.18 (1.00)</td>
<td>-0.65 (0.94)</td>
</tr>
<tr>
<td>PATR</td>
<td>0.45 (0.72)</td>
<td>-0.65 (0.93)</td>
<td>-1.42 (0.28)</td>
</tr>
<tr>
<td>SAST</td>
<td>0.38 (0.83)</td>
<td>-1.04 (0.81)</td>
<td>-0.57 (0.77)</td>
</tr>
</tbody>
</table>

N = 163. Note. RA = Regular Achievement; LD = Learning Disability; LA = Low Achievement; Irregular = Irregular word reading; Nonword = Nonword reading; CRAN = Rapid Automatic Naming Colour; LRAN = Rapid Automatic Naming Letter; DSF = Digit Span Forward, DSB = Digit Span Backward; RPM = Raven’s Progressive Matrices; PAT-R = Progressive Achievement Test in Reading; SAST = South Australian Spelling Test

In order to ascertain what students were identified on the basis of the multiple deficit operationalisation of LD, the new groupings were compared to that of the regression-adjusted discrepancy method of identification. This particular discrepancy operationalisation was selected out of the two utilised in this study, as it was considered to be closer to the ‘real’ construct of LD because it controls for regression to the mean (Willson & Reynolds, 2002). The correspondence between the two groups, shown in Table 32, is fairly good, with 79% of the students allocated the same way by either method as compared to the 52% that would be expected by chance. Unfortunately, the low cell sizes in some of the cells prevent a chi-square analysis, but the theory that the two methods would identify similar groups of students is tentatively supported by this result. Furthermore, unlike the DFA defined groups, in which similarity to the original groups mainly stemmed from many students being identified in the larger RA group, this method seemed to provide more accuracy for the LD and the LA groups.
Table 32

Cross Tabulation of Students with Regular Achievement, Learning Disabilities and Low Achievement via the Regression-Adjusted Discrepancy Method and the Multiple Deficits Method.

<table>
<thead>
<tr>
<th></th>
<th>RAD Groups</th>
<th>MD Groups</th>
<th>RA</th>
<th>LD</th>
<th>LA</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>109</td>
<td>LD</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>112</td>
</tr>
<tr>
<td>LD</td>
<td>16</td>
<td>LA</td>
<td>14</td>
<td>8</td>
<td>2</td>
<td>29</td>
</tr>
<tr>
<td>LA</td>
<td>0</td>
<td>Total</td>
<td>28</td>
<td>10</td>
<td></td>
<td>163</td>
</tr>
</tbody>
</table>

N=163. Note. RAD Groups = Regression-adjusted discrepancy-defined groups; MD Groups = Cognitive Deficit Defined Groups; LD = Learning Disability; LA = Low Achievement.

Summary

In this chapter the cognitive processes of phonological processing, orthographic processing, naming speed and verbal memory have been assessed so that their relationship to reading in general and LD in particular, can be determined. A cluster analysis examining the relationship between these variables and spelling was also performed. However, the results indicated that these processes were not those that differentiated a group of struggling students from their average or high performing counterparts. In contrast, a similar analysis on reading found that irregular and nonword reading, letter RAN and digit span forward, differentiated between poor and average readers well.

A comparison of LD, LA and RA groups, as allocated by the traditional and regression-adjusted discrepancy methods followed. The first traditional discrepancy-defined groups found that there were no significant differences in means between LD and LA groups on any of the cognitive processes. The second analysis found that the only difference between the LD and LA groups when they were allocated by the regression-adjusted discrepancy method was in irregular word reading, which was an unexpected result.

In the first DFA analysis using nonword reading, RAN, and digit span tests to predict membership in the traditional discrepancy-defined groups, the cognitive
processes correctly predict membership in 77% of cases. Furthermore, a series of ANOVAs comparing the newly-defined LD, LA and RA groups found that the only difference between the LD and LA groups was that the LA group scored significantly lower than the LD group on average on the digit span forward test. This suggests that although the LD group does suffer from a double deficit, the LA group may have a triple deficit.

In the second DFA analysis, using the same five variables to predict membership in the regression-adjusted discrepancy groups, prediction into the LD, LA and RA groups was successful in 82% of cases. This was despite not allocating any students into the LA group. Because this analysis resulted in only two groups, namely LD and RA, these two groups were compared through independent sample t-tests and were found to be significantly different on all measures except for the colour RAN and RPM.

Finally, a new method of screening, based only on cognitive processing deficits was suggested. Students in the double deficit or triple deficit groups had lower mean reading scores than their single or low deficit counterparts. Students with phonological or multiple deficits were selected in the LD group, while the LA and RA criterion remained the same. Although a very liberal 18% of the cohort was identified as possibly LD, it is important to remember that this is a screening tool. Comparisons with the regression-adjusted discrepancy method of identification suggest that this method is consistent with the current literature and, as such, this may be a more efficient manner in which to screen for LD students. The efficacy of this method, and all other results, are discussed in Chapter 7.
Chapter 7

Overview
In Chapters 5 and 6 the results from the current study were presented. In this chapter these results are discussed in regards to how they relate to the research questions and hypotheses put forward in Chapter 1 and expanded in Chapter 3. The cognitive processes tested in the current study, namely phonological processing, naming speed, verbal memory and orthographic processing, and their relationship to both reading and Learning Disabilities (LD) are each discussed. This is primarily done through an examination of the results from the comparisons of sets of groups dictated by various methods of identifying students with LD, Low Achievement (LA) or Regular Achievement (RA).

In Chapters 2, 3 and 6, a multiple deficits model of LD identification was put forward. In order to generalise the results of this model with that of previous research, the discrepancy groupings of the same sample were also examined. In this chapter, a new interpretation of previous results found using the discrepancy method of identification is put forward. This reinterpretation is based primarily on the theory that many of the students identified as LA in the past were actually LD students. Their cognitive processing deficits resulted in a reduced IQ score, which in turn decreased the likelihood of those students being identified as LD. Furthermore, many of those identified as LD through the discrepancy method did not have the cognitive processes inherent to the disorder. Consequently, their difficulty in reading could be attributed to other causes.

The next section begins with a discussion of the multiple deficits model of LD. Support for this model, as well as the areas in need of further testing are discussed. This is then followed by an examination of how this model of identification compares to previous models, including discrepancy, Response To Intervention (RTI), phonological deficit and double deficit models. The implications for practice, policy and theory of a multiple deficits model of LD identification are then discussed. Finally the limitations of the current study, along with suggested areas of future research, are presented.
**Research Questions and Hypotheses**

*Can the psychometric validity of the newly-developed irregular word and nonword lists be supported as per the tenets of both Classical Test Theory and Item Response Theory?*

As discussed in Chapter 4, the psychometric validation of the word lists was conducted within the Item Response Theory (IRT) and Classical Test Theory (CTT) paradigms. CTT, the more common test theory, states that all test scores are equal to the true score on the test plus error (Novick, 1966). In IRT, the true score on the construct rather than the test is measured, and participants and items are measured on the same scale (Bechger et al., 2003). IRT is thought to be superior to CTT in test measurement, however, the use of both is recommended (Edelen & Reeve, 2007). Although both tests were found to be psychometrically sound via the tenets of IRT, the irregular word list did not fit all the assumptions inherent to CTT, primarily due to the ceiling effect in this scale.

Similar to many scales that have been developed with the aim of screening for abnormal behaviour in a normal sample, skew was required if the word lists were appropriately targeted to the population of interest. A normally distributed scale ensures that there are enough questions to separate those students at a higher functioning level, as well as those at the lower end of the continuum. However, these extra items would not be beneficial as it is only the lower performing students that the test was developed to identify. Nevertheless, as the scales were used in analyses based on the tenets of CTT, it is important to ensure that extreme skew is unlikely to affect these analyses by violating statistical assumptions. Ultimately, both lists gained psychometric support as per the tenets of IRT, and beside the ceiling effect in the irregular word list, as per CTT as well.

The irregular word list had good model fit and person separation index. Furthermore, the test was shown to be unidimensional both through exploratory factor analysis and Rasch analysis. Regardless of its eventual use as a predictor of LD or as an alternative explanation to poor reading in the case of non LD students, the irregular word list will be useful in identifying students with poor orthographic processing.
The nonword list was less skewed than the irregular word list. Furthermore, 20 items rather than 15 were retained. This was because the identification of phonological deficits was based on reading, rather than chronological, age. The use of reading-age based deficits meant that differentiation needed to be made at varying points on the nonword list. In the irregular word list there was only one cut-off score, regardless of reading proficiency. Consequently the cut-off points varied between reading-ages and it was important that there was sufficient spread at a higher level of the nonword list than that of the irregular word list. The nonword list demonstrated unidimensionality and reliability, as well as meeting the psychometric requirements of Rasch analysis, in which model fit and the PSI were both good.

An interesting result to come out of the Rasch analysis was that the three-level scoring structure was not appropriate. Originally, the idea was to award students who had managed to get the stem of the word correct with an error in the suffix. However, the analyses indicated that this extra level of scoring provided no information that the other two scores did not already provide. However, it is possible that differentiating between these kinds of suffix-based errors and other errors may provide more information in a younger cohort.

Similar to the work of Edwards and Hogben (1999), simple irregular and nonword lists were found to be psychometrically sound measures. However, in contrast to the Edwards and Hogben scales, these scales were specifically targeted to poor Grade 3 readers, resulting in shorter lists and more differentiation between targeted students. Overall, the results from Chapter 5 provided support for the psychometric validity of the irregular and nonword lists, with the caveat that the ceiling effect of the irregular word list needed to be taken into account in CTT analyses. If the irregular word list is required in an analysis that is not robust in regards to normality, then it may not be suitable for use.

*Can tests of cognitive processing deficits be used to identify a similar cohort of LD students as would be identified via the discrepancy method?*

This hypothesis was supported by the Discriminant Function Analysis (DFA) predicting both traditional and regression-adjusted discrepancy group membership. Phonological processing, naming speed and verbal memory could successfully predict 77% of students into their traditional discrepancy-defined groups, as
compared to the 56% that would be expected by chance. Furthermore, these same processes could successfully allocate 82% of students into regression-adjusted discrepancy-defined groups, despite no LA students being identified via this method. Finally, when students were allocated via cognitive processing deficits only, 79% of the students were still allocated into the same groups that they would have been via the regression-adjusted discrepancy method. It would be expected that both of these allocations, if dictated by chance alone, would correctly allocate 63% of students.

The finding that cognitive processing deficits could predict the presence of possible LD is consistent with previous studies that have used similar cognitive processes to the same end (e.g., Chung et al., 2010; Hatcher et al., 2002; Johnson et al., 2010; Puolakanaho et al., 2007). This provides convergent validity for both LD as a neurological disorder and the ability of the discrepancy method to identify it. Furthermore, longitudinal studies that have used these same processes to predict future reading problems (Adlof et al., 2010) indicate that this model would work in the long term as well.

The fact that this multiple deficit model could be used to identify LD students can be considered as either support for, or in contradiction of, other multiple deficit models. For instance, the triple deficit hypothesis (Badian, 1997; Stage et al., 2003) was similar to the model put forward in the current study but used orthographic processing deficits instead of verbal memory deficits to identify LD. Furthermore, the cognitive hypothesis testing model (Fiorello et al., 2006) and the multiple deficit hypothesis (Ho et al., 2002) also utilised orthographic processing deficits as a marker variable as well other deficits. Although the deficits utilised in the current study were different, the theory put forward that LD can be potentially identified using the cognitive processes was supported. The most consistent difference between the model put forward in the current study and previous models was the use of verbal memory deficits as a predictor of LD, and orthographic processing as an alternative explanation for reading difficulties.

Are the identification methods and cognitive processes currently researched in regards to reading-based LD relevant to spelling-based LD?

At first glance, there was little support for the applicability of the theories put forward for reading-based LD as relevant to spelling. Primarily, the low correlation
of the South Australian Spelling Test (SAST) and the Raven’s Progressive Matrices (RPM) indicated that a spelling-based discrepancy analysis was not appropriate. Significant discrepancy between two variables with a correlation of .21 could be considered an outcome of orthogonal variables rather than a neurological disorder.

Additionally, the results from the spelling-based cluster analysis were not as informative as those from the reading-based analysis. In the latter, there was a small group of poor readers with correspondingly low cognitive processing skills. In the former, the cohort was split roughly in two, providing much less information. Although the good spelling group scored significantly higher on all variables than the poor spellers, this is to be expected when the cohort is split into simple high and low performing groups. These two factors combined led to the conclusion that these methods, while appropriate for reading-based LD, were not appropriate for the spelling equivalent.

Interestingly, this research question on spelling-based LD does need to be revisited in the context of the descriptive statistics of the groups identified via the multiple deficits method. The LD cohort identified via the multiple deficits method had very low spelling scores, lower than those of the LA cohort. This was the opposite of what was found for the Progressive Achievement Test in Reading (PAT-R). Therefore the method utilised to identify future reading difficulties effectively identified current poor spellers. This finding has limited applied utility for identifying poor spelling, as a simple spelling test can identify current spelling difficulty. However, there are some theoretical interpretations of this finding that deserve scrutiny. One interpretation is that some students, who have masked their reading difficulties in early reading years, may experience trouble with spelling before their difficulties in reading surface. This is an interpretation worthy of examination in future studies as the SAST is freely available and already administered in many Australian schools. The SAST may not only be of use in assessing current spelling ability, but may also provide some insight into future problems with reading-based LD.

Although the model put forward for reading-based LD did not apply to the spelling equivalent, this does not mean that these cognitive processes are not relevant to spelling-based LD at all. The finding that low scores on the SAST for the multiple
deficit model-identified LD students supports previous results indicating that naming
speed (Savage, Pillay et al., 2007) and phonological processing (Pennington et al.,
2001) can predict spelling-based LD. It may be that a different method of
identification, on which convergent validity could be based, needs to be devised as
the correlation between spelling and aptitude, as measured in the current study, were
too low for a discrepancy model to be applied.

Are some of the students identified as LA by the discrepancy method of identification
actually facing many of the same cognitive processing difficulties that are faced by
LD students?

In the traditional and regression-adjusted discrepancy-defined groups, there were no
significant cognitive processing differences between LD and LA students, with one
exception. Surprisingly, LD students had significantly lower irregular word reading
scores than their LA counterparts when the groups were regression-adjusted
discrepancy-defined. However, there were no differences between LD and LA
students in phonological processing and naming speed found in the current study,
both of which are traditionally linked to LD.

The significantly lower irregular word reading score in the DFA-defined LD cohort
was wholly unexpected. One possible explanation is that the line separating LD and
non LD students on an aptitude/achievement scatter plot flattens as a consequence of
the regression-adjusted discrepancy method. This results in the identification of
more students with low IQ and low reading as LD, instead of high IQ students with
average reading scores. Consequently, the reading achievement of the LD students
identified is lower than it would be if they were identified through a traditional
discrepancy method. Because irregular word reading is possibly a reflection of
reading achievement, rather than a causal factor, this could be considered a simple
reflection of the changed groupings if it were not for the fact that all LA students are,
by definition, poor readers. This would explain a decrease in reading achievement
scores in a regression-adjusted discrepancy-defined LD cohort as compared to a
traditional discrepancy-defined cohort, but not compared to an LA group.

However, to say that low irregular word reading scores in regression-adjusted
discrepancy-defined LD groups support the theory that orthographic processing is a
deficit causing LD is also not necessarily justified. This is because none of the more
widely accepted deficits that cause LD were significantly lower in the LD cohort in this analysis. For instance, given longitudinal research on the lack of long-term reading problems stemming from orthographic deficits (Byrne et al., 1992), it could be argued that this is a reflection of orthographic deficits resulting in a discrepancy between aptitude and achievement, rather than the deficits causing a neurological disorder. Therefore, the role of orthographic processing, as a reflective or formative variable of LD, needs further investigation.

Orthographic processing aside, and in line with similar studies in the past, no differences between discrepancy-defined LD and LA groups were found in the current study (Fletcher et al., 1994; Gough & Tunmer, 1986; Share, 1996; Stanovich & Stanovich, 1997). This is in contrast to those that have found that there are significant differences between the two groups (Decker, 2008; Dickerson et al., 2007) or those that have found that the LA group has significantly lower naming speed scores than the LD group (Bone et al., 2002). One interpretation of this finding is that deficits in these processes are actually a symptom of poor reading achievement rather than a neurological disorder. However, the utility of phonological deficits as a predictor of disordered reading, as compared to orthographic processing (Byrne et al, 1992), indicates that this may not be the case. Another interpretation is that there are students with LD in both cohorts and as such, the discrepancy method is not differentiating between the two cohorts.

*That deficits in verbal memory will be indicative of LD in the same manner as phonological processing and naming speed.*

The predictive value of verbal memory of discrepancy defined groups was similar to that of naming speed deficits. Furthermore the reading abilities of students with multiple deficits were consistent, regardless of the presence of verbal memory deficits. Therefore, overall, this hypothesis was indeed supported. Verbal memory deficits resulted in similar reading patterns as naming speed deficits. Furthermore, both processes seemed to have additive value when paired with other cognitive processing deficits. Although the verbal memory deficit group was not considered LD in the final multiple deficits model, the naming speed deficit group was not either.
This supports previous research that has led the authors to conclude that memory deficits may play a role in LD (Brown & Greaves, 2001; Gang & Siegel, 2002; Swanson et al., 2010). Worthy of note is the finding that not all students identified as possibly LD through the multiple deficits model had verbal memory deficits. Furthermore, there were students identified as LD that did not have phonological or naming speed deficits as well. This heterogeneous nature of LD can make results difficult to interpret and could make the role of processes such as verbal memory more difficult to establish.

*That increasing numbers of cognitive processing deficits will be positively associated with reading difficulty.*

This hypothesis was strongly supported by the results of the current study in which those students with deficits in all three cognitive processes tested had the lowest average score in reading. Additionally, those students with deficits in two of the three processes had higher average reading scores than those with three deficits, but lower average reading scores than those with one. This is in line with previous research that has found that multiple deficits result in compounded difficulties in reading (Shaywitz & Shaywitz, 2005; Wolf & Bowers, 1999). This result lends support to the theory that all of these processes rely on both shared and unique genetic variance (Haworth et al., 2009). For instance, naming speed and phonological deficits would be the result of some common and unique genetic input. Consequently, it would make sense that deficits in more than one of these processes would not only co-occur more than would be dictated by chance alone, but also that the difficulty with reading would be compounded by multiple deficits. This interpretation could also go a long way to explaining the difficulty in establishing the unique contribution of multiple processes. Studies that show that the variance in reading between cognitive processes is shared (Phinney et al., 2007), or different processes provide unique variance (Kirby et al., 2010; Savage & Frederickson, 2005), could be accommodated by this interpretation.

*That some of the students not identified as having cognitive processing difficulties will still be poor readers.*

After the students that would be designated as possible LD through the multiple deficits model were allocated, there were still 29 out of the 141 remaining students...
with reading scores that were one standard deviation or more below the mean. Therefore this hypothesis was supported. For the purposes of comparing the results from this analysis with those identified via the discrepancy method, these students were designated as LA.

Not only was this hypothesis supported by the results of the current study, but it is also an important distinction to make in all LD research. As discussed in Chapter 2, it is not enough to pick the poorest readers in a class when looking to identify a LD cohort; it must be those whose reading difficulties are caused by a cognitive processing deficit or deficits. Therefore, it should follow that some of those students not identified by this process will still be poor readers and this was the case in the current study.

Of course the problem with this finding is that there is no way in the current study to differentiate between those students who were correctly and incorrectly identified as LA. This is because definitive identification is not possible for a disorder without a widely accepted operationalisation, and especially for a large cohort without access to individual testing. Consequently, it should be said that this hypothesis received limited support because even though the results of the current study are congruent with this hypothesis, there are other plausible explanations for these results, especially if this model of identification is not accepted as accurately identifying genuine LD students.

*That orthographic processing deficits will provide an alternative explanation for poor reading if LD is not found.*

This hypothesis was only partially supported by the data. Nearly three-quarters, 73%, of the multiple deficit-defined LA group had below average irregular word reading scores. However, only 55% had scores more than one standard deviation below the mean, the standard at which deficits have been identified in the current study. Furthermore, although the lowest average score for the multiple deficit-defined LA group was in irregular word reading, the LA group had higher irregular word reading scores than the LD group, once again indicating that there may be orthographic deficits in LD.
Therefore, despite nearly three quarters of poor readers without LD having below average reading scores, the theory that orthographic processing deficits could provide a simple contrast for LD identified students has not been supported. As will be discussed later in the chapter, the role of orthographic processing in LD is unclear and worthy of future research.

This section has reviewed the results from Chapter 5 and 6 in light of the hypotheses and research questions put forward in Chapter 3. In the next section the cognitive processes examined in the current study, namely phonological processing, naming speed, verbal memory and orthographic processing, and how they relate to both reading and LD, are discussed.

*Reading, Learning Disabilities and Cognitive Processing Deficits*

In Chapter 6, there was an examination of the relationship between cognitive processes and reading achievement, with a focus on poor readers in general and LD students in particular. However the aim here is not to predict present poor readers, the PAT-R alone could be utilised for this purpose. Instead this model was developed to pick up students with LD, even if their disorder has not yet begun to affect their reading scores. This next section will examine each of the cognitive processes tested, starting with phonological processing, and look at how they relate to both reading and LD.

**The role of phonological processing in reading and learning disabilities**

Phonological processing, as measured by the refined nonword list, was the most important cognitive processing variable in the current study, as would be expected given the previous research in this area (e.g., Bach et al., 2010; Bekebrede et al., 2009; Caylak, 2010). There was a high correlation between nonword reading and PAT-R scores in the current study. Interestingly, the correlation was not as high as it was between irregular word reading and PAT-R scores. This could be possibly attributed to the similarity in the skills required for the irregular word list and the PAT-R, as there would have been few novel words in the PAT-R. Alternatively, this finding could be explained by previous longitudinal studies that have found that phonological processing is a good long term predictor of reading (e.g., David et al., 2007; Linklater et al., 2009; Muter & Snowling, 2009; Storch & Whitehurst, 2002; Svensson & Jacobson, 2006; Torgesen & Wagner, 1994).
As expected, phonological processing was significantly lower in the poor readers, as identified by the cluster analysis, than in the average readers. Furthermore, phonological processing was significantly higher in the RA group than in the LD and LA groups, in both the traditional and regression-adjusted discrepancy-defined analyses. There was a non-significant trend for LD students to be significantly lower in phonological processing than their LA counterparts in the regression-adjusted discrepancy groups. In the traditional-discrepancy DFA-defined groups, the RA group was significantly better at nonword reading than their LD and LA counterparts; however there was no significant difference between LD and LA groups in nonword reading in this set of groups. Overall, these results support research that states that phonological processing is lower in poor readers than in good readers (David et al., 2007; Linklater et al., 2009). However, there is only evidence for phonological processing deficits specific to LD if they are used in the operationalisation of the disorder, not when LD students are discrepancy-defined. However, recent research on phonological deficits in LD (e.g., Adlof et al., 2010; Linklater et al., 2009; Muter & Snowling, 2009) has provided theoretical justification for the use of phonological deficits as a marker of LD.

When identifying students with phonological deficits, reading-age, rather than grade-age based deficits, were used in this study. This may have impacted on the results from the multiple deficit model dictated groupings. The phonological deficit only group had average rather than low reading scores. However, if grade-age based deficits were used instead of reading-age based deficits, the phonological deficit group would have presumably had lower reading scores. This is because there would not have been lower deficit cut-offs put in place for those who were reading at a Grade 1 or 2 level. Students with phonological deficits only were selected in the LD cohort, even though those with only one of naming speed or verbal memory deficits were not. Furthermore phonological deficits along with all deficits had a cumulative effect on reading scores when there was more than one deficit present.

**The role of naming speed in reading and learning disabilities**

Both tests of naming speed, the colour and letter Rapid Automatic Naming (RAN) tests, were found to have good test-retest reliability, but internal consistency could not be tested for the two single item scales. Colour RAN had a low correlation with the PAT-R, and the correlation between letter RAN and the PAT-R was moderate.
The letter RAN test was consistently a better predictor than the colour RAN test in the analyses in Chapter 6.

In both sets of discrepancy-defined groups, there were significant differences between groups on the letter RAN test, but not the colour RAN. Furthermore in the planned contrasts for both sets of discrepancy-defined groups, there were significant differences in the letter RAN between RA students and their LD and LA counterparts combined, but not between LD and LA students. The same result was found in both DFA-defined sets of groups.

Support was found for previous research demonstrating that RAN tests made up of non-alphabetic stimuli, such as colours, become less effective predictors of reading once a certain level of reading proficiency is reached (Misra et al., 2004; Wolf et al., 2000). Colour RAN was not significantly different between poor and average readers as allocated by the cluster analysis, nor was it significantly different within traditional or regression-adjusted discrepancy-defined groups. It should be noted in the case of the two discrepancy-defined group analyses that there was a trend towards differences in colour RAN but, after Bonferroni correction, these differences were not significant. Meanwhile, in the same analyses, there were significant differences in letter RAN scores between groups. Furthermore, colour RAN was not significantly different between the two sets of DFA-defined LD, LA or RA groups. Consequently, only letter RAN deficits were used to signify a naming speed deficit. However, this may not be the best course of action if the same study was replicated on a younger cohort.

Overall, naming speed deficits contributed well to the multiple deficit model, both double deficit groups with naming speed deficits received low reading scores. The naming speed deficit only group, however, had a similar mean reading score to the no deficit group. This could indicate that, similar to phonological processing, naming speed deficits are indicative of future difficulties, or, alternatively, that on their own, naming speed deficits are not sufficient to warrant a designation of a neurological disorder. In this study, the latter was assumed correct as there is little previous research available on the long term effects of naming speed deficits, controlling for verbal memory deficits. Previous research on the naming speed deficit only students would have included participants with verbal memory deficits.
As per the results of the current study, this is a different group of students and their presence could have confounded results. Therefore, these students were not included in the LD cohort; however this decision would need to be confirmed by further research.

The Role of verbal memory in reading and learning disabilities

The correlation between verbal memory and reading was not as strong as it was between naming speed or phonological processing and reading, verbal memory still fits into a multiple deficits model well. The correlations between both digit span measures and the PAT-R and SAST were significant, however they were weak. Poor readers, as designated by the cluster analysis, were significantly worse at the digit span forward task than average readers. Conversely, the digit span backwards task did not differentiate between poor and average readers.

In both the traditional and regression-adjusted discrepancy-defined groups, RA students outperformed LD and LA students combined on the digit span forward task. In the DFA-predicted traditional discrepancy-defined groups, LD students significantly outperformed LA students on the digit span forward. Overall these results support the theory that verbal memory deficits lead to poor reading achievement, but there is some evidence that verbal memory is lower in LA than in LD cohorts. This could be interpreted as evidence for verbal memory deficits having less to do with LD than poor reading, but it could also be taken as a challenge to the claim that the discrepancy method is appropriately delineating the two. If this first interpretation, which will be discussed later in the chapter, is assumed correct, then the results from the current study support previous research that found that verbal memory deficits play a role in LD (Brown & Greaves, 2001; Gang & Siegel, 2002; Kibby & Cohen, 2008; Siegel & Ryan, 1988; Swanson & O'Connor, 2009). The finding in previous research that LD students have both forward and digit span backward deficits (Helland & Asbjørnsen, 2004; Shaywitz et al., 1999) was only partially supported by the current study, as the results from the digit span backwards were less conclusive.

Overall, the digit span forward task seemed more relevant than the digit span backward task in the current study. In both the discrepancy-defined groups, significant differences were found in the one-way ANOVAs between the groups for
digit span forward, but not backward. In the traditional discrepancy DFA-predicted groups, LD students performed significantly better on digit span forward than their LA counterparts. Although the explanation for the success of letter RAN as compared to colour RAN had strong basis in previous research, the reason for this difference between forward and backward digit span is less clear and worthy of future research.

If the theory that digit span forward is a measure of short term memory and digit span backward a measure of working memory (e.g., Baddeley, 1986; Baddeley et al., 1998; Gathercole, 1999; Kauffman, 1994; Savage, Lavers et al., 2007) is accepted, then the results from the current study could be interpreted as indicating that short term memory deficits are more important to LD than working memory. However, an alternative explanation is based on the contrasting research that has suggested that digit span backwards, similar to digit span forward, is actually a measure of short term memory, not working memory (Cantor et al., 1991; Conway et al., 2005; Engle et al., 1999; Richardson, 2007) or influenced by other factors such as attention (Hale et al., 2002). If either of these theories are correct then it may be that a more direct working memory measure may still have a role to play in LD identification.

The role of orthographic processing in reading and learning disabilities

In the current study orthographic processing, as measured by the newly refined irregular word list, had a very high correlation with both the PAT-R and the SAST. This could be interpreted as orthographic processing contributing to or merely being a reflection of reading achievement. The fact that the difference between poor and average readers, as identified by the cluster analysis, was highly significant could also be interpreted in the same way. It should be noted that these results were found despite the ceiling effect in the irregular word list that would presumably result in a reduced correlation between irregular word reading and any other variables.

In both sets of the DFA-defined groups, the RA group was significantly better at irregular word reading than the LD and LA counterparts. In the traditional discrepancy-defined groups, no significant mean differences were found between the LD and LA groups on the irregular word list. However when testing for differences between regression-adjusted discrepancy groups, the RA group was significantly
better at irregular word reading than the LD and LA group combined and the LD group was significantly worse than the LA group.

The fact that the irregular word list was the only task that was significantly different for the LD and LA groups in the regression-adjusted discrepancy-defined groups suggests one of four scenarios. The first option is that the separation as per the discrepancy theory is not as it should be and, as such, the two groups are not representing the constructs that they are meant to represent. As such it may be that those students that do have a discrepancy between aptitude and achievement, who are not necessarily LD, are also those that are more likely to have orthographic processing deficits. While this interpretation is not particularly likely as there are quite a few assumptions made in this explanation, it is possible nonetheless.

Another explanation is that orthographic processing is a cognitive processing deficit that causes LD, similar to phonological processing, naming speed or verbal memory. This interpretation would be consistent with previous research (e.g., Badian, 1997; Fiorello et al., 2006; Ho et al., 2002; Stage et al., 2003). If this is the case, then orthographic processing deficits can easily be added to the multiple deficits model. However confirming this would be difficult, as orthographic processing is also considered to be a reflection of exposure to print or environmental disadvantage (Braten et al., 1999; Castles et al., 1999; Griffiths & Snowling, 2002), a confounding factor in a lot of LD research. Alternatively, it could be that orthographic processing deficits are important to LD, but that a different test, that does not have the confounding factor of exposure to print, should be used to test for this.

This leads to the third explanation, and that is that poor orthographic processing is an outcome of LD, not a cause. Ultimately longitudinal studies would be needed to confirm the causality of the relationship between LD and orthographic processing. However the longitudinal study by Byrne and colleagues (1992) that found that students with orthographic deficits caught up to their non-deficit peers, without intervention, provides some support for this interpretation.

Finally, another explanation is that LD students actually have difficulty with a specific aspect of orthographic processing, and both this and exposure to print is assessed by the irregular word list. For instance, some researchers have suggested that LD students experience difficulty with integrating phonological and
orthographic information (Cao et al., 2008; Desroches et al., 2010). If this is the case then it may be that a test that can examine this ability, without the confounding factor of exposure to print, could be of use in a multiple deficits model. This issue with the measurement of orthographic processing is a real one, and therefore more studies using different measures of orthographic processing, for instance the orthographic choice task (Bowey & Muller, 2005) are needed to determine which of these scenarios is most likely.

The finding that the LD students had such low irregular word reading scores does seemingly contradict the theory put forward by some researchers that orthographic processing is a compensatory strength in LD readers (e.g., Lavidor et al., 2006; Price & Mechelli, 2005). However, once the heterogeneous nature of LD is taken into account, it is difficult to be certain about even this, as it may be that orthographic processing is high in a minority of the LD readers. For instance, there may be a group with a certain deficit or combination of deficits that enable orthographic processing to be developed as a compensatory strength. If this was the case then studies that found high orthographic scores in LD groups may have been using a method of identification that made identifying students with those particular deficits more likely.

The interpretation that the irregular word reading list reflects to some extent exposure to print does not mean it is not worth measuring in an LD screening tool. In order to ensure that the battery is as useful as possible in a classroom setting, the irregular word list is important for providing an alternative explanation for any reading difficulties that a student may face. Although not all LA students identified were below average on irregular word reading, the majority were. It could be that, if a student has poor reading scores and an orthographic processing deficit, but does not have other cognitive processing deficits, that their problems may be of a more environmental nature. Therefore, it is reasonable to suggest that they may be more readily remediated than their LD peers.

This section has examined the influence of the cognitive processes tested in the current study, on both reading and LD. These results, especially as they pertain to cognitive processing features of LD and LA students, can only be accommodated
with a reinterpretation of the discrepancy theory. This will be outlined in the next section.

**A Reinterpretation of the Discrepancy Theory**

In the current study, the discrepancy method of identification was used as the basis of comparison for an investigation into a multiple deficits model of LD. There were multiple comparisons made to gauge the uniformity of the two methods of identification. First, a DFA was calculated with phonological processing, naming speed and verbal memory used as predictor variables. These cognitive processes predicted traditional and regression-adjusted discrepancy-defined groups, with 77% and 82%, respectively, of students correctly allocated. Finally, there was an agreement of 79% of LD, LA and RA groupings as classified by the regression-adjusted discrepancy method and the multiple deficits method. This is quite good given that other studies have found a correspondence between methods of 48-70% (Proctor & Prevatt, 2003).

However, it is important to keep in mind the fact that only 77 to 82% of discrepancy-defined students were correctly allocated into groups using cognitive variables. The success, or lack thereof, of using cognitive processing deficits to gain similar LD, LA and RA allocations to the discrepancy method can be viewed in a number of ways. First of all, it could be seen as a failed attempt. There are no LA students in the multiple deficit model that are then allocated as RA in via either discrepancy method, or vice versa. This is because the allocation of these two groups is based, in part, on reading scores being above or below a $z$-score of negative one. Consequently, the only difference lies in how the LD students are identified. Therefore, the fact that 18% to 23% of the sample were incorrectly allocated could be seen as a failure to identify LD via the multiple deficits method.

The major problem with any of these interpretations is that what is being investigated here is LD and LA groups as they should be, representing groups with different neurological make ups, rather than as they are, which is an approximation of these same groupings. There are LD and LA readers in any given sample of poor readers and intelligence alone is not what actually separates them, instead it is the presence of a neurological disorder. However, those with LD are more likely to have an IQ inconsistent with their reading achievement than their LA counterparts. This
is the basis of the discrepancy method. In reality, within a group of poor readers, there are probably LD and LA readers on both sides of a discrepancy based cut-off. This is of course the reason for many of the conflicting results in this field of research. But, if for a moment, the discrepancy groupings as they are are discussed, it could be suggested that many of those in the LD group are made up of the traditional phonological or double deficit students that have been identified in the past as LD students. Deficits in these two cognitive processes may have less impact on aptitude test scores than verbal memory. Despite the average mean RPM scores in the naming speed deficit only or naming speed and phonological deficit groups, the RPM score of those with both naming speed and verbal memory deficits or both phonological processing and naming speed deficits was low. This finding could help explain why verbal memory deficits have not been identified in the past when discrepancy-defined groups were the basis for comparison. Verbal memory deficits, when combined with one or more deficits, seem to bring aptitude scores down too low for a significant discrepancy between aptitude and achievement to be viable. Consequently, students with naming speed and phonological deficits could fit the traditional discrepancy model. This is supported by the high average RPM score in the double deficit cohort identified via the multiple deficits method.

If the discrepancy method of identification is seen as a starting point, rather than a definitive way of differentiating LD and LA students, the identification of low IQ students with verbal memory deficits can be interpreted as the multiple deficits model correctly allocating students who were misclassified by the discrepancy method. This, of course, would need longitudinal verification. However there is justification for the theory, as the differently-classified students may have been misclassified on the basis of established criticisms of the discrepancy method. In order to place this model into context with previous research on the discrepancy method, it is worthwhile examining those students who were differently classified by the multiple deficit and regression-adjusted discrepancy methods in order to speculate as to why that might be.

There were three students identified as LD via the discrepancy method, but allocated into the RA group via the multiple deficits model. It is possible that these students had a significant discrepancy between aptitude and achievement for reasons other than LD. In the discrepancy method, students with average reading scores, and no
underlying issues causing this, were identified as LD if their non verbal IQ was sufficiently high. The only reason that this average achievement was considered cause for concern is because the discrepancy method of identification dictated that it was a problem. However, there were no cognitive processing deficits in these students indicating the presence of LD. Therefore these students may be those who are not living up to their potential for any of a wide range of reasons other than LD, such as problems at home, motivation or issues at school.

The same explanation generally applies to the 14 students identified as LD via the discrepancy method but LA via the multiple deficits model. The only real difference lies in the fact that these students are poor readers, but as previously discussed there are a multitude of explanations for poor reading achievement besides LD (Coutinho et al., 2002; Davis-Kean, 2005; Metsala et al., 1996). For instance, 55% of these students had irregular word reading scores one standard deviation or more below the mean, a finding that has been linked to a lack of exposure to print (Braten et al., 1999).

Sixteen students were identified as RA via the discrepancy method, but LD via the multiple deficit model. It seems that these students have been able to, so far at least, compensate for their cognitive processing deficits sufficiently to avoid a significant discrepancy between aptitude and achievement. This is a group of students that have been identified in the past as those that may be overlooked by the discrepancy method of identification (Gustafson & Samuelsson, 1999). The ability to identify these students before their reading skills decrease in relation to their aptitude is one of the primary reasons that this method of identification is preferable to discrepancy or RTI. This is because it is important that students are identified before compensating for their difficulties becomes untenable and their reading abilities slip in comparison to their peers and IQ.

The two students identified as LA via the discrepancy method, but LD via the multiple deficits model, probably represent the most interesting cohort in the current study. These students were presumably not eligible for LD diagnosis as their aptitude scores were too low for a significant discrepancy between aptitude and achievement. These are the students whose deficits may or may not be contributing to their low IQ but, more importantly, are suffering from the same deficits as their
traditionally-defined LD counterparts. It is argued here that the students would benefit from remediation based on accommodating these deficits as much as their traditionally-defined LD counterparts. This will also reduce the perceived unfairness in countries with federal recognition of LD that allocate funds and remediation only to those that are considered to be intelligent enough to have LD (Stanovich, 1999).

The low number of students that were LA via the discrepancy method and LD via the multiple deficits model, or LD via the discrepancy method but RA via the multiple deficits model, is a reflection of the type of discrepancy method selected and not a reflection of the types of students identified by the multiple deficits model. If the traditional discrepancy method was selected as the basis of comparison, more high aptitude students would be identified as LD, increasing the number of students moving from LD to RA designations between the two models. Furthermore, more students were identified as LA in the traditional discrepancy method which, in turn, would have resulted in more students in the LA group being designated LD via the multiple deficits model.

Finally, there were 11 students who were identified as LD by both the regression-adjusted discrepancy method and the multiple deficits method. These are the students that the discrepancy model was designed to identify, namely those students whose cognitive processing deficits have resulted in a significant discrepancy between aptitude and achievement. Given the low RPM scores of all students with verbal memory deficits identified as possibly LD within the multiple deficits model, these 11 students most likely had a traditional phonological processing and naming speed deficit or a phonological processing deficit only. This provides further support for the theory put forward here that verbal memory deficits have been somewhat overlooked in LD research, as naming speed and phonological processing deficits do not appear to have the same effect on IQ. Consequently, this then effects eligibility for LD identification via the discrepancy model.

As discussed in Chapter 2, a discrepancy in aptitude and achievement is a likely, but not necessary, outcome of LD. There is no reason why a child with low intelligence cannot have the same cognitive processing deficits as a student with a significant discrepancy between aptitude and achievement. There is also a myriad of reasons as to why a student may not have a reading achievement level that is commensurate
with their aptitude, and only one of these reasons is LD. Identifying students on the basis of cognitive processing deficits would transcend these issues, as identification would be orthogonal to intelligence, rather than despite, or because of it. Consequently, high IQ students may still be identified, but only if they have the cognitive deficits that cause LD, and not because they have a high IQ. Additionally, low IQ students with low reading achievement will not be identified as LD via a multiple deficits model if they do not have the processing deficits to warrant identification.

The reason for this lack of clarity is the previously mentioned problem of the interpretation of discrepancy groupings *as they should be* rather than *as they are*. Many assumptions about the construct of LD have been based on an inaccurately identified cohort of students. It is assumed that all students with a significant discrepancy between aptitude and achievement have the neurological deficits inherent to LD and that none of those in the LA groups have these deficits. Consequently, when no differences are found between the two groups, this finding has been interpreted incorrectly as no differences existing between LD and LA students, rather than a result of operationalisation. Nevertheless, there is much to support the notion that there are LD and LA groups that should be differentiated. The method of doing so, however, needs to change. In the next section the multiple deficits model, which is designed for this purpose, is discussed.

**A Multiple Deficit Model of Learning Disabilities**

In this study a multiple deficits model of LD identification has been put forward. Similar to other multiple deficit models of identification, the model is based on cognitive deficits rather than reading achievement. Students identified as possibly LD with the multiple deficits screening tool put forward in the current study belong to five different groups based on their cognitive processing deficits. Those with deficits in phonological processing, naming speed and verbal memory form the triple deficit group. Those with deficits in two of the three processes all form three types of double deficit groups and, finally, those with phonological deficits only make up the last group.

Students who only had phonological deficits in the current study were flagged as possibly having LD, despite similar reading achievement scores to the RA group.
This decision is due to the high predictive value of phonological processing in previous research (David et al., 2007; Linklater et al., 2009; Muter & Snowling, 2009; Storch & Whitehurst, 2002; Svensson & Jacobson, 2006; Torgesen & Wagner, 1994). However, given the fact that there was less literature on the predictive value of naming speed or verbal memory deficits, along with the high reading scores of these groups, it was decided not to include students with a single deficit in naming speed or verbal memory in the LD cohort.

With three cognitive deficits tested in the current study, there were three types of double deficit groups, all of which were associated with reduced reading and spelling scores. Support was found for previous research that found a memory and phonological deficits phenotype within an LD cohort (Bonifacci & Snowling, 2008) as well as for the traditional double deficit theory (Wolf & Bowers, 1999). Furthermore, the finding that the naming speed and verbal memory deficits group, had lower reading scores than those with deficits in only one of these processes, is concordant with the finding that students with naming speed deficits tend to rely more heavily on working memory when reading (Baddeley, 1986). Students with deficits in only one of these cognitive processes may be able to compensate, but students with deficits in both may not.

The triple deficit group had the lowest scores on the PAT-R, SAST and RPM. This supports the theory put forward in Chapter 2 that these deficits can, in certain circumstances, result in lowered IQ tests scores as well as in tests of achievement. The low reading scores of the triple deficit group in the current study are compatible with those of previous studies that used phonological processing, short term memory and naming speed tasks, as well as others, to successfully predict LD (e.g., Chung et al., 2010; Hatcher et al., 2002). Furthermore, the suggestion that multiple deficits would result in compounded difficulty with reading (e.g., Baddeley, 1986; Wolf & Bowers, 1999) has been supported by the triple deficit group having lower achievement scores than their double deficit counterparts, who had lower scores than their single or no deficit counterparts.

The heterogeneous nature of LD has also been accounted for in a multiple deficits model. The range of deficits, combined with the accommodation offered with more lenient cut-offs for those students, also fits in well with recent genetic research
suggesting different chromosomes partially contributing to the existence of LD (Berninger et al., 2008; Pennington, 2009; Wood & Grigorenko, 2001) resulting in a wide range of types. The overlapping variance of some of these cognitive processing deficits (Pennington & Bishop, 2009) can make finding significant differences difficult, as this overlap increases the within group variance. This confounding factor needs to be taken into account when conducting significance testing with an LD cohort.

Another confounding aspect of LD research that a multiple deficits model can account for is that there are a wide range of ages at which LD can manifest itself. The reading difficulties of many students do not become apparent until the fourth grade, and many students who struggle with reading at younger ages go on to catch up to their peers without any intervention (Lipka et al., 2006). This can be a confounding factor for methods of identification that rely on reading achievement (Semrud-Clikeman, 2005; Shaywitz et al., 1991). Within the discrepancy paradigm, students with reading difficulties that are severe enough to affect their IQ scores cannot be designated as LD once their LD is severe enough to reduce a difference score to non-significance. Within a RTI paradigm, students cannot be identified until their reading difficulties have reduced their reading scores sufficiently for them to meet the performance-based cut-off for Tier 1 identification. Therefore, using reading-based measures to identify LD, as intuitive as that may be, is not going to pick up future problems accurately.

The screening tool, used in this study, or any tools developed within the model, could be amended to reflect the different needs of different grade levels. The most appropriate tests and cognitive processes could be selected specifically for each year level. A prime example of this is the fact that the letter RAN test was much more closely aligned with reading and spelling ability in the Grade 3 cohort in the current study. However, previous studies have demonstrated that, for younger students, the colour RAN would be a more appropriate measure (Misra et al., 2004). Consequently, a similar screening tool developed for a younger cohort could use symbols, rather than alphanumeric, stimuli. Furthermore, by specifically targeting all tests to one grade level, the difficulty of the test can be specifically targeted to a given population, avoiding the length and varying skill levels faced in tests that are suitable for a wider age range of students (e.g., Edwards & Hogben, 1999; Martin &
Pratt, 2001). In this regard, Grade 3 was selected for the focus of the current study, as it is thought to be the year when latent LD issues become apparent (Woolley, 2007). One of the exciting parts of the system of identification proposed here is that the battery of tests can be tailored to different grade levels.

The use of the multiple deficits model will also reduce the Type I error that stems from identifying students that have a significant discrepancy between aptitude and intelligence for reasons other than LD (Coutinho et al., 2002; Davis-Kean, 2005; Metsala et al., 1996). This same reduction would occur when comparing this model to RTI, in that those students failing to respond to intervention for reasons other than LD would not be identified (C. Reynolds & Shaywitz, 2009). Furthermore, the Type II error associated with under-identifying either low aptitude students via the discrepancy method (Stanovich, 1999) or average readers who are currently compensating for their difficulties through RTI (Gustafson & Samuelsson, 1999) will also be avoided.

Most importantly, both RTI and discrepancy have been accused of failing to inform remediation (C. Reynolds & Shaywitz, 2009; Silver et al., 2008). This is an important criticism as the aim of identifying the cause of reading problems is in order to more effectively address these problems. All student reports stemming from the screening tool piloted in the current study, even those without any literacy difficulties, would have a list of strengths and weaknesses in terms of the cognitive processes that contribute to reading. This report could be used to highlight the exact cognitive processes that LD or LA students are having problems with, and thus inform remediation or accommodation. Furthermore if relatively homogenous sub-groups of LD students can be identified, research on appropriate remediation programs would be made simpler and more direct.

An outline of the model and the groups of LD students identified within it has been put forward in this section. In the next section the theoretical implications of the multiple deficits model, and the use of a screening tool within it, will be discussed.

Implications for the Theory
There is a well documented difficulty in generalising results in this area of research due to the wide ranging methods of operationalisation of LD (Flanagan et al., 2006; Hoskyn & Swanson, 2000; Suhr, 2008). Furthermore, many of the concepts within
this field of research have been framed by the more popular operationalisations. The prime example is the concept of low achievement, which has come to mean a low IQ poor reader, rather than a poor reader who does not have LD. Therefore, it is important that any new method of identification is compared to older established methods so that differences in results can be explained and examined. How the multiple deficit model can account for some of the documented issues with older methods of identification will be discussed in this section.

Compared to phonological deficit models (e.g., Snowling, 1981), the wide range of possible subtypes within the multiple deficits model better explains the well documented heterogeneous nature of LD (Fiorello et al., 2006; Haworth et al., 2009; Pernet et al., 2009; Pierce et al., 2007). Supporting previous research, naming speed and verbal memory also played an important role in identifying LD, as it has been suggested that cognitive processes other than phonological processing are implicated (Johnson et al., 2009). How this provides support, or a lack thereof, for the phonological deficit models of LD depends solely on how verbal memory and naming speed are conceptualised. If it turns out that naming speed and verbal memory are both part of phonological processing, rather than separate processes (Carroll & Snowling, 2004; Catts & Hogan, 2003; Ramus, 2003), this will have little bearing on the relevance of this model, aside from some changes in nomenclature. The importance of the three processes to the theory does not change if these processes are then deemed sub-processes. If it turns out that all three, including the decoding aspect of the nonword reading test, belong under a phonological processing umbrella term, this does not change the value of the unique variance that all three processes provide.

When comparing multiple and double deficit models (Wolf & Bowers, 1999) of LD, the results of the current study supported the double deficit model inasmuch as both phonological and naming speed deficits were found in those designated as having LD. The results from the current study could be considered an extension of research that is able to predict the presence of LD through a double deficit model (Adlof et al., 2010; Puolakanaho et al., 2007). However, as these deficits were part of the method of identification, this is to be expected. The difference is in the number of deficits; in the current study it was demonstrated that verbal memory deficits also played a role. The results from the current study support the suggestion made in the
past that although naming speed and phonological processing deficits are a cause of LD, there were other deficits that also need to be taken into account (Ackerman et al., 2001). This includes those studies that also found verbal memory deficits as well as the traditional double deficit (Vukovic & Siegel, 2006).

The decision to not designate those with naming speed only deficits as LD does not necessarily conflict with the double deficit theory that posits that there is a naming speed only deficit group in LD (Badian, 1997; Wolf & Bowers, 1999). However, it could help explain inconsistent findings surrounding the naming speed deficit group within a double deficit paradigm. Research into the double deficit theory has proffered up conflicting results on the difficulties faced by a naming speed deficit only group (Vukovic & Siegel, 2006). By incorporating other cognitive processing deficits inherent to LD into this theory it could be that this group is representing those people who have deficits in other areas, such as verbal memory, in addition to naming speed. If some of the students in a double deficit-defined naming speed deficit group had other, untested deficits, such as in verbal memory, this could be sufficient cause for LD, even though others in the same group had only naming speed deficits. These intra-group differences would then confound significance testing.

The primary difference between the multiple deficit model put forward in the current study and those put forward previously (e.g., Badian, 1997; Fiorello et al., 2006; Ho et al., 2002; Pennington, 2006) is the use of verbal memory deficits rather than deficits in orthographic processing. The results of this study indicate that, while the inclusion of a verbal memory deficit is justified, the role of orthographic processing as a deficit that causes LD requires more research. Furthermore, there are other deficits put forward in these models that may require more investigation in conjunction with the model presented in the current study. One of the biggest advantages of any of these multiple deficit models is that such additions are congruent with the multiple deficits paradigm and, as such they can be incorporated into future research.

The main theoretical implication of a multiple deficit model replacing RTI as an operationalisation of LD is how remediation and or accommodations can be viewed. Instead of regarding remediation as something that will not work for LD students,
specific interventions compensating for different cognitive deficits can be formulated. Consequently, the aim of research may be generating effective intervention for LD students, rather than formulating hurdle that only their LA counterparts can overcome. That a multiple deficits model would be more accurate than RTI is difficult to ascertain due to the well documented difficulty of ascertaining a hit rate of RTI based LD identification (Kavale et al., 2008). However, previous research into the long term outcomes of cognitive deficits (Byrne et al., 1992; Svensson & Jacobson, 2006; Torgesen & Wagner, 1994) indicates that a multiple deficits model will be accurate.

Another advantage of the multiple deficits model, as compared to RTI and discrepancy, is that it is not based on reading skills and is, instead, based on the processes underlying these skills. This is also an advantage for the screening tool put forward in this study as compared to those screening tools based on reading skills (Good III et al., 2007; M. Reynolds et al., 2009; Speece et al., 2010; van Kraayenoord et al., 1999). Paradoxically, directly testing reading ability for reading-based LD reduces accuracy as there are a wide range of causal factors in reading ability. Furthermore, tests based on reading skills are obviously going to predict current, rather than future, deficits. This is, of course, something that a simple reading test alone can do. Cognitive processing skills are thought to be better predictors of future reading difficulties (Byrne et al., 1992). By testing cognitive processes it is possible to enable teachers to address LD before the Matthew effect takes hold.

One of the confounding factors involved in testing reading ability rather than cognitive processes is that this process will also identify LA readers. By using cognitive deficits as the identifying factors for LD students it should be possible to more accurately delineate LD and LA students. All methods of LD identification, except for those that are based on low achievement alone, have been developed with the aim of appropriately separating these two groups. However, as methods in the past have been based on likely outcomes of the disorder, this has resulted in those cognitive deficits that are less likely to result in these secondary symptoms, such as a significant discrepancy or failure to respond to intervention, being overlooked. The prime example of this is the role of verbal memory in discrepancy-defined LD and LA students.
The multiple deficit-defined groups identified in the current study with verbal memory deficits as one of the deficits, had lower average RPM scores than those without verbal memory deficits. Working memory plays a role in IQ, as supported by the working memory index of the Weschler Intelligence Scale for Children (WISC; Weschler, 2004). Consequently, within a discrepancy paradigm, these students would be less likely to be identified as LD as their verbal memory deficits are likely to reduce their IQ scores. However these students are just as likely as their discrepancy-defined LD counterparts to have phonological and naming speed deficits. Therefore both groups should be potentially LD and both groups require the same assistance and remediation.

It should be noted that this interpretation can only go so far, as there is also a processing speed index in the WISC. However, it may be that naming speed is sufficiently independent of processing speed for this to not confound results as compared to verbal and working memory. Given that those with naming speed deficits rely more heavily on memory capacity (Baddeley, 1986), this deficit may only be important when other deficits are present. This interpretation would account for both the confounding nature of verbal memory deficits and discrepancy-defined LD and inconsistent results on the cognitive profile of double deficit-defined students. Furthermore, a theory like the one proposed here can provide an explanation as to why naming speed and phonological processing differences were often not found between discrepancy defined LD and LA students. It could be because these deficits are common to both groups.

Because the same cognitive deficits thought to cause LD can also reduce IQ it is reasonable to be concerned that this method of identification will ultimately be a low achievement based model. Therefore it is worth noting that only two of the regression-adjusted discrepancy-defined LA students were identified as LD via the multiple deficits model. As noted before this is, in part, a reflection of the type of discrepancy method selected. However, it is also an indicator that the implementation of this method would not result in an influx of identification of LA students via the multiple deficits method, an ongoing concern (McPhail & Palincsar, 1998). Furthermore, this low number, along with the 22 students identified LA via this method, also supports, to a certain extent, the theory that these processing deficits are indicative of something other than mere poor reading skills.
It is important that this theoretical justification of operationalisation of LD does not occur in isolation of policy and practice, if this new model is not well suited to policy changes, it will not be implemented in schools. The implications of the current study on policy are discussed in the next section.

**Implications for Policy**

The need for new methods of improving literacy in Australian schools has been highlighted by the Australian Government’s recent smarter schools program in which 150 million dollars has been allocated to facilitate literacy or numeracy reform within schools (DEEWR, 2010). The focus in this program was on schools to better assist those with poor literacy in innovative ways. By screening all children for LD it should be possible in some cases to pick up problems before reading proficiency is affected, thus avoiding the Matthew Effect. The large amount of time and money needed to successfully identify LD students via RTI or discrepancy has been well documented and is not congruent with universal screening. The method of LD identification put forward in the current study provides information on the cognitive processes used in reading as well as identifying possible cases of LD. Consequently, the reports generated from this method of identification could provide useful information that could instigate the kind of reform that the Australian Government is currently advocating.

The Victorian Department of Education and Early Childhood Development (DEECD) recently highlighted educational outcomes in low socio-economic status communities as a focus in their Education Blueprint (2008). Furthermore, discrepancy and RTI seem susceptible to socio-economic bias, with the discrepancy method overlooking those with cultural disadvantage (Brooks-Gunn et al., 1996) and RTI over-identifying the same cohort (Al Otaiba & Fuchs, 2006). By basing operationalisation on definable and easily tested cognitive deficits, many of the political issues that influence current LD identification (McPhail & Palincsar, 1998) may be avoided.

The decision to develop a screening rather than diagnostic tool was made, in part, because there is less incentive to officially diagnose LD when there is no federal recognition of the disorder, as is the case in Australia (Jenkinson, 2007). Therefore it was more important to develop a tool that could give both parents and teachers
information on the reading progress of students, and to identify possible future problems, than it was to develop a tool that provided definitive definition. In countries with funding for LD students, this screening tool could provide a method of more accurately identifying students in need of diagnostic assessment.

Furthermore, it is hoped that a multiple deficits model of LD identification may provide support for the public perception that LD is a neurological disorder, not a socially invented construct. By basing identification on cognitive deficits, rather than secondary symptoms that can be confused with cultural disadvantage or other factors, the concern that LD diagnosis is used as an interim solution to wider school problems (Kavale et al., 2005) can be somewhat addressed. In turn, this approach could facilitate more widespread support for federal recognition of LD in Australia.

That the multiple deficits model and corresponding screening tool can be readily accepted in a school setting, regardless of changes in policy, is also important. In the next section the applied implications of this multiple deficits model in general, and the findings from the current study in particular, are discussed.

**Implications for Practice**

This section discusses the implications that the results of the current study might have in an applied setting. As discussed in Chapter 3, efforts were made to ensure that the screening battery piloted in the current study was well suited to the classroom, so that it was more likely to be implemented. The importance of ensuring that a test can be easily used in the setting that it is designed for is frequently ignored in LD research (Freebody, 2007).

The tests used for data collection for the current study were all extremely easy to administer and there is no reason why these tests could not be administered by a teacher or teacher’s aide. The word lists are the kinds of tests that both teachers and teacher’s aides administer in the classroom on a regular basis. The RAN tests were also easy to administer. A stopwatch was sufficient for timing the RAN tests and enough variance was provided when the tests were timed in whole seconds. Finally, digit span tests are traditionally administered by psychologists, however this is in order to ensure that results are appropriately interpreted, and not because the task itself requires specialist training. If interpretation is programmed into a database for
the current study, there is no reason why these three new tests could not be administered, scored and interpreted in a classroom setting.

Such a flexible method of identification could be easily moulded to fit in well with current programs, such as Reading Recovery, already in place in Australian schools. It is hoped that this will reduce the chance that the program will not be accepted in schools, as has happened in the past (e.g., D. Fuchs et al., 1996). Unlike other cognitive based systems of identification (Singleton et al., 1996), this battery would not require interpretation by a professional. Given that students can compensate for their reading difficulties long before they are outwardly manifested as reading difficulties (Lipka et al., 2006), global testing of all students would be the only way to ensure that all LD students are potentially identified. However it is not feasible to provide professional identification for all students, if current, more labour-intensive methods are used. Therefore, by implementing a global screening tool and by giving parents information on their child’s reading progress and likely factors that might hinder future progress, requests for costly individual assessments may reduce as a result (Long & McPolin, 2009).

When examining the use of clinical tests for LD, the usability of these tests is more likely to influence the chance of their selection by a practitioner than any psychometric concerns (Olm-Madden, 2008). Furthermore, even well established programs with demonstrated efficacy have been omitted or ignored by teachers if they cannot fit in with their day to day routine (D. Fuchs et al., 1996). Without ensuring that tests can be readily used and will not be an unrealistic time burden on classrooms teachers, they may never be fully embraced by the very people for whom they are designed. By using the measures selected for the current study, it was possible to automatically generate reports such as those shown in Appendix G without writing each of them manually. Individual interpretation and generation of reports would be an unrealistic burden on already high teacher workloads.

This aim of automatically generating reports is not an abstract one, nor is it untenable. First of all, it is important to ensure that the test battery fits into a system that can be easily scored and information easily entered into a database. The method of programming an Excel worksheet so that reports could be automatically generated is outlined in Appendix I. Although tools such as this may not seem directly
relevant to the field of LD at first glance, it could be argued that the dismissal of such pragmatic considerations is one of the factors holding LD research back in applied settings. These reports will be generated with the deficits clearly outlined so that the school can decide how to proceed on the basis of the information required. One of the advantages of this is that the reports can be used regardless of whichever paradigm is being implemented at schools at any given time.

The removal of the intelligence test from LD diagnosis will not impact on the ability of the test in this study to inform remediation, as it has been noted on multiple occasions that IQ tests do not contribute to remediation (e.g., Fletcher et al., 2004; Willis & Dumont, 2006). Current remediation development is confounded by the diverse group of students that researchers are hoping to accommodate. By diagnosing on the basis of cognitive processes it should be possible to also group students by their processing deficits, resulting in relatively homogeneous groups under the heterogeneous LD umbrella. This can then lead to tailored remediation or accommodations that are appropriate for the cognitive deficits identified in individual LD students.

This is not to say that non-LD poor readers should not receive extra assistance with their reading. Conversely, these students could be highlighted as those most likely to respond to an appropriately designed intervention. These students may be more likely to benefit from an intervention with a focus on reading in general, as their reading problems are not neurologically determined. Without the types of deficits that are inherent to LD, a remediation program with a focus on intensive phonological training may not be suitable. This is important at a Grade 3 level where, for example, a remediation program based on explicit phonological training would not challenge non-LD poor readers.

This multiple deficit model has been put forward as capable of overcoming many issues inherent to other methods of LD identification. This assertion has been extrapolated from the results of the current study. However, this study is not without limitations, and these limitations are acknowledged and discussed in the next section.

**Limitations**

Although significant contributions to the field were made in the current study, it is important to acknowledge the limitations of the study. First and foremost, it is
important to note that as the tests developed formed a screening tool, not a diagnostic one, there were high levels of false positives or Type II error. In the current study, 18% of students were flagged as possibly having LD, obviously much higher than accepted prevalence rates that are closer to 5%. Although this is an accepted aspect of screening tools, it is vital that it is recognised (Gredler, 1997).

A major consideration when interpreting the results of the current study is that even though the results are compatible with a multiple deficits model, there are other possible interpretations. For instance, the fact that 79% of students could be correctly grouped into the regression-adjusted discrepancy groups could be explained by the fact that this method is actually more effective at picking up “true” LD students and that the other 21% represents error in the discrepancy method. However there are of course many other ways that this finding could be interpreted and longitudinal studies are needed to confirm this explanation of the results.

Furthermore, given the highly specific method of splitting the LD cohort according to cognitive deficit groups, sample size is another limitation of the current study that must be recognised. For instance, in the double and triple deficit groups identified in the current study, only three to six students met the criterion for each group. Consequently, the results stemming from these analyses need to be interpreted with caution. This splitting of the LD cohort into more homogenous groups will dictate that higher sample sizes in future research will be needed to ensure that each group is adequately represented for statistical analyses.

The thorny issue of different types of memory and their operationalisation was outside the scope of the current study. The results from both the current study and other studies on memory indicate that verbal memory deficits do play a role in LD (Gang & Siegel, 2002; Swanson & Beebe-Frankenberger, 2004). However, investigation of memory issues in LD identification warrants further investigation. The best kind of memory test for predicting reading achievement, and whether that test will be assessing working memory or short term memory, is currently unclear. The current study has utilised digit span as a measure of verbal memory due to its common use in the past, however this does not mean that it is the most appropriate measure.
There were a few procedural issues surrounding the timing and location of testing that need to be taken into account when interpreting the results from the current study. The testing for the current study was done over a four month period covering most of the second semester of 2009. Consequently, those students who were tested last had received close to an extra half year of instruction as compared to those students tested first. Additionally, the norms would only be appropriate in an Australian setting given that norms on literacy tests do not translate well internationally (Hempenstall, 2009).

Another limitation of the current study was the ceiling effect in the irregular word list. The reduced correlation resulting from the skew (Kendall & Stuart, 1958) has made the irregular word list less suited to a cluster analysis than the other variables. However the separation between the cluster analysis-defined groups of irregular word reading was good in both analyses. Because of this skew, Levene’s test of homogeneity of variance was conducted on all variables before significance testing. Ultimately it is hoped that enough steps were taken to ensure that the ceiling effect in the irregular word list did not affect results but it is important that the possibility that it has is acknowledged.

Another limitation in regards to the sampling framework in the current study that needs to be taken into account is that the reading achievement test was not the same at the Grade 1 level as it was in Grades 2 and 3. The current study had to utilise the students’ teacher designated reading levels as the teachers decided that the PAT-R was not appropriate for their Grade 1 students. By utilising a test that is appropriate at all levels, more precise reading-age measures could be calculated, potentially providing more accuracy in any screening tool.

**Directions for Future Research**

Similar to most research, the results from the current study have led to more questions that deserve further investigation. The most informative research that needs to be undertaken in this area is longitudinal studies that follow the reading progress of students that have deficits in the cognitive processes outlined in the current study. Although studies in the past have looked at the ability of some of these processes to predict reading achievement, they have studied the cognitive processes of students on a continuum instead of as dichotomous deficit/no deficit
variables. Although lower cut-offs were put in place for students with more than one deficit, this was the only concession made to the fact that there are interactions between the two deficits. It has been suggested that while deficits can be indicative of LD, high levels of these processes may not result in different outcomes to average or above average readers (Anderson, 2008). Primarily, any regression based equation will attempt to compensate for this, reducing accuracy at the lower end of the reading spectrum. Therefore a model based only on deficits, or a lack thereof, may actually be more informative than one that treats these cognitive processes as continuous variables.

Furthermore, by following student achievement over the long term, after testing them on these cognitive variables, further information can be gathered as to what kinds of issues students with specific combinations of cognitive processing deficits may face in the future. This approach could highlight areas in need of remediation. The other reason that longitudinal studies are needed in this area is to provide support for the multiple deficits model.

As noted in this chapter, students with naming speed or verbal memory deficits only were not identified as possibly LD in the current study. This decision was based on a lack of research on the outcomes of students with deficits such as these without the presence of other deficits. It was not because there was research indicating that these deficits alone do not play a role in LD. Consequently, this is an area of future research that could prove interesting. Furthermore, as previously mentioned, there is a need for research into what type of memory deficits are best used for identifying LD. For instance, it has been recommended that memory tasks involving phonologically novel input may be more appropriate for studies on the role of memory in reading (Baddeley et al., 1998). If it is decided that working memory is an LD deficit, then the use of the digit span backward could be revisited as there is still controversy over whether it is a measure of short term or working memory (Richardson, 2007). Alternatively, a different test that is more widely accepted as a measure of working memory, such as reading span (Daneman & Carpenter, 1980), could be used instead in order to more directly assess the role of working memory in LD.
If students are to be identified as possibly LD via a multiple deficits model, then the next step is to develop remediation programs based on the different types of deficits that any given student is hoping to overcome. As noted in Chapter 3, remediation programs more relevant to the causal cognitive processing deficits of LD students have been requested for many years (e.g., Hay et al., 2007; Mather & Gregg, 2006). In this regard, some programs have been developed with specific cognitive processing deficits in mind (Foorman et al., 1998; Hayward et al., 2007; Hempenstall, 2008; Swanson et al., 2010). Incorporating the multiple deficits model with this line of research should assist in equipping teachers with the tools they need to help LD students in their classroom. Furthermore, by identifying these deficits explicitly and developing remediation programs that address them, more information can be gathered on the permanence, or lack thereof, of these neurological deficits.

A currently ignored psychometric problem that warrants further study is the use of screening tools in statistical research. The primary aim of screening tools is to identify a small cohort of individuals who are unusually high or low on a targeted construct. Consequently, when given to the general population, the majority of people taking the test score at one extreme end of the scale. Although the applied advantages of screening tools are well documented (e.g., Glover & Albers, 2007), the effect of skew on any analyses that these scales are used in is rarely addressed, despite their prolific use. Consequently, this is an area in urgent need of further research, not just within educational psychology but in any area that uses screening tests on a general population.

Given the unclear results in the current study on the role of orthographic processing as a reflective or formative variable of LD, this is an area in which future research is needed. The irregular word reading list is thought to reflect, to some extent at least, exposure to print. Because of this, the irregular word list is of use in a screening tool targeted to poor readers, regardless of the link with LD. This is because as a measure of exposure to print, it can provide information on a student’s reading progress, regardless of the presence or lack thereof of other cognitive processing deficits. However, as recent research is demonstrating that there may be an issue with the integration of orthographic and phonological information (Cao et al., 2008; Desroches et al., 2010), research into how deficits in the integration of different cognitive processes can predict the presence of LD may be worthwhile.
This brings us to the most important caveat and suggestion for future research from this study. The identification model proposed here is by no means definitive. The role of orthographic processing is unclear, and other processes are still being studied as possibly contributing to LD. For instance, the role of more direct measures of working memory and morphological processing in predicting LD needs to be examined. It is possible that there are other deficits that contribute to LD and the inclusion of them into a model such as this one could result in more accuracy. Although the wide range of possible cognitive processing deficits that could be responsible for LD make the identification of these same processes difficult to find (Kohli et al., 2005), it is a worthwhile concentration of resources, given all the inherent advantages to this model.

Furthermore, it is possible that some of these other deficits, not examined in the current study, could still be confounding results. This means that some of the students not identified on the basis of only having a naming speed or verbal memory deficit may actually have another deficit as well that could in the future result in them being identified as possible LD. If more cognitive processing variables were added in the future then the cut-off scores for identification would need to be reviewed in order to ensure that the Type II error was not too high, even for a screening tool.

As noted in Chapter 2 there are negative long term implications of early literacy difficulties (McGee et al., 2002; Young et al., 2002) and once literacy difficulties set in, they often compound themselves (Morgan et al., 2008; Stanovich, 1986), as indicated by the Matthew Effect. Therefore, these are worthy avenues of future research as picking up the cognitive processing deficits inherent to LD early can result in more effective remediation and increased literacy rates.

**Conclusion**

The primary aim of the current study was to develop a screening tool suited to classroom administration that could identify possible cases of LD. In order to do this more effectively, a multiple deficits model of LD has been proposed. By testing for phonological processing, naming speed and verbal memory deficits it was possible to identify a similar cohort to that identified via the discrepancy method. It could also be argued that as these students were identified on the basis of the cognitive
processing deficits thought to cause LD, that the multiple deficits model is a better representation of LD than that provided by the discrepancy model.

The first contribution of the current study is the theoretical justification of the use of the discrepancy method as a basis for ascertaining the convergent validity of more direct methods of identification. The vast range of identification methods currently in use make it vital that the identification method chosen as the basis of convergent validity is chosen carefully and made explicit. This is because the effect of this baseline method on the new method of identification needs to be acknowledged in the interpretation of the results. Furthermore, the concept of unexpected underachievement that underlies the discrepancy method of identification is a good representation of LD. Therefore, by ascertaining which cognitive processing deficits are aligned with a discrepancy, the deficits themselves can be used to identify LD.

The second contribution is the development of the irregular and nonword lists that were psychometrically refined as per CTT and IRT. Both of these lists are specifically tailored towards an Australian Grade 3 cohort, with the express purpose of identifying poorly performing students with significant negative skew. The norms for these tests and those for the digit span and RAN tests, found in Appendix I, are freely available for use in Australian classrooms.

The third contribution of the current study is the finding that there is probably a role for verbal memory deficits as a cause of LD, a finding that has much less support in previous research than the role of phonological processing or naming speed. This was supported by the importance of the verbal memory measures in the DFA as well as the decreasing reading achievement of students with multiple deficits, including verbal memory. For instance, the naming speed and verbal memory double deficit group had lower scores on both reading and non-verbal IQ than any of the single deficit groups.

It has been proposed in the current study that the long sought difference between LD and LA students cannot be found when using the discrepancy method because verbal memory deficits have affected the allocation of students into these groups. In previous research, the convergence of the double deficit theory and the discrepancy theory, inasmuch as these deficits do not seem to adversely effect IQ, may have resulted in the importance of verbal memory in LD identification being ignored.
possible interpretation of the results in the current study could be that the effect that verbal memory deficits have on measures of fluid intelligence in some LD students has resulted in a diagnosis of LA. This is because deficits in verbal memory were affecting intelligence as well as reading. When IQ decreases, the corresponding deficit between aptitude and achievement also decreases.

The fourth contribution of the current study is the formulation of a new multiple deficits model of LD, by which phonological processing, naming speed and verbal memory deficits can be used to identify LD. An important part of this is the acknowledgement that this list of three processes may not be definitive, as there may be other deficits causing LD that could aid in identification. Furthermore, it is important to ensure that work in previous studies can be interpreted through this new model; as such the current study has provided a reinterpretation that consolidates many of the issues inherent to the discrepancy method of identification. This has been achieved by explaining the results through a multiple deficits paradigm. If it is accepted that the discrepancy based LD cohort is only an approximation, and that some of the LA students have LD and the corresponding deficits that go with the disorder, then many of the problems with separating the two in the first place can be addressed.

Finally, and most importantly, this research has culminated in a prototype screening tool for LD that can be utilised in the classroom and in future research. This was made possible by ensuring firstly, that the new model was developed based on theoretical concerns surrounding current LD identification and operationalisation. Secondly, the manner in which students are identified is flexible and could accommodate further cognitive processes such as orthographic processing if they are found to contribute to LD. This is important, as the diagnosis of LD on the basis of cognitive processes is a relatively new idea, and some processes such as orthographic and morphological processing merit further research.

**Final Comment**

Research into the early identification of LD is clouded by inconsistent operationalisation of the disorder. By utilising a method of identification that is focused on the cause of the disorder, rather than a possible secondary outcome, it should be possible to more accurately identify those students with an actual
neurological disorder rather than literacy difficulties with a different cause. Furthermore it is vital that new methods of identification are readily applied to the classroom, with technically accurate methods of identification. Ultimately, it is important that methods of identification such as the multiple deficits model are implemented so that research can reveal the best way to remediate the reading difficulties facing LD students and the adverse long term outcomes that currently accompany this disorder can be avoided.

References


Sets (Argentina, France, Hong Kong, Spain, USA and UK). In N. Bezruczko (Ed.), Rasch Measurement in Health Sciences (pp. 411-431). Maple Grove: JAM Press.


Fletcher, J. M., & Vaughn, S. (2009b). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives, 3*(1), 30-


Individuals with Disabilities Education Act (IDEA), 20 USC 1401 §30 (2004).


doi:10.1207/s15327035ex1304_3


Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the
garden-variety poor reader: The phonological-core variable-difference model.
doi:10.1177/002221948802101003

intelligence led us astray? *Reading Research Quarterly, 26*, 7-29. Retrieved from
http://www.jstor.org/pss/747729

Psychology and Psychiatry Review, 3*(1), 17-21. doi:10.1111/1475-
3588.00203

Stanovich, K. E. (1999). The sociopsychometrics of learning disabilities. *Journal of

Stanovich, K. E. (2005). The future of a mistake: Will discrepancy measurement
continue to make the learning disabilities field a pseudoscience? *Learning

with reading disabilities: A regression-based test of the phonological-core
variable-difference model. *Journal of Educational Psychology, 86*(1), 24-53.
Retrieved from http://goo.gl/Zwe8T

phonological and surface subtypes of reading disability. *Journal of
Educational Psychology, 89*(1), 114-127. Retrieved from
http://goo.gl/nKZYb

Stanovich, K. E., & Stanovich, P. J. (1997). Further thoughts on
aptitude/achievement discrepancy. *Educational Psychology in Practice,
13*(1), 3-8. doi:10.1080/0266736970130101

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic
http://www.jstor.org/pss/747605


Swanson, H. L. & O’Connor, R. (2009). The role of working memory and fluency practice on the reading comprehension of students who are dysfluent readers.


Hambleton (Eds.), *Handbook of modern item response theory*. New York: Springer-Verlag.


Appendix A

South Australian Spelling Test

1. ON
   Please put your shoe ON. Write ON.

2. HOT
   The water in the bath is HOT. Write HOT.

3. CUP
   I drink from a CUP. Write CUP.

4. VAN
   The lady can drive the VAN. Write VAN.

5. JAM
   I like JAM on my bread. Write JAM.

6. MUD
   I got MUD on my shoes when it rained. Write MUD.

7. SIT
   Please SIT on this chair. Write SIT.

8. BEG
   I taught my dog to BEG for a biscuit. Write BEG.

9. ME
   This present is not for ME. Write ME.

10. GO
    I will GO to the shops after school. Write GO.

11. DO
    What will you DO next? Write DO.

12. OF
    I am not sure OF your name. Write OF.

13. THE
    Is this THE toy you want? Write THE.

14. SO
    You did that job SO quickly. Write SO.

15. PLAN
    I used a PLAN to make this model. Write PLAN.

16. SHIP
    A SHIP is on the sea. Write SHIP.

17. CHOP
    The butcher will CHOP the meat. Write CHOP.

18. FROM
    Our new teacher comes FROM Sydney. Write FROM.

19. THIN
    The THIN cat squeezed under the furrow. Write THIN.

20. LOST
    I LOST my key. Write LOST.

21. DART
    I throw a DART at the dartboard. Write DART.

22. SEEM
    The shop did not SEEM to be open. Write SEEM.

23. FOOD
    We must take FOOD to the picnic. Write FOOD.

24. FOR
    Is this letter FOR me? Write FOR.

25. ARE
    Animals ARE in the field. Write ARE.

26. WHO
    WHO was that knocking at the door? Write WHO.

27. HERE
    Put the box over HERE. Write HERE.

28. FIRE
    We need dry sticks to start the FIRE. Write FIRE.

29. DATE
    What IS the DATE today? Write DATE.

30. LOUD
    Your velcro is too LOUD. Write LOUD.

31. EYE
    Please shut one EYE and look at this. Write EYE.

32. FIGHT
    I saw two dogs FIGHT in the park. Write FIGHT.

33. FRIEND
    She is my best FRIEND. Write FRIEND.

34. DONE
    What have you DONE with your book? Write DONE.

35. ANY
    Are there ANY cakes left? Write ANY.
| 36 | GREAT | I was chased by a GREAT big dog. Write GREAT. |
| 37 | SURF | I am not SURF how to spell this. Write SURF. |
| 38 | WOMEN | Two WOMEN went for a swim. Write WOMEN. |
| 39 | ANSWER | Please ANSWER my question. Write ANSWER. |
| 40 | BEAUTIFUL | The flowers in the garden look BEAUTIFUL. |
| 41 | ORCHESTRA | I play the piano in the ORCHESTRA. |
| 42 | EQUALLY | They shared the money EQUALLY. |
| 43 | APPRECIATE | Thank you. I APPRECIATE your help. |
| 44 | FAMILIAR | His face seemed FAMILIAR. Had we met before? |
| 45 | ENTHUSIASTIC | The student was an ENTHUSIASTIC player. |
| 46 | SIGNATURE | She wrote her SIGNATURE on the paper. |
| 47 | BREATHE | Fresh air is good to BREATHE. |
| 48 | PERMANENT | Will that sign be taken away or is it PERMANENT? |
| 49 | SUFFICIENT | We have SUFFICIENT food to last for the weekend. |
| 50 | SURPLUS | We will sell the SURPLUS apples. We have too many. |
| 51 | CUSTOMARY | It is CUSTOMARY to shake hands. |
| 52 | ESPECIALLY | This gift is ESPECIALLY for you. |
| 53 | MATERIALLY | This story is not MATERIALLY different from the one in your book. |
| 54 | CEMETERY | The funeral took place at the CEMETERY. |
| 55 | LEISURE | She spent her LEISURE time in the garden. |
| 56 | FRATERNALLY | FRATERNALLY means the same as brotherly. |
| 57 | SUCCESSFUL | The fund-raising was very SUCCESSFUL. |
| 58 | INDEFINITE | I agreed on a INDIFICITE time to meet her. |
| 59 | EXHIBITION | There is an art EXHIBITION at the gallery. |
| 60 | APPARATUS | We use the APPARATUS in the science lab. |
| 61 | MORTGAGE | I bought the house by taking a MORTGAGE. |
| 62 | EQUIPPED | The campers were EQUIPPED with new tents. |
| 63 | SUBTERRANEAN | SUBTERRANEAN means under the ground. |
| 64 | POLITICIAN | Did you vote for that POLITICIAN? |
| 65 | MISCELLANEOUS | Mixing different items together makes a MISCELLANEOUS set. |
| 66 | EXAGGERATE | The fish wasn't that big! Don't EXAGGERATE. |
| 67 | GUARANTEE | My washing machine has a two-year GUARANTEE. |
| 68 | CREDENTIALS | I find it CREDENTIALS to give a speech. |
| 69 | CONSCIENTIOUS | Students who work hard are said to be CONSCIENTIOUS. |
| 70 | SEISMOGRAPH | A SEISMOGRAPH is an instrument to measure the strength of earthquakes. |
## Appendix B

### Psychometric Properties of the Irregular Word Lists

**Table B1**

*Number of Syllables, Letters and Phonemes and Longest Consonant Cluster in the Full Irregular Word List*

<table>
<thead>
<tr>
<th></th>
<th>Syllables</th>
<th>Letters</th>
<th>Phonemes</th>
<th>Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>to</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>we</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>of</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>use</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>sea</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>put</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>sew</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>who</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>boat</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>want</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>four</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>said</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>have</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>done</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>high</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>blue</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>eight</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>juice</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>towel</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>cough</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>laugh</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>emu</td>
<td>2</td>
<td>9</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>zero</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>people</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>stomach</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>tortoise</td>
<td>2</td>
<td>8</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>frightened</td>
<td>2</td>
<td>10</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>excited</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>exhausted</td>
<td>3</td>
<td>9</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>jewellery</td>
<td>3</td>
<td>10</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>characters</td>
<td>3</td>
<td>10</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>somersault</td>
<td>3</td>
<td>10</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>rhinoceros</td>
<td>4</td>
<td>10</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>especially</td>
<td>4</td>
<td>10</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>certificates</td>
<td>4</td>
<td>12</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>cooperation</td>
<td>5</td>
<td>11</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>
# Appendix C

Psychometric Properties of the Nonword List

Table C1

*Tally of Letters used in the Full Nonword List*

<table>
<thead>
<tr>
<th>Letters</th>
<th>Tally</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>13</td>
</tr>
<tr>
<td>b</td>
<td>4</td>
</tr>
<tr>
<td>c</td>
<td>3</td>
</tr>
<tr>
<td>d</td>
<td>3</td>
</tr>
<tr>
<td>e</td>
<td>16</td>
</tr>
<tr>
<td>f</td>
<td>3</td>
</tr>
<tr>
<td>g</td>
<td>2</td>
</tr>
<tr>
<td>h</td>
<td>6</td>
</tr>
<tr>
<td>i</td>
<td>9</td>
</tr>
<tr>
<td>j</td>
<td>4</td>
</tr>
<tr>
<td>k</td>
<td>5</td>
</tr>
<tr>
<td>l</td>
<td>7</td>
</tr>
<tr>
<td>m</td>
<td>4</td>
</tr>
<tr>
<td>n</td>
<td>6</td>
</tr>
<tr>
<td>o</td>
<td>16</td>
</tr>
<tr>
<td>p</td>
<td>9</td>
</tr>
<tr>
<td>q</td>
<td>1</td>
</tr>
<tr>
<td>r</td>
<td>10</td>
</tr>
<tr>
<td>s</td>
<td>4</td>
</tr>
<tr>
<td>t</td>
<td>7</td>
</tr>
<tr>
<td>u</td>
<td>8</td>
</tr>
<tr>
<td>v</td>
<td>1</td>
</tr>
<tr>
<td>w</td>
<td>2</td>
</tr>
<tr>
<td>x</td>
<td>3</td>
</tr>
<tr>
<td>y</td>
<td>6</td>
</tr>
<tr>
<td>z</td>
<td>2</td>
</tr>
</tbody>
</table>

N = 36
Table C2

*Tally of Phonemes used in the Full Word List*

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Consonants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonemes</td>
<td>Tally</td>
</tr>
<tr>
<td>ə</td>
<td>6</td>
</tr>
<tr>
<td>ɪ</td>
<td>4</td>
</tr>
<tr>
<td>e</td>
<td>5</td>
</tr>
<tr>
<td>aɪ</td>
<td>2</td>
</tr>
<tr>
<td>Λ</td>
<td>2</td>
</tr>
<tr>
<td>eɪ</td>
<td>3</td>
</tr>
<tr>
<td>i</td>
<td>2</td>
</tr>
<tr>
<td>œʊ</td>
<td>2</td>
</tr>
<tr>
<td>æ</td>
<td>2</td>
</tr>
<tr>
<td>ø</td>
<td>4</td>
</tr>
<tr>
<td>ɔ</td>
<td>3</td>
</tr>
<tr>
<td>u</td>
<td>4</td>
</tr>
<tr>
<td>u</td>
<td>1</td>
</tr>
<tr>
<td>a</td>
<td>1</td>
</tr>
<tr>
<td>aʊ</td>
<td>1</td>
</tr>
<tr>
<td>ð</td>
<td>2</td>
</tr>
<tr>
<td>eə</td>
<td>1</td>
</tr>
<tr>
<td>ə</td>
<td>1</td>
</tr>
<tr>
<td>œ</td>
<td>3</td>
</tr>
<tr>
<td>uə</td>
<td>0</td>
</tr>
<tr>
<td>ʊə</td>
<td>2</td>
</tr>
<tr>
<td>η</td>
<td>2</td>
</tr>
</tbody>
</table>

N = 36
Table C3

*Number of Syllables, Letters and Phonemes and Longest Consonant Cluster in the Full Nonword List*

<table>
<thead>
<tr>
<th>Nonword</th>
<th>Syllables</th>
<th>Letters</th>
<th>Phonemes</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>ab</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>ek</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>ra</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>zay</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>aip</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>foy</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>oot</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>vun</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>huk</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>yex</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>kour</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>nied</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>jite</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>reng</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>nawk</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>crex</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>zatch</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>joid</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>thunf</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>rilch</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>quelp</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>oji</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>eero</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>wopa</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>kory</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>olloy</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>habay</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>joota</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Nonword</td>
<td>Cluster</td>
<td>1st Cons</td>
<td>2nd Cons</td>
<td>3rd Cons</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>bupode</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>shumba</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>poslip</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>geasure</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>frimpler</td>
<td>2</td>
<td>8</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>strimpex</td>
<td>2</td>
<td>8</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>lantople</td>
<td>3</td>
<td>8</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: Cluster = length of the longest string of consecutive consonants within the nonword.
Appendix D

Ethical Clearance from the Swinburne University, Department of Education

From: Keith Wilkins
To: Callinan, Sarah; Cunningham, Everarda
Date: 11/03/2009 5:39 pm
Subject: SUHREC Project 2008/111 Ethics Clearance

CC: Thelier, Stephen
To: Assoc Prof Everarda Cunningham/Ms Sarah Callinan, FHIL

Dear Arda and Sarah,

SUHREC Project 2008/111 Development of an Australian-normed measure of reading
Assoc Prof Everarda Cunningham, FHIL; Ms Sarah Callinan, Dr Stephen Thelier

Approved Duration: 11/03/2009 To 30/06/2011

I refer to the ethical review of the above project protocol undertaken by Swinburne's Human Research Ethics Committee (SUHREC). Your responses to the review were as emailed on 6 December 2008 and 25 February 2009, the latter attaching further revised consent information statements, the former attaching the consent form considered to be used. The responses were put to the Chair of SUHREC for consideration.

I am pleased to advise that, as submitted to date, the project has approval to proceed in line with standard on-going ethics clearance conditions here outlined. As regards involvement of schools operating under a formal overarching authority, eg, Government or Catholic schools, evidence of authority to involve the schools concerned should be forwarded to my office for the record as soon as practicable.

- All human research activity undertaken under Swinburne auspices must conform to Swinburne and external regulatory standards, including the National Statement on Ethical Conduct in Human Research and with respect to secure data use, retention and disposal.

- The named Swinburne Chief Investigator/Supervisor remains responsible for any personnel appointed to or associated with the project being made aware of ethics clearance conditions, including research and consent procedures or instruments approved. Any change in chief investigator/supervisor requires timely notification and SUHREC endorsement.

- The above project has been approved as submitted for ethical review by or on behalf of SUHREC. Amendments to approved procedures or instruments ordinarily require prior ethical appraisal/ clearance. SUHREC must be notified immediately or as soon as possible thereafter of (a) any serious or unexpected adverse effects on participants and any remedial measures; (b) proposed changes in protocols; and (c) unforeseen events which might affect continued ethical acceptability of the project.

- At a minimum, an annual report on the progress of the project is required as well as at the conclusion (or abandonment) of the project.

- A duly authorised external or internal audit of the project may be undertaken at any time.

Please contact me if you have any queries about on-going ethics clearance. The SUHREC project number should be quoted in communication.

Best wishes for the project.

Yours sincerely,

Keith Wilkins
Secretary, SUHREC

***************************************************************
Keith Wilkins
Research Ethics Officer
Swinburne Research (F68)
Swinburne University of Technology
P O Box 218
HAWTHORN VIC 3122
Tel +61 3 9214 5218
Fax +61 3 9214 5267
Ethical Clearance from the Victorian Department of Education and Early Childhood Development

Department of Education and Early Childhood Development

Office for Policy, Research and Innovation

2 Treasury Place
East Melbourne, Victoria 3002
Telephone: +61 3 9037 2000
DX 210863
GPO Box 4307
Melbourne, Victoria 3001

RIS09039

Ms Goreh Callinan
Swinburne University of Technology
Faculty of Higher Education
Locked Bag 218
LILYDALE 3140

Dear Ms Callinan,

Thank you for your application of 2 March 2009 in which you request permission to conduct a research study in government schools titled: Development of an Australian-Normed Measure of Reading.

I am pleased to advise that on the basis of the information you have provided your research proposal is approved in principle subject to the conditions detailed below.

1. Should your institution's ethics committee require changes or you decide to make changes, these changes must be submitted to the Department of Education and Early Childhood Development for its consideration before you proceed.

2. You obtain approval for the research to be conducted in each school directly from the principal. Details of your research, copies of this letter of approval and the letter of approval from the relevant ethics committee are to be provided to the principal. The final decision as to whether or not your research can proceed in a school rests with the principal.

3. No student is to participate in this research study unless they are willing to do so and parental permission is received. Sufficient information must be provided to enable parents to make an informed decision and their consent must be obtained in writing.

4. As a matter of courtesy, you should advise the relevant Regional Director of the schools you intend to approach. An outline of your research and a copy of this letter should be provided to the Regional Director.

5. Any extensions or variations to the research proposal, additional research involving use of the data collected, or publication of the data beyond that normally associated with academic studies will require a further research approval submission.

6. At the conclusion of your study, a copy or summary of the research findings should be forwarded to Education Policy and Research Division, Department of Education and Early Childhood Development, Level 2, 33 St Andrews Place, GPO Box 4367, Melbourne, 3001.

Every child, every opportunity
I wish you well with your research study. Should you have further enquiries on this matter, please contact Chris Warne, Senior Policy and Research Officer, Education Policy and Research, by telephone on (03) 9637 2272 or by email at <warne.chris@edumail.vic.gov.au>.

Yours sincerely

Elizabeth Hartnell-Young
Group Manager
Education Policy and Research

24/3/2009

enc
17 November 2009

Ms Sarah Callinan
Investigator
Swinburne University
Mail L100- research and Development
Locked bag 218
LILYDALE VIC 3140

Dear Ms Callinan

Thank you for your letter dated 10 November, 2009 in which you have requested permission to conduct research titled “Learning problems in young students and how they can be most efficiently identified” for students in Grade 3 in Primary schools in the Diocese of Sale.

I am happy for you to approach the schools in this diocese. It is important that you understand that the final permission for you to undertake this work rests with the Principal.

This approval in principle is subject to the attached Research - Standard Conditions. Should you require further information please contact Mr Peter Ryan at this Office, email pryan@ceo.sale.catholic.edu.au or phone 5622 6603.

With best wishes

Yours sincerely

Peter Ryan
DIRECTOR OF CATHOLIC EDUCATION
DIOCESE OF SALE
RESEARCH – STANDARD CONDITIONS

This approval in principle is subject to certain standard conditions:

1. The decision as to whether or not a project can proceed in a school rests with the school principal. You will need to obtain approval directly from the principal of each school that you wish to involve.

2. You should provide the principal with an outline of your proposal and indicate what will be asked of the school. A copy of this letter of approval should be included.

3. If your study involves one-to-one contact with a child, you are required to provide the principal with a certified copy of your registration with the Victorian Institute of Teaching or a current “Working with Children” check.

4. You should provide to the director of this office, a list of schools which have agreed to participate in the research project.

5. Any substantive modifications to the research proposal, or additional research involving use of the data collected will require a further research approval submission to this office.

6. Data relating to individuals or schools is to remain confidential.

7. The collection and use of information from schools must conform with the Privacy Amendment (Private Sector) at 2000 as indicated in the School’s Privacy Policy.

8. Since participating schools have an interest in the findings, you should discuss with each principal ways in which the results of the study could be made available for the benefit of the school community.

9. At the conclusion of the study, a copy of the summary of the findings should be forwarded to the Director of this Office.
Appendix E

Information Sheet and Consent Form Signed by all Participants

PARENT/GUARDIAN INFORMATION SHEET

Development of an Australian-normed measure of reading

Researchers: Dr Everarda Cunningham, Sarah Callinan and Dr Stephen Theiler
Swinburne University of Technology

Your son/daughter is invited to participate in a research project that is explained below.

If you would like this information and consent form in another language, please ask your son or daughter's teacher.

Thank you for taking the time to read this Information Sheet.

What is an information statement?
An information statement explains what the research project is about and what you are being asked to do if you want your son or daughter to take part. Please read this information carefully. You can also ask your school or your child's teacher more about the project. Once you have understood what the project is about and if you would like your child to take part please sign the consent form at the end of this statement and return it to your child's school.

What is the research project about?
The aim of this study is to come up with an affordable test that teachers can give to students in order to check for learning difficulties. During the study, your child will take part in reading, spelling and general ability tests designed to identify students who have early stage problems with reading and spelling or signs of a learning problem.

The child of each parent who consents to their child taking part will have an individual report written up summarising their results that will be forwarded on to their teacher. This study is very important and will look at how children with reading and spelling problems can not only be easily identified but how we can then give both teachers and parents better tools to help children overcome problems with reading and spelling.

Who is in conducting of the research?
Sarah Callinan is a PhD student at Swinburne University. She is studying in the field of Educational Psychology and is supervised by Dr Stephen Theiler and Dr Everarda Cunningham. Dr Everarda Cunningham is in charge of the research and is the Associate Dean of Research in the Faculty of Higher Education at Swinburne University. She has conducted research in schools for the past ten years and was also a teacher for more than 20 years.

Why are you being asked to participate?
We hope that the findings from this study will be helpful to your school and other schools in helping teachers and schools to identify children with reading and spelling problems so that they can better catered for at school. In particular, we hope that these tests can be administered by teachers in the classroom so that expensive testing is not required. We also hope that by doing this we can help provide better educational programs and policies that meet the needs of all students.

What are you being asked to do?
As a normal part of school teachers give students tests that can help identify those students that have trouble with reading and spelling. In this study your son or daughter will be asked to complete some reading, spelling and cognitive tests that their teacher and a researcher will give them during normal class time. These tests will consist of four computer based tests of approximately 10 to 15
minutes each. Your child will complete these tests one at a time spread out throughout the day. There will be two other tests that will be administered to class as a group and will take 30 to 45 minutes.

We are asking for your consent to let your son or daughter sit these tests and for us to access certain additional information the school may already have collected about students, such as the South Australian Spelling Test, Progressive Achievement Test in Reading and the Literacy section of their National Assessment Program and whether or not they have participated in the Reading Recovery program. If these tests have not yet been given to your child’s class, these tests will also be administered. We are also asking permission to come back and administer some of these tests again in three to six months time.

What are the possible risks?
There are no perceived risks associated with the research. Taking part in this study is completely voluntary. If you give your consent for your son or daughter to take part, you may withdraw your consent at any time by contacting either the school, your child’s teacher, or the university researchers.

What can you receive from this study?
Researchers will work with teachers to develop individual reports for all children who participate in the study. These reports will be the only circumstance in which your child’s individual information will be used.

Will the information be confidential and how will you find out the results of the study?
All information obtained in connection with this study that can identify your son or daughter will remain confidential. Your child’s results will be kept in locked cabinets and stored in password protected databases. In accordance with Swinburne policy and regulatory requirements, this information will be securely retained until your child is 25 years after which it will then be destroyed. At the end of the study, reports will be prepared for each of the schools taking part and for the Department of Education and Early Childhood Development. The results of this study will be published as a PhD thesis and may also be published in scientific journals and newspapers. In all instances, only group data will be presented and no individual will be identifiable. The overall results from this study will be available through your school at the end of the project.

Where to from here?
If you agree to your son/daughter taking part in this study, please fill-in the attached consent form and then ensure that the form is returned to your school. Without the completed form your son or daughter will not be permitted to participate.

Questions about the study?
If you have any queries or would like more information about this study, please contact your school or contact Sarah Callinan on 9215 7233 or by email on scallinan@swin.edu.au. Alternatively you can contact Dr. Everanda Cunningham on 9215 7316 or by email on ecunningham@swin.edu.au.

This project has been approved by or on behalf of Swinburne’s Human Research Ethics Committee (SUHREC) in line with the National Statement on Ethical Conduct in Human Research. If you have any concerns or complaints about the conduct of this project, you can contact:
Research Ethics Officer, Swinburne Research (H68),
Swinburne University of Technology, P O Box 218, HAWTHORN VIC 3122.
Tel (03) 9214 5218 or +61 3 9214 5218 or resethics@swin.edu.au

Please keep this information sheet for your own records.

In anticipation of your participation, thank you for your assistance.
Development of an Australian-Normed Measure of Reading

Dr Everarda Cunningham
Sarah Callinan
Dr Stephen Theiler

CONSENT FORM

1) I/We consent to my/our child/dependent here named to participate in the project named above. I have been provided a copy of the project information statement and this consent form and any questions I have asked have been answered to my satisfaction.

Name of Child/Dependent: .................................................................

2) Please circle your response to the following:

- We agree that s/he can be tested by the researcher: Yes No
- We agree that the test results can be recorded by electronic device: Yes No
- We agree to allow the researchers to access data from previously completed tests (Progressive Achievement Test in Reading, South Australian Spelling Test, and the National Assessment Program in Literacy): Yes No

3) I/We acknowledge that:

a) My/our child/dependent's participation is voluntary and that s/he is free to withdraw from the project at any time without explanation.

b) The project is for the purpose of research and not for profit.

c) Any information gathered in the course of and as the result of my/our child/dependent participating in this project will be (i) collected and retained for the purpose of this project and (ii) accessed and analysed by the researcher(s) for the purpose of conducting this project;

d) My/our child/dependent's anonymity is preserved and s/he will not be identified in publications or otherwise without my express written consent.

By signing this document I/we agree to your child/dependent's participation in this project.

Name of Parent(s)/Guardian: .................................................................

Signature & Date: .................................................................
Appendix F:
Administration Instructions for all the Cognitive Processing Tests

SAY: OK so all we are going to do today is a couple of ‘brain exercises’ that don’t take very long. Now do you remember when we did the test with the different patterns and you had to find the missing piece? Do you remember how they started off with easy ones and then went to really hard ones towards the end of the test? Well these exercises are the same as that, they start off with easy ones and go on to really hard ones but it’s never anything to worry about because they are hard for everyone. Does that make sense?

DO: Wait for response; make sure the student is comfortable, if not elaborate

SAY: OK so the first brain exercise is all about your memory. Now what is going to happen is I am going to say some numbers and all I want you to do is say the same numbers back to me in exactly the same order. For instance if I said 6, 2 – what would you say?

DO: Wait for student to give correct answer, if they do not, let them know what the correct answer would be and then try with the second question, 8, 1 – if they get this wrong you should continue making up number combinations and correcting incorrect responses until they get one right.

SAY: So we’ll get started then, but remember that just like in the test with the patterns, they start off with easy ones and then they go on to hard ones, so it’s nothing to worry about OK?

DO: Continue with the two digit string set and then go on to the three digit string set. Each set is made up of three separate digit strings, with the number of digits in each string increasing by one per set. The testing is stopped as soon as a student gets all three strings in one set incorrect. When this occurs:

SAY: They get really hard towards the end don’t they? OK what we are going to do next is the same thing except this time instead of saying the numbers in the SAME order that I say them, I want you to say the same numbers BACKWARDS from the order I said them. For instance if I said 7, 4 what would you say?
DO: Wait for the student to give the correct answer, if they do not, let them know what the correct answer would be and then try with the second question, 2, 5 – if they get this wrong you should continue making up number combinations and correcting incorrect responses until they get one right. Once they do, go on to the third digit set. Each set is made up of three digit strings, with the number of digits in each string increasing by one per set. The testing is stopped as soon as a student gets all three strings in one set incorrect. When this occurs:

SAY: Fantastic work! Well done! OK what we are going to do next is something a bit more fun. In a second I am going to give you this card

DO: Hold up the Colour RAN Card so that the student can see it.

SAY: And when I do all I want you to do is read out the colours on this card, left to right, line by line as fast as you can. Does that make sense?

DO: Wait for response, ensure that student understands. Then flip the card over so it is face down on the table

SAY: Great but before we do that I’m just going to have to turn on this recorder so that I can time you later

DO: Turn on the recorder and put it near the student.

SAY: Can you say your first and last name into the recorder?

DO: Wait for student to say their name. If they speak too softly or too loudly let them know. Once this is done hand over the colour RAN card

SAY: OK when you’re ready go for it

DO: Start timing when they start reading from the card. Wait for student to finish the card. When they do, take the Colour RAN card and hold up the Letter RAN card.

SAY: OK so this time we are going to do exactly the same thing but this time we are going to do it with letters instead of colours. So when you’re ready start!
DO: Hand over the Letter RAN card, start timing when they start reading from the card. Wait for the student to finish the card. When they do take the card from them and hand over the Word List booklet.

SAY: That was great, well done! OK the last thing we are going to do is some word reading. As you can see here on this first page this is just a list of words, and all I want you to do is read them out loud, one by one. But remember that just like everything else this starts off with easier words and goes on to hard ones. It’s hard for everyone so it’s nothing to worry about, I just want you to try and read each word, but if there is one you get to that you can’t read, just say pass and go on to the next one, OK?

DO: Wait for the student to get to the end of the list. If the student is struggling with the list and starts saying “pass” to whole pages, ask them to quickly read through each of the words to make sure that there are none that they could get. Once they get to the end of the list:

SAY: That was great, well done! OK one final list to go – this is just another word list for you to read out loud, but this time they aren’t real words! So what I want you to do is read each of these words out loud as best as you can, as though they were a brand new word, or a name that you hadn’t seen before.

DO: Make sure the student understands the idea of reading the nonwords. Once they do hand over the list, make sure they get two of the first three words correct in order to make sure that they understand the concept. If not go through the idea of reading new words out loud, even sounding out the words with them if necessary. Students may ask for confirmation that they are reading the words correctly as they get a little longer, if this happens just tell them that there can be different ways of pronouncing different words, and that you are interested in what they think they should sound like. Wait until the student has finished reading the entire list.

SAY: Fantastic! You’re all finished, well done!

All individual tests were marked at a later date in order to keep the time that a student spent away from class as short as possible. Please note that when administering the tests to Grade 1 and 2 students the following adjustment was made to the script:
Before the word lists were administered the following sentence was inserted in the script. “The one thing I need you to know about this word list is that it is actually a list for Grade 3 students, so if it gets a bit hard that’s nothing to worry about”.
Appendix G

Sample Report Automatically Generated through Microsoft Excel

Name: Bart Simpson
School: Manchester Primary School
D.O.B: 6/1/2000

Cognitive Ability Tests

Raven’s Progressive Matrices (RPM)
The RPM is a test that is free from language or text so that non-verbal reasoning can be tested without a student’s reading ability also being tested. Results from this test can be coded into one of the four following categories:

Category 1: 0-5: Indicates significant difficulties
Category 2: 6-10: Indicates some difficulties
Category 3: 11-25: Indicates possible difficulties
Category 4: 25 +: No indication of difficulties

Bart’s score placed him in Category 4 which indicates that he has no difficulties with non-verbal reasoning.

Processing Speed
Processing speed was tested using a task commonly known as Rapid Automatic Naming (RAN). In a RAN test students are asked to read aloud a list of colours, then letters, as quickly as they can. Bart’s letter reading time places him in the 70th percentile of students; this indicates that his processing speed is in the normal range. Bart’s colour naming time places him in the 50th percentile of students his age; this indicates that his processing speed is in the normal range. Furthermore Bart is equally good at recognising both letters and colours.

Working Memory
One of the most commonly used tests of working memory is the Digit Span Test. In this test students listen to a string of numbers and attempt to recall them, first in the
same order that they heard them and then in backwards order. Bart’s forward digit
span of 5 digits puts him in the top 40% of children his age. Bart’s backward digit
span of 3 digits puts him in the top 70% of children his age.

Educational Tests

South Australian Spelling Test (SAST)
The SAST is a long established Australian Spelling test that gives a spelling age that
can be compared to a student’s actual age. Bart’s spelling age of 9.5 was 0.32 years
above the spelling ability of someone his age, which indicates that Bart is in a
normal range for spelling when compared to children his age.

Progressive Achievement Test in Reading (PAT-R)
The PAT-R is a widely used, well respected test of reading comprehension that is
administered in schools all across Australia. Bart’s score placed him in the 91st
percentile which means that he performed better than 91% of most students his age.
Bart scored above the cut-off used to screen for possible learning problems, which
suggests that he does not have difficulty with reading comprehension.

Word Reading Tests
In the word reading tests students were presented with three lists of words that they
were asked to read aloud.

Regular words are those words that are pronounced the same way that they are spelt,
for instance cat pack or galaxy. These are the easiest words for children to learn as
they can sound them out when they are unfamiliar and use these sounds as cues
when they are trying to recall them. Bart’s score in the regular word test placed him
in the bottom 35% of readers his age.

Irregular words are words that are not pronounced the way that they are spelt, for
instance of, yacht or laugh. These words can be harder for your child to learn as they
cannot necessarily be sounded out if they do not recognise the word immediately.
However if the words are in a sentence your child can use the context that word is in
to help them work out what it is. For instance, in the sentence ‘the boy laughed so hard he nearly cried’ the word laughed cannot be sounded out, however it is possible to guess what the word is when the rest of the sentence is taken into account. Bart’s score in the irregular word list placed him in the bottom 30% of children his age.

Nonwords are words that are made up and do not have any meaning in English, for instance *nawk*, *gop* and *blurk*. These words can be tricky for children as they have not seen them before so they cannot use the context in which they come in as a clue on how to read the word. However as they are totally new words to students they are an excellent way of ascertaining how well students are able to sound out words and not how well they are able to memorise words that they have been shown earlier. Bart’s score in the nonword reading test placed him in the bottom 30% of children his age.

Because Bart was equally proficient at reading irregular words and nonwords he may benefit from practising reading both irregular and nonwords. Please find attached a word list that Bart could practise reading out loud in his spare time.

**Are you concerned about any part of this report?**

Please remember that no one test is always going to be perfect representation of your child’s abilities. These tests do not take in to account the time of day your child was tested, the kind of day your child was having when being tested or when the testing was done in the school year. Please note that testing for this report took place at different times of the year in different schools; from July to December.
## Practice Word List

<table>
<thead>
<tr>
<th>is</th>
<th>state</th>
<th>ef</th>
<th>parce</th>
</tr>
</thead>
<tbody>
<tr>
<td>ago</td>
<td>threw</td>
<td>pe</td>
<td>rilch</td>
</tr>
<tr>
<td>hoe</td>
<td>whose</td>
<td>kry</td>
<td>rouls</td>
</tr>
<tr>
<td>aunt</td>
<td>bubble</td>
<td>rup</td>
<td>scown</td>
</tr>
<tr>
<td>bear</td>
<td>cattle</td>
<td>yof</td>
<td>skisk</td>
</tr>
<tr>
<td>both</td>
<td>charge</td>
<td>bick</td>
<td>sourt</td>
</tr>
<tr>
<td>cell</td>
<td>except</td>
<td>ceam</td>
<td>stawm</td>
</tr>
<tr>
<td>dare</td>
<td>helped</td>
<td>dwaf</td>
<td>tards</td>
</tr>
<tr>
<td>four</td>
<td>island</td>
<td>goin</td>
<td>twirk</td>
</tr>
<tr>
<td>grow</td>
<td>kicked</td>
<td>halp</td>
<td>wucts</td>
</tr>
<tr>
<td>half</td>
<td>minute</td>
<td>kire</td>
<td>bousts</td>
</tr>
<tr>
<td>hour</td>
<td>smiled</td>
<td>mesk</td>
<td>bunped</td>
</tr>
<tr>
<td>lose</td>
<td>tongue</td>
<td>momp</td>
<td>fazell</td>
</tr>
<tr>
<td>mile</td>
<td>careful</td>
<td>olib</td>
<td>grized</td>
</tr>
<tr>
<td>rule</td>
<td>picture</td>
<td>pont</td>
<td>kached</td>
</tr>
<tr>
<td>sure</td>
<td>science</td>
<td>serk</td>
<td>nurfew</td>
</tr>
<tr>
<td>tube</td>
<td>tonight</td>
<td>sleb</td>
<td>plokim</td>
</tr>
<tr>
<td>baked</td>
<td>dinosaur</td>
<td>zelt</td>
<td>rakken</td>
</tr>
<tr>
<td>begin</td>
<td>February</td>
<td>zoud</td>
<td>screft</td>
</tr>
<tr>
<td>chalk</td>
<td>headache</td>
<td>broop</td>
<td>splirk</td>
</tr>
<tr>
<td>child</td>
<td>ferocious</td>
<td>corbs</td>
<td>stoult</td>
</tr>
<tr>
<td>gauge</td>
<td>shoulder</td>
<td>dwinf</td>
<td>swozed</td>
</tr>
<tr>
<td>guard</td>
<td>question</td>
<td>froke</td>
<td>skathed</td>
</tr>
<tr>
<td>house</td>
<td>tomorrow</td>
<td>heafe</td>
<td>throute</td>
</tr>
<tr>
<td>might</td>
<td>vacation</td>
<td>jimab</td>
<td>treached</td>
</tr>
<tr>
<td>paper</td>
<td>Wednesday</td>
<td>kwich</td>
<td>nefts</td>
</tr>
<tr>
<td>scene</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix H.

Method used to Program Microsoft Excel to Automatically Generate Reports

In keeping in line with the applied ethos of the current study, the method by which reading reports could be automatically generated is outlined here. This was vital as the expectation on many classroom tests is that the teachers will not only be able to, but have the time, to interpret results. This is an unreasonable burden on already high teacher workloads, as such the following steps could be taken to ensure that this does not happen.

Once all data entry was completed, reports were generated using Microsoft Excel and Microsoft Word. The template developed is intended as a blueprint for any report generating programs that would stem from this project. As such, the focus is on time-efficiency and ease of use for the teachers, so that the process remains as user-friendly as possible.

The first step was for the relevant personal information to be entered into Microsoft Excel, with one row per student. All that was required was first and last names, school, date of birth, gender and date of testing. After this, the raw scores from each test needed to be entered into the spreadsheet, this includes the raw scores from the PAT-R and Raven’s as the percentiles could be generated from these raw figures, thus saving the teachers one extra step. The spreadsheet was then programmed so that all the information needed to then form a comprehensible Word Merge function with Microsoft Word was already entered by this point. For instance, the raw score of a Grade 3 PAT-R score entered into cell A2 could be transformed into a percentile score by:

\[=\text{LOOKUP}(A2,\{0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28\},\{1,1,1,1,1,1,2,4,5,7,9,12,15,18,21,25,29,34,39,45,50,56,63,70,77,84,91,97,99\})\]

This LOOKUP function designates the appropriate percentile score to a student based on their raw score; this function was used for both established and new tests. The next challenge was to use the program to generate readable reports that did not seem stilted or table-based and were therefore user-friendly and suitable for parent
distribution if this was what the teacher wished to do. The IF function, also used in Microsoft Excel, was the primary tool used for reaching this aim.

By inserting equations such as:

=IF(B2-C2>19,"has more difficulty reading irregular words than nonwords",IF(B2-C2<-19,"has more difficulty reading nonwords than irregular words","is equally proficient at reading irregular words and nonwords"))

Where B2 represents the cell where the nonword score is entered and C2 represents the cell where the irregular score is entered. It was possible for statements regarding a student’s relative proficiency in reading irregular words and nonwords to once again be automatically generated. In this example, the equation simply examined whether a student’s percentile score for reading irregular words was significantly higher than, lower than, or roughly equal to their percentile score in nonword reading and generated the appropriate sentence that would then be inserted into the appropriate section of the reading report in Microsoft Word.
Appendix I
Normative Data for Grades Three Students on All Cognitive Processing Tests

Table II
*Grade 3 Irregular Word Preliminary Reading Norms*

<table>
<thead>
<tr>
<th>Correct Responses</th>
<th>Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>37</td>
</tr>
<tr>
<td>11</td>
<td>45</td>
</tr>
<tr>
<td>12</td>
<td>54</td>
</tr>
<tr>
<td>13</td>
<td>64</td>
</tr>
<tr>
<td>14</td>
<td>76</td>
</tr>
<tr>
<td>15</td>
<td>92</td>
</tr>
</tbody>
</table>
Table I2

*Grade 3 Digit Span Forward and Backwards Backward Preliminary Norms*

<table>
<thead>
<tr>
<th>Digit Span</th>
<th>Forward Percentile</th>
<th>Backward Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>66</td>
</tr>
<tr>
<td>5</td>
<td>37</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>74</td>
<td>97</td>
</tr>
<tr>
<td>7</td>
<td>95</td>
<td>99</td>
</tr>
<tr>
<td>8</td>
<td>99</td>
<td>99</td>
</tr>
</tbody>
</table>
Table I3

*Grade 3 Nonword Reading Preliminary Norms*

<table>
<thead>
<tr>
<th>Correct Responses</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td>13</td>
<td>40</td>
</tr>
<tr>
<td>14</td>
<td>47</td>
</tr>
<tr>
<td>15</td>
<td>54</td>
</tr>
<tr>
<td>16</td>
<td>63</td>
</tr>
<tr>
<td>17</td>
<td>72</td>
</tr>
<tr>
<td>18</td>
<td>83</td>
</tr>
<tr>
<td>19</td>
<td>91</td>
</tr>
<tr>
<td>20</td>
<td>97</td>
</tr>
<tr>
<td>Seconds</td>
<td>Percentile</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>22</td>
<td>99</td>
</tr>
<tr>
<td>23</td>
<td>99</td>
</tr>
<tr>
<td>24</td>
<td>98</td>
</tr>
<tr>
<td>25</td>
<td>97</td>
</tr>
<tr>
<td>26</td>
<td>96</td>
</tr>
<tr>
<td>27</td>
<td>94</td>
</tr>
<tr>
<td>28</td>
<td>92</td>
</tr>
<tr>
<td>29</td>
<td>90</td>
</tr>
<tr>
<td>30</td>
<td>88</td>
</tr>
<tr>
<td>31</td>
<td>85</td>
</tr>
<tr>
<td>32</td>
<td>80</td>
</tr>
<tr>
<td>33</td>
<td>74</td>
</tr>
<tr>
<td>34</td>
<td>68</td>
</tr>
<tr>
<td>35</td>
<td>60</td>
</tr>
<tr>
<td>36</td>
<td>56</td>
</tr>
<tr>
<td>37</td>
<td>51</td>
</tr>
<tr>
<td>38</td>
<td>45</td>
</tr>
<tr>
<td>39</td>
<td>40</td>
</tr>
<tr>
<td>40</td>
<td>37</td>
</tr>
<tr>
<td>41</td>
<td>32</td>
</tr>
<tr>
<td>42</td>
<td>26</td>
</tr>
<tr>
<td>43</td>
<td>20</td>
</tr>
<tr>
<td>44</td>
<td>16</td>
</tr>
<tr>
<td>45</td>
<td>15</td>
</tr>
<tr>
<td>46</td>
<td>13</td>
</tr>
<tr>
<td>47</td>
<td>12</td>
</tr>
<tr>
<td>48</td>
<td>10</td>
</tr>
<tr>
<td>49</td>
<td>8</td>
</tr>
<tr>
<td>50</td>
<td>8</td>
</tr>
<tr>
<td>51</td>
<td>7</td>
</tr>
<tr>
<td>52</td>
<td>6</td>
</tr>
<tr>
<td>53</td>
<td>4</td>
</tr>
<tr>
<td>54</td>
<td>3</td>
</tr>
<tr>
<td>55</td>
<td>3</td>
</tr>
<tr>
<td>56</td>
<td>2</td>
</tr>
<tr>
<td>57</td>
<td>2</td>
</tr>
<tr>
<td>58</td>
<td>2</td>
</tr>
<tr>
<td>59</td>
<td>1</td>
</tr>
<tr>
<td>60</td>
<td>1</td>
</tr>
<tr>
<td>Seconds</td>
<td>Percentile</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>15</td>
<td>99</td>
</tr>
<tr>
<td>16</td>
<td>99</td>
</tr>
<tr>
<td>17</td>
<td>98</td>
</tr>
<tr>
<td>18</td>
<td>96</td>
</tr>
<tr>
<td>19</td>
<td>92</td>
</tr>
<tr>
<td>20</td>
<td>88</td>
</tr>
<tr>
<td>21</td>
<td>83</td>
</tr>
<tr>
<td>22</td>
<td>76</td>
</tr>
<tr>
<td>23</td>
<td>69</td>
</tr>
<tr>
<td>24</td>
<td>64</td>
</tr>
<tr>
<td>25</td>
<td>59</td>
</tr>
<tr>
<td>26</td>
<td>54</td>
</tr>
<tr>
<td>27</td>
<td>45</td>
</tr>
<tr>
<td>28</td>
<td>38</td>
</tr>
<tr>
<td>29</td>
<td>33</td>
</tr>
<tr>
<td>30</td>
<td>27</td>
</tr>
<tr>
<td>31</td>
<td>22</td>
</tr>
<tr>
<td>32</td>
<td>18</td>
</tr>
<tr>
<td>33</td>
<td>15</td>
</tr>
<tr>
<td>34</td>
<td>13</td>
</tr>
<tr>
<td>35</td>
<td>11</td>
</tr>
<tr>
<td>36</td>
<td>8</td>
</tr>
<tr>
<td>37</td>
<td>7</td>
</tr>
<tr>
<td>38</td>
<td>6</td>
</tr>
<tr>
<td>39</td>
<td>5</td>
</tr>
<tr>
<td>40</td>
<td>4</td>
</tr>
<tr>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td>42</td>
<td>1</td>
</tr>
<tr>
<td>43</td>
<td>1</td>
</tr>
<tr>
<td>44</td>
<td>1</td>
</tr>
<tr>
<td>45</td>
<td>1</td>
</tr>
</tbody>
</table>