A Cluster Analytic Study Of The Wisc-Iv In Children Referred For Psychoeducational Assessment Due To Persistent Academic Difficulties

Corinne Hale
University of Windsor

Follow this and additional works at: https://scholar.uwindsor.ca/etd

Recommended Citation

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000 ext. 3208.
A CLUSTER ANALYTIC STUDY OF THE WISC-IV IN CHILDREN REFERRED FOR PSYCHOEDUCATIONAL ASSESSMENT DUE TO PERSISTENT ACADEMIC DIFFICULTIES

By
Corinne Hale

A Dissertation
submitted to the Faculty of Graduate Studies through the Department of Psychology in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the University of Windsor

Windsor, Ontario, Canada
2010
© Corinne Hale
NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l’Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L’auteur conserve la propriété du droit d’auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n’y aura aucun contenu manquant.
Author’s Declaration of Originality

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone’s copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.
Abstract

The purpose of this investigation was to determine the extent to which reliable and valid WISC-IV subtest patterns, consistent with those identified in taxonomic research using previous WISC editions, could be empirically derived in a sample of children referred for psychoeducational assessment. Two cluster analytic studies were conducted. In the first study, only the ten core subtests were used to form clusters. In the second study, ten core plus three supplemental subtests (i.e., Information, Picture Completion, and Arithmetic) were used to derive clusters. Through two-stage cluster analyses employing Ward’s hierarchical method followed by k-means iterative partitioning analyses, virtually the same three clusters emerged in both studies. These clusters were characterized by (1) low scores on all subtests; (2) low scores on subtests associated with the VCI; and (3) low scores on subtests associated with the WMI and PSI. These clusters were internally valid in the sense that they remained stable across first and second stages of the initial cluster analysis, were derived using four distinct clustering algorithms, and were well-replicated across various samples. The Globally Low cluster identified in Study 1 differed from the other two clusters on three WIAT-II subtests. The external validity of the other clusters remains unclear. As hypothesized, the clusters derived in this investigation have been identified in taxonomic research using previous WISC editions. Unexpectedly, clusters characterized by poor performance on subtests historically associated with the VIQ and PIQ did not emerge, nor did clusters suggesting weaknesses in perceptual reasoning (i.e., PRI), visual processing (i.e., Gv), or nonverbal fluid reasoning (Gf). Curiously, the Picture Concepts subtest represented the highest score in every cluster, failing to vary in a predictable manner with the other PRI subtests. Theoretical and clinical implications of this investigation are discussed and suggestions for future research are provided.
Dedication

I dedicate this work to the memory of my parents, Duane E. and Judith K. Hardy. There are no words to express the love, admiration, and respect I feel for you. Though not in physical form, I know we’ve walked this long journey together. To my dad, thank you for instilling in me a love of learning and an understanding that without humility, true success is unachievable. I have yet to meet a better example of a gentleman. To my mom, thank you for your unwavering support and uncanny ability to know exactly when to yell, when to cry, when to push, when to hug, when to advise, and when to listen. The world is far too quiet without your laugh. I owe you both a debt of gratitude. I hope I’ve made you proud.
Acknowledgements

I would like to take this opportunity to express my sincere gratitude to the many people who have guided me through this taxing, yet rewarding, process. First, to my advisor, Dr. Joseph Casey: Thank you for your insight, wisdom, and steadfast confidence in my abilities. Your intellect, integrity, groundedness, and wit are only part of what make you an outstanding mentor. I thank you for always having my best interest at heart, for believing in me, and for intuitively knowing when to provide a gentle nudge. It has been an honour and a pleasure to be your student.

I would also like to thank the members of my committee. To Dr. Philip Ricciardi: Thank you for providing the data that made this investigation possible. I have appreciated your unwavering enthusiasm, countless words of encouragement, and frequent reminders to "set a date." Your exceptionally calm demeanor carried me through many stressful data-coding days, and your focus on clinical relevance has kept my research meaningful. To Dr. Christopher Abeare: Thank you for your encouragement and advice throughout this process, and for the support you provided during my internship experience. Your sense of humour and down-to-earth nature are invaluable, and to some degree offset your tendency to raise challenging defense questions. To Dr. Kara Smith: Thank you for your many supportive comments, and your suggestions on earlier drafts of this document. I appreciate your interest in my research topic, and will forever remember your encouraging smiles during my presentations. To Dr. Geri Salinitri: I appreciate your willingness to chair my defense with such short notice. Thank you.

To Dr. Gerry Taylor: I feel honoured to have had you as my external examiner. Thank you for sharing your ideas and your time. Your passion for pediatric neuropsychology is apparent, and there is no question how you have achieved such a long and impressive list of accomplishments.

To Dr. Erin Picard: Thank you for providing the flexibility I needed to complete this work, and for teaching me the true meaning of the word grit. You impress me daily, and I am fortunate to have you as a mentor.

To Michelle Petherick: I treasure our friendship, and I am certain that I would not be in this position without you. I will forever cherish our exceedingly long telephone calls ranging in content from regression analysis to child birth. For your unflagging patience, advice, tolerance, and encouragement, I am profoundly grateful.

To my big sister Christine: I love you. From bags full of food, to vitamins, to work-out programs, to perfectly timed phone calls, to late-night and early morning walks, to defense attire...you have tirelessly nurtured, loved, and supported me. I could never have accomplished this without you. I hope you know how important you are to me.

To my sons Jack and Bradley: Thank you for your unparalleled love and tolerance, and your willingness to overlook my mental and physical absence. As I have consistently told you, the most important title I will ever hold is ‘Mom’ – I stand by that statement. I will secretly miss hearing your footsteps coming down the stairs at 3:30 am, blankets and pillows in tow. Thank you for those early morning hugs and for your desire to sleep on the floor next to me while I worked. It made this process bearable. I appreciate your unwavering confidence in me and your willingness to remind me “we’re okay mom...keep going” . You are my inspiration. I love you.

Finally, to my husband John: We did it! Thank you for selflessly sacrificing a normal life for a crazy life, and for putting your goals on hold to help me achieve mine. I appreciate the many ways that you have supported me, from the grandest to the smallest of gestures. Thank you for standing firmly behind me, prepared to provide whatever was needed to see me through. Your unqualified love and commitment have not gone unnoticed. You are unmistakably my better half. I love you.
Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiii</td>
</tr>
<tr>
<td>List of Appendices</td>
<td>xiv</td>
</tr>
<tr>
<td><strong>CHAPTER 1: INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Criticisms of the Discrepancy Model</td>
<td>1</td>
</tr>
<tr>
<td>1.2 The Changing Role of Intelligence Tests</td>
<td>3</td>
</tr>
<tr>
<td>1.3 The Wechsler Intelligence Scale for Children</td>
<td>5</td>
</tr>
<tr>
<td>1.4 WISC Profile Analysis</td>
<td>6</td>
</tr>
<tr>
<td>1.5 WISC Patterns of Performance in Group Research</td>
<td>9</td>
</tr>
<tr>
<td>1.6 Cluster-Analytic Methodology</td>
<td>10</td>
</tr>
<tr>
<td>1.6a Internal Validation</td>
<td>11</td>
</tr>
<tr>
<td>1.6b External Validation</td>
<td>12</td>
</tr>
<tr>
<td>1.7 Empirically-Derived WISC Performance Patterns</td>
<td>13</td>
</tr>
<tr>
<td>1.7a The WISC (1949)</td>
<td>15</td>
</tr>
<tr>
<td>Organization and Content</td>
<td>15</td>
</tr>
<tr>
<td>Construct Validity</td>
<td>17</td>
</tr>
</tbody>
</table>
The Perceptual Reasoning Index ............................................. 54
The Working Memory Index .................................................. 55
The Processing Speed Index ................................................... 56
Construct Validity ............................................................... 56
WISC-III versus WISC-IV ...................................................... 58
Summary of Changes Apparent in the WISC-IV ......................... 59
WISC-IV Patterns of Performance ........................................... 60
Summary of WISC-IV Patterns of Performance .......................... 66
Limitations of existing WISC-IV studies .................................. 67

1.8 Summary of Literature Review and Rationale for the Current Study .......... 68
1.9 Overview of the Current Study .......................................... 71

CHAPTER 2: METHOD ............................................................. 73

2.1 Study 1 .................................................................. 73
2.1a Participants .............................................................. 73

Full Sample .................................................................. 74
Split Samples .................................................................. 74
2.1b Measures ............................................................... 75

WISC-IV .................................................................. 75
WIAT-II^{CDN} .............................................................. 76
2.1c Procedures ............................................................... 77

Data Screening ............................................................. 77
Initial Cluster Analysis ....................................................... 78
Internal Validation .......................................................... 80
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical and K-means Stability</td>
<td>90</td>
</tr>
<tr>
<td>Multiple-method Reliability</td>
<td>92</td>
</tr>
<tr>
<td>Split-half Reliability</td>
<td>92</td>
</tr>
<tr>
<td>3.1d Description of Clusters</td>
<td>94</td>
</tr>
<tr>
<td>3.1e External Validation</td>
<td>101</td>
</tr>
<tr>
<td>3.2 Study 2</td>
<td>105</td>
</tr>
<tr>
<td>3.2a Data Screening</td>
<td>105</td>
</tr>
<tr>
<td>3.2b Initial Cluster Analysis</td>
<td>107</td>
</tr>
<tr>
<td>3.2c Internal Validation</td>
<td>107</td>
</tr>
<tr>
<td>Hierarchical and K-means Stability</td>
<td>107</td>
</tr>
<tr>
<td>Multiple-method Reliability</td>
<td>108</td>
</tr>
<tr>
<td>3.2d Description of Clusters</td>
<td>108</td>
</tr>
<tr>
<td>3.2e External Validation</td>
<td>115</td>
</tr>
<tr>
<td><strong>CHAPTER 4: DISCUSSION</strong></td>
<td>121</td>
</tr>
<tr>
<td>4.1 Study 1</td>
<td>121</td>
</tr>
<tr>
<td>4.1a Summary of Findings</td>
<td>121</td>
</tr>
<tr>
<td>4.1b Implications</td>
<td>124</td>
</tr>
<tr>
<td>4.2 Study 2</td>
<td>128</td>
</tr>
<tr>
<td>4.2a Summary of Findings</td>
<td>128</td>
</tr>
<tr>
<td>4.2b Implications</td>
<td>130</td>
</tr>
<tr>
<td>4.3 Limitations</td>
<td>132</td>
</tr>
<tr>
<td>4.4 Future Directions</td>
<td>135</td>
</tr>
</tbody>
</table>
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kappa values and ICCs comparing prospective solutions generated during Stages 1 and 2 of the Initial Ward’s Cluster Analysis in Study 1</td>
<td>91</td>
</tr>
<tr>
<td>2</td>
<td>Kappa values and ICCs comparing Ward’s solutions to solutions derived using three alternative hierarchical clustering techniques in Study 1</td>
<td>93</td>
</tr>
<tr>
<td>3</td>
<td>ICCs comparing mean WISC-IV profiles generated in each split-half sample in Study 1</td>
<td>95</td>
</tr>
<tr>
<td>4</td>
<td>WISC-IV means and standard deviations for the final 3-cluster solution in Study 1</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>Age and gender distribution for the final 3-cluster solution in Study 1</td>
<td>102</td>
</tr>
<tr>
<td>6</td>
<td>Means, standard deviations, and univariate F-scores for differences on WIAT-II subtests (standard scores) based on cluster membership in Study 1</td>
<td>103</td>
</tr>
<tr>
<td>7</td>
<td>Post-hoc analyses (Games-Howell Test) and effect size estimates (Cohen’s d) for mean differences between WISC-IV clusters on WIAT-II subtests in Study 1</td>
<td>106</td>
</tr>
<tr>
<td>8</td>
<td>Kappa values and ICCs comparing prospective solutions generated during Stages 1 and 2 of the Initial Ward’s Cluster Analysis in Study 2</td>
<td>109</td>
</tr>
<tr>
<td>9</td>
<td>Kappa values and ICCs comparing Ward’s solutions to solutions derived using three alternative hierarchical clustering techniques in Study 2</td>
<td>110</td>
</tr>
<tr>
<td>10</td>
<td>WISC-IV means and standard deviations for the final 3-cluster solution in Study 2</td>
<td>116</td>
</tr>
<tr>
<td>11</td>
<td>Age and gender distribution for the final 3-cluster solution in Study 2</td>
<td>117</td>
</tr>
<tr>
<td>12</td>
<td>Means, standard deviations, and univariate F-scores for differences on WIAT-II subtests (standard scores) based on cluster membership in Study 2</td>
<td>119</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Internal validation procedures for Study 1</td>
<td>81</td>
</tr>
<tr>
<td>2</td>
<td>Mean WISC-IV profile for Cluster 1 in Study 1 (Globally Low)</td>
<td>96</td>
</tr>
<tr>
<td>3</td>
<td>Mean WISC-IV profile for Cluster 2 in Study 1 (Low VCI)</td>
<td>97</td>
</tr>
<tr>
<td>4</td>
<td>Mean WISC-IV profile for Cluster 3 in Study 1 (Lower WMI and PSI)</td>
<td>98</td>
</tr>
<tr>
<td>5</td>
<td>Mean WISC-IV profiles for all clusters in Study 1</td>
<td>99</td>
</tr>
<tr>
<td>6</td>
<td>Mean WIAT-II profiles for the WISC-IV cluster derived in Study 1</td>
<td>104</td>
</tr>
<tr>
<td>7</td>
<td>Mean WISC-IV profile for Cluster 1 in Study 2 (Globally Low(^2))</td>
<td>111</td>
</tr>
<tr>
<td>8</td>
<td>Mean WISC-IV profile for Cluster 2 in Study 2 (Low VCI(^2))</td>
<td>112</td>
</tr>
<tr>
<td>9</td>
<td>Mean WISC-IV profile for Cluster 3 in Study 2 (Low WMI and PSI(^2))</td>
<td>113</td>
</tr>
<tr>
<td>10</td>
<td>Mean WISC-IV profiles for all clusters in Study 2</td>
<td>114</td>
</tr>
<tr>
<td>11</td>
<td>Mean WIAT-II profiles for the WISC-IV cluster derived in Study 2</td>
<td>120</td>
</tr>
</tbody>
</table>
List of Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>Summary of empirical clustering studies using the WISC-R</td>
</tr>
<tr>
<td>B.</td>
<td>Summary of empirical clustering studies using the WISC-III</td>
</tr>
<tr>
<td>C.</td>
<td>Relationship of the WISC-IV subtests to the Indexes and FSIQ</td>
</tr>
<tr>
<td>D.</td>
<td>Description of selected WISC-IV subtests</td>
</tr>
<tr>
<td>E.</td>
<td>Description of selected WIAT-II subtests</td>
</tr>
<tr>
<td>F.</td>
<td>Figure F1. Abbreviated example of an agglomeration schedule</td>
</tr>
<tr>
<td></td>
<td>Figure F2. Abbreviated example of a dendrogram</td>
</tr>
<tr>
<td>G.</td>
<td>Table G1. Kappa value interpretation system used in present investigation</td>
</tr>
<tr>
<td>H.</td>
<td>Figure H1. WISC-IV profile of Multivariate Outlier 1 removed from analyses</td>
</tr>
<tr>
<td></td>
<td>Figure H2. WISC-IV profile of Multivariate Outlier 2 removed from analyses</td>
</tr>
<tr>
<td></td>
<td>Figure H3. WISC-IV profile of Multivariate Outlier 3 removed from analyses</td>
</tr>
<tr>
<td></td>
<td>Figure H4. WISC-IV profile of Multivariate Outlier 4 removed from analyses</td>
</tr>
<tr>
<td></td>
<td>Figure H5. WISC-IV profile of Multivariate Outlier 5 removed from analyses</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Intelligence tests have long been used in the psychological assessment of children exhibiting persistent academic difficulties\(^1\) (Flanagan & Kaufman, 2009). Over the last decade, however, the manner in which these tests are used has changed dramatically. Within the context of psychoeducational assessment, intelligence tests have historically been employed to rule out the presence of a broad-based intellectual disability or identify a learning disability based on the existence of a significant discrepancy between global intellectual functioning and academic achievement (Dombrowski, Kamphaus, & Reynolds, 2004). In recent years, this latter practice, referred to as the discrepancy model, has faced staunch criticism in the scientific literature due to a lack of theoretical, psychometric, and empirical support (Dombrowski et al., 2004; Kavale & Forness, 2003; Meyer, 2000; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Fletcher, Denton, & Francis, 2005; Reschly & Hosp, 2004; Sternberg & Grigorenko, 2002; Stuebing et al., 2002). Following is a brief review of this issue.

1.1 Criticisms of the Discrepancy Model

Numerous criticisms have been leveled at the discrepancy model over the last decade. One of the most widely cited problems with this practice relates to the assumption that intelligence tests assess an individual's capacity for learning (Meyer, 2000). According to this assumption, IQ necessarily sets an upper limit on a person's educational outcome, as it is conceptually impossible to achieve beyond that which one is

---

\(^1\) To avoid confusion related to the wide variety of labels and definitions used in the literature, unless otherwise specified, the term persistent academic difficulties will be used throughout this manuscript to describe conditions in which significant and persistent academic weaknesses are apparent despite adequate educational opportunities and in the absence of a broad-based intellectual disability.
capable of achieving. That many children perform better on achievement tests than on measures of intelligence argues against this inherent assumption (Siegel, 1999).

Another criticism of the discrepancy model involves the belief that intelligence and academic achievement reflect distinct constructs and can, therefore, be measured independently (Siegel, 1989). A considerable body of research suggests that intelligence tests and achievement tests are highly intercorrelated leading many psychologists to conclude that they likely measure similar processes (Flanagan & Kaufman, 2004; Francis et al., 1996; Sternberg & Grigorenko, 2002). Thus, underlying cognitive weaknesses would be expected to affect intellectual and academic test performance in a similar manner, thereby diminishing the ability-achievement discrepancy.

Similarly, many psychologists have criticized the discrepancy model based on a phenomenon referred to as the Matthew Effect, a biblical reference to the idea that the rich get richer and the poor get poorer (Stanovich, 1986). According to this perspective, children who are skilled readers are in a better position than unskilled readers to develop an adequate vocabulary, acquire general knowledge, and develop the ability to comprehend information about the world. This in turn is reflected in better performance on standardized measures of intelligence. Conversely, children who experience reading difficulties are less likely to develop the requisite knowledge to perform well on intelligence tests, which results in diminished IQ-achievement discrepancies and difficulty qualifying for special education services (Dombrowski et al., 2004).

Myriad other criticisms of the discrepancy model have been articulated in the literature, including psychometric issues related to the limited reliability of difference scores and failure to account for statistical regression to the mean (Sternberg &
Grigorenko, 2002). Further, validity studies have found that children with significant academic difficulties formed on the basis of the presence or absence of an ability-achievement discrepancy do not differ appreciably with respect to response to instruction (Fletcher, Denton, & Francis, 2005), nor do they differ in terms of long-term prognosis (Francis, Shaywitz, et al., 1996).

Perhaps the most devastating criticism of the discrepancy model is that it delays the provision of special education services until a time when the discrepancy criterion is finally met (Stuebing et al., 2002; Taylor, Anselmo, Foreman, Schatschneider, & Angelopoulos, 2000). Statistically, it is difficult for children with even severe academic difficulties to meet the discrepancy criterion until grade 3 or 4 when they fall far enough behind peers. At this time, effective interventions are much more difficult to implement (Stuebing et al.). This ‘wait-to-fail’ model limits the possibility of a good outcome for these struggling students and can have a negative impact on self-esteem and self-concept (Meyer, 2000).

1.2 The Changing Role of Intelligence Tests

Reflecting growing concern regarding the discrepancy model, the latest re-authorization of the Individuals with Disabilities Education Act (IDEA, 2004) permits schools to abandon this model of assessment in favor of a response to intervention approach (RTI; Machek & Nelson, 2007; Smith, 2005). Although a thorough discussion of the RTI approach is beyond the scope of the current work, the interested reader is referred to Fletcher, Lyon, Fuchs, & Barnes (2007) and Hollenbeck (2007) for comprehensive reviews of this topic. Briefly, RTI is a process involving the provision of early evidence-based instructional intervention for all students, frequent progress
monitoring, and increasingly intensive and individualized interventions for children who exhibit persistent academic difficulties. According to this model, failure to show improvement to this series of interventions (i.e., failure to respond) leads to learning disability classification (Reschly, 2004) or to a comprehensive neuropsychological assessment conducted to determine the socioemotional, behavioral, and cognitive profiles of these children (Feifer, 2008; Hale, Kaufman, Naglieri, & Kavale, 2006; LDA, 2010; Schmitt & Wodrich, 2008).

The changes apparent in IDEA (2004) have left many psychologists questioning the role of intelligence tests in the assessment of children with persistent academic difficulties (Mather & Gregg, 2006). While some scholars have argued that intelligence testing should be abandoned entirely (Siegel, 1989; 1999; Stanovich, 1999) or used solely to rule out the presence of an intellectual disability (Meyer, 2000), most professionals have called for a more balanced approach to this issue, which views intelligence testing as an integral part of a comprehensive assessment conducted to elucidate the psychological profiles of children who fail to respond to evidence-based academic interventions (Berninger & O'Donnell, 2005; Dombrowski et al., 2004; Fletcher-Janzen & Reynolds, 2008; Hale et al., 2006; Kaufman, 2009; Machek & Nelson, 2007; Mascolo, 2004; Wodrich, Spencer, & Daley, 2006). Although not universally accepted (e.g., Dana & Dawes, 2007; McDermott, Fantuzzo, & Glutting, 1990; Watkins, Glutting, & Lei, 2007), proponents of the latter perspective generally agree that the value of intelligence tests lies in their ability to capture a wide range of cognitive strengths and weaknesses, which when integrated with other sources of information (e.g., data from other norm-referenced tests, parent/teacher reports, direct observations, curriculum-based
assessments, and informed clinical judgment), has the potential to facilitate the
development of effective interventions and accommodations (Dombrowski et al., 2004; Flanagan & Kaufman, 2009; Hale Fiorello, Kavanaugh, Holdnack, & Aloe, 2007; Kavale, Holdnack, & Mostert, 2006; Mather & Gregg, 2006; Naglieri & Paolitto, 2005; Schrank, Miller, Catering, & Desrochers, 2006).

1.3 The Wechsler Intelligence Scale for Children

Despite the availability of myriad individually-administered intelligence tests, the Wechsler Intelligence Scale for Children (WISC) 2 unequivocally represents the most widely used measure of children’s intelligence worldwide (Flanagan & Kaufman, 2009; Kaufman et al., 2006). This instrument, which surveys a broad range of cognitive abilities, has guided the science and practice of childhood intellectual assessment for over 50 years (Flanagan, McGrew, & Oritz, 2000). Since its original publication in 1949, the WISC has undergone three revisions (WISC-R, 1974; WISC-III, 1991; WISC-IV, 2003). The latest edition of this instrument, the WISC-IV, is structurally distinct from its predecessors, and represents the most extensive revision ever made to a Wechsler scale (Flanagan & Kaufman, 2009; Keith, Fine, Taub, Reynolds, & Kranzler, 2006).

Despite the widespread use of this instrument within school settings, few studies have investigated WISC-IV performance in children exhibiting persistent academic difficulties. In the face of scant research, it has been recommended that practitioners interpret the WISC-IV on the basis of empirical findings generated on previous WISC versions (e.g., Gabel & Shaughnessy, 2006; Prifitera, Saklofske, & Weiss, 2005; Wechsler, 2003c; Williams, Weiss, & Rolphus, 2003a). This practice is disconcerting

---

2 Unless otherwise specified, the term WISC is used to refer to the WISC lineage which includes the original WISC (1949), the WISC-R (1974), the WISC-III (1991), and the WISC-IV (2003).
considering that the manifold changes apparent in the WISC-IV likely alter, in a complex fashion, the type of information obtained from intellectual assessment (Strauss, Spreen, & Hunter, 2000). Thus, the extensive changes made to the major scales of this instrument may limit the generalizability of previous WISC research to this latest measure (Flanagan & Kaufman, 2009). Considering this, understanding how the WISC-IV compares to its predecessors is important in that failure to do so may lead to erroneous interpretations and misguided recommendations (Strauss et al.).

1.4 WISC Profile Analysis

Since its creation, clinicians and researchers have debated the manner in which the WISC should be interpreted (Fiorello et al., 2007; Zachary, 1990). While a number of psychologists espouse the use of the global IQ score, contending that this is the only level at which there is adequate reliability, validity, and clinical utility (Dana & Dawes, 2007; McDermott, Fantuzzo, & Glutting, 1990), others suggest that the Index scores represent the most appropriate interpretive level (Donders, 1996; Naglieri & Paolitto, 2005). Finally, many practitioners argue that the most clinically meaningful approach to interpreting the WISC involves examining patterns of performance on the individual subtests that compose the more global scores (Fiorello et al., 2007; Flanagan & Kaufman, 2009; Hale & Fiorello, 2004; Hynd, Riccio, Cohen, & Arceneaux, 1998; Groth-Marnat, Gallagher, Hale, & Kaplan, 2000; Kaufman, 1994; Lezak, 1995). Proponents of the latter approach contend that patterns of subtest scores provide valuable clinical information about an individual’s strengths and weaknesses that is lost when analyses are limited to the more global Index scores or the omnibus IQ (Flanagan & Kaufman, 2009;
This popular but controversial practice, commonly referred to as profile analysis, has received considerable attention in recent years, particularly with respect to its use in the evaluation of children with persistent academic difficulties (Groth-Marnat, 2001; McDermott, Fantuzzo, & Glutting, 1990; Reynolds, 2007). Over the years, many criticisms have been leveled at the practice of profile analysis. Those in opposition admonish this interpretive approach primarily on the grounds of insufficient reliability and specificity of subtest scores (Livingston, Jennings, Reynolds, & Gray, 2003; McDermott et al., 1990; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992) and limited diagnostic utility (Daley & Nagle, 1996; Dana & Dawes, 2007; Kavale & Forness, 1984). Conversely, proponents of profile analysis argue that “this interpretive strategy is both clinically justified and empirically supported” (Hale & Fiorello, 2004 pg. 23), and although many of the WISC subtest reliability and specificity coefficients may not be adequate for high-stakes diagnostic purposes, they are adequate for generating hypotheses regarding cognitive strengths and weaknesses (Groth-Marnat, 2001; Hale & Fiorello, 2004; Kaufman, 1994; Sattler, 2001; Yeates & Donders, 2005). In an effort to summarize the ongoing debate in this area, Stanton and Reynolds (2000) suggest that “clinicians who focus on the relationship between groups of participants appear to support the clinical practice of profile analysis, whereas statisticians who focus on the relation between variables strongly oppose the practice” (p. 434).

Notwithstanding the foregoing controversy, it is clear that profile analysis is the preferred method of WISC interpretation. Nearly 90% of school psychologists report
using profile analysis to interpret the WISC, and almost 70% believe that the ability to conduct this level of analysis represents one of the most valuable features of this instrument (Pfeiffer, Reddy, Kletzel, Schmelzer, & Boyer, 2000). Seventy-four percent of school psychology graduate training programs emphasize the use of profile analysis (Alfonso, Oakland, LaRocca, & Spanakos, 2000), and many preeminent assessment textbooks devote considerable attention to this level of interpretation (e.g., Flanagan & Kaufman, 2004; Hale & Fiorello, 2004; Lezak, 1995; Prifitera, Saklofske, & Weiss, 2005; Sattler, 2004).

In the early days of the WISC, psychologists hoped that profile analysis would be useful with respect to differentiating children with learning difficulties from typically-achieving children. Research, however, has not supported the use of WISC profile analysis in isolation for diagnostic purposes (Sattler & Dumont, 2004; Prifitera, Saklofske, Weiss, & Rolfhus, 2005). In contemporary practice, profile analysis has a much different purpose. It is used in combination with other sources to: “(a) clarify the functional nature of a child’s learning problems; (b) arrive at treatment recommendations, (c) develop educational programs, and (d) recommend a vocational placement” (Sattler, 2001, p. 299). Reflecting this, a best-practice approach to psychoeducational assessment currently includes the use of WISC profile analysis as a tool for generating individualized hypotheses regarding the cognitive ability structures of children with persistent academic weaknesses. This data is then integrated with information from other sources, with the overarching goal of developing effective interventions and appropriate recommendations for these children (Berninger & O'Donnell, 2005; Berninger, Dunn, & Alper, 2005;
1.5 WISC Patterns of Performance in Group Research

Given the popularity of the WISC, it is no surprise that over the years there has been considerable interest in understanding how individuals within specific groups perform on this instrument. Reflecting this, there is an extensive body of scientific literature exploring WISC profiles in clinical, and referred samples (e.g., Allen, Thaler, Donohue, & Mayfield, 2010; Anderson & Gordon, 1992; Mayes & Calhoun, 2008; Moore & Wielan, 1981; Robinson & Harrison, 2005; Romi & Marom, 2007; Waxman & Casey, 2006; Williams, Weiss, & Rolfhus, 2003c). Much of this research has involved forming groups based on a common characteristic (e.g., clinical diagnosis or reason for referral) and reporting group mean data across the range of WISC variables (i.e., mean subtest or index score profiles; e.g., Bolte, Rudolf, & Poustka, 2002; Hubble & Groff, 1980; Moore & Wielan, 1981; Romi & Marom, 2007; Wechsler, 2003c). Unfortunately, this research design suffers from inherent limitations that render it inadequate for exploring WISC performance patterns (Lange, 2007). Specifically, no matter how clearly defined a sample may be, group mean data obscures performance variability. Thus, this type of research fails to take into consideration the heterogeneity of the individuals within a given sample, thereby providing only minimal information regarding performance patterns of a particular group (Lange).

Cluster analysis offers an alternative method for exploring patterns of performance, and has made a significant contribution to research exploring Wechsler Scale profiles in group data (Lange, 2007). In contrast to the above described group-means method,
Cluster analysis is specifically designed to capture the performance heterogeneity within a sample, arguably making it the most appropriate method for delineating WISC patterns of performance in group studies (Glutting, McDermott, & Konold, 1997; Konold, Glutting, McDermott, Kush & Watkins, 1999; Lange, 2007). Following is a review of this methodology.

1.6 Cluster-Analytic Methodology

Cluster analysis refers to a family of nonlinear multivariate statistical techniques that are used to assign objects (e.g., children) into subgroups on the basis of a set of selected variables (e.g., WISC subtest scores). The goal of cluster analysis is to produce subgroups that evidence high within-cluster homogeneity and high between-cluster heterogeneity (Hair & Black, 2000). Thus, the primary objective is to assemble distinct subgroups of individuals who are similar to one another across a predetermined set of variables (Aldenderfer & Blashfield, 1984).

With the use of cluster analysis, the researcher is required to make a series of decisions regarding the manner in which object similarity will be defined, and the specific clustering algorithm that will be used to determine whether an individual belongs to a particular cluster (Morris, Blashfield, & Satz, 1981; Milligan & Cooper, 1987). Further, the researcher must determine the most appropriate cluster solution (i.e., the number of clusters that best represents the dataset); a process which requires numerous objective and subjective considerations (Aldenderfer & Blashfield, 1984; Hair & Black, 2000; Milligan & Cooper). Once the groups have been established, variable means are plotted in order to derive representative profiles for each cluster, which in turn guides the interpretation of the cluster solution (Hair & Black; Speece, 2003).
As with any technique, cluster analysis has its strengths and weaknesses. Many researchers see the availability of multiple clustering algorithms a benefit of this technique in that it affords research-design flexibility (Everitt, 1980) and permits replication of findings using a variety of algorithms (Fletcher & Satz, 1985). Further, given that cluster analysis does not assume a linear model, it has been suggested that this method has an advantage over other empirical clustering techniques (e.g., Q-type factor analysis) in that it is sensitive to profile variations in level and shape (Hair & Black, 2000; Konold et al., 1999), both of which represent areas of interest in research on children with academic difficulties (Glutting et al., 1997).

In terms of limitations, with the use of cluster analysis, all participants in a sample are forced into clusters on the basis of relative similarity to other participants without consideration of similarity in an absolute sense (DeLuca, Adams, & Rourke, 1991; Everitt, 1984; Hair & Black, 2000). Thus, subgroups generated through cluster analysis are likely to include some individuals who bear only minimal similarity to the mean profile derived for that cluster. It is also important to note that, through the use of this technique, clusters will be formed even in randomly-generated data (Morris et al., 1981). As such, the investigator is required to take painstaking efforts to ensure the internal validity (i.e., stability and reliability) of the final cluster solution (Aldenderfer & Blashfield, 1984; Skinner, 1981).

1.6a Internal Validation

Critical to identifying meaningful performance patterns using cluster analytic methodology is the process of internal and external validation (Fletcher & Satz, 1985; Hair & Black, 2000). Essentially, the process of internal validation involves evaluating
the reliability of the derived typology. For a typology to be considered reliable it should not be dependent on a specific clustering method or algorithm for its derivation and it should be able to be replicated in multiple samples (Skinner, 1981). Further, it should successfully classify an adequate number of individuals (Skinner).

One way to evaluate the internal validity of an empirical typology is to apply numerous clustering algorithms to the same sample in order to determine the extent to which similar representative profiles are formed (Speece, 2003). Similarly, two distinct clustering techniques (e.g., cluster analysis and Q-type factor analysis) can be applied to the same data and the resulting profiles compared (Fuerst, Fisk, & Rourke, 1989).

With respect to evaluating the replicability of a cluster solution across samples, one of two methods is generally employed in research on children with academic difficulties: split-sample or single-sample designs (Speece, 2003). Split-sample replication involves randomly dividing a given dataset in half, clustering each half separately, and statistically comparing the results (Hair & Black, 2000). Single-sample replication, which is considered to be less powerful than split-sample replication, involves either clustering a subset of the original dataset and comparing the results to the cluster solution derived using the full sample or adding additional subjects or variables to the analysis in an attempt to create 'noise' in the sample. In theory, if true structure exists in the data, the same mean profiles will be derived despite these additional subjects or variables (Speece).

1.6b External Validation

Once it is established that the empirically-derived subgroups are adequately reliable (i.e., internally valid), it is important to evaluate the extent to which they are externally
valid. That is, in order to be clinically meaningful, the derived subgroups should differ on relevant variables not used to establish the typology (Skinner, 1981). This external validation process should be predicated, at least in part, on a priori hypotheses regarding the manner in which the derived subgroups would be expected to differ based on previous research or theory (Fletcher, 1985; Speece, 2003). For example, in the case of empirical clustering research using the WISC in children with academic difficulties, the derived subgroups would be expected to bear an empirical relationship to variables such as academic achievement patterns, neuropsychological profiles, socioemotional status, or response to specific interventions and accommodations (Fiorello et al., 2007; Mascolo, 2004; Berninger & O'Donnell, 2005).

1.7 Empirically-Derived WISC Performance Patterns

Numerous studies have used cluster analysis to explore WISC patterns of performance in groups of children exhibiting persistent academic difficulties (e.g., Holcomb, Hardesty, Adams, & Ponder, 1987; Snow, Cohen & Holliman, 1985; Waxman & Casey, 2006). Whereas initially it was anticipated that profiles unique to these children would emerge (Sattler, 2001), research has not supported this notion. Studies exploring this issue consistently report that between 70% and 90% of children with academic difficulties display subtest profiles matching those found in the WISC standardization sample (e.g., Borsuk, Watkins, & Canivez, 2006; Watkins and Kush, 1994). Considering this, it is generally agreed that WISC profiles should not be used in isolation for diagnostic purposes (Sattler, 2001; Watkins 2003); however, as suggested earlier, these patterns may be useful in terms of generating hypotheses regarding effective educational interventions and accommodations (Berninger & O’Donnell, 2005; Fiorello,
Hale, & Snyder, 2006; Flanagan & Kaufman, 2009; Geary, 2005; Hale, Fiorello, Miller, Wenrich, Teodori, & Henzel, 2008; Reynolds, 2008; Riccio, 2008, Schmitt & Wodrich, 2008). That is, similar to what would be expected for typically-achieving students (Groth-Marnat, 2001), individualized educational programming that incorporates knowledge of cognitive strengths and weaknesses would be expected to maximize the performance of children with persistent academic difficulties (Allen & Hancock, 2008; Decker, 2008; Fiorello et al., 2006; Taylor, 2008). Although this position is not universally embraced (e.g., Fletcher, Lyon, Fuchs, & Barnes, 2007; Reschly, 2004), it has been endorsed by numerous professional organizations (e.g., National Academy of School Psychologists, 2007; Learning Disabilities Association of America, 2010; Schrank et al., 2006) and supported by research (Allen & Hancock, 2008; Goldstein & Naglieri, 2008; Hale et al., 2008). Thus, identifying empirically-derived WISC patterns in children with academic difficulties is important in that it has the potential to inform instructional research aimed to optimize psychoeducational outcomes.

In contrast to a substantial body of literature using empirical clustering methods with previous WISC editions, this methodology has not been widely used to explore patterns of performance using the WISC-IV. Despite the widespread use of this measure with children referred for psychoeducational assessment, no study has attempted to empirically derive patterns of WISC-IV performance in this population. The current study will address this gap in the literature. Before moving on to a description of the current investigation, however, a review of the evolution of the WISC will be provided focusing on the structural changes that were made to this instrument during each revision, and the empirical clustering research that has been conducted using each of its versions.
Considering the WISC-IV within its historical context is important in terms of making determinations regarding the extent to which previous cluster analytic research involving children with academic difficulties can or should be generalized to this new measure.

1.7a The WISC (1949)

In 1949, the WISC was developed as a downward extension of the adult Wechsler-Bellevue Intelligence Scale (W-B; Wechsler, 1939). The majority of items on the WISC were taken directly from the W-B, with easier items added to the low end of each subtest in order to make the instrument applicable to children ages 5 to 15 years (Boake, 2002; Littell, 1960; Wechsler, 1949b).

Organization and Content

The original WISC comprised 12 subtests equally divided into two scales labeled *Verbal* and *Performance*. Three deviation Intelligence Quotients (i.e., Verbal, Performance, and Full Scale; VIQ, PIQ, and FSIQ, respectively) and 12 scaled subtest scores were derived on the basis of an individual’s raw scores. Consistent with all subsequent versions of this instrument, mean composite scores were set to 100, with a standard deviation of 15; scaled score means were set to 10, with a standard deviation of 3 (Wechsler, 1949b).

The Verbal Scale comprised six subtests: (1) *Information* (a subtest requiring the child to answer a broad range of questions related to general knowledge topics); (2) *Comprehension* (a subtest requiring the child to respond to questions and solve everyday problems on the basis of his/her understanding of general principles and social situations); (3) *Arithmetic* (a subtest requiring the child to mentally solve a series of orally presented arithmetic problems within a specific amount of time); (4) *Similarities* (a
subtest requiring the child to identify similarities between two words or concepts); (5) *Vocabulary* (a subtest requiring the child to provide definitions of words read aloud by the examiner); and (6) *Digit Span* (a subtest requiring the child to repeat a series of numbers presented by the examiner in forward and reverse order).

Like the Verbal Scale, the Performance Scale comprised six subtests: (1) *Picture Completion* (a subtest requiring the child to identify missing parts of pictures within a specified number of seconds); (2) *Picture Arrangement* (a subtest requiring the child to arrange a set of pictures in logical order); (3) *Block Design* (a subtest requiring the child to rapidly recreate designs presented as either a model or a picture using blocks); (4) *Object Assembly* (a subtest involving putting puzzle pieces together rapidly to form common objects); (5) *Coding* (a subtest requiring the child to rapidly copy novel symbols paired with more familiar or meaningful symbols); and (6) *Mazes* (a subtest requiring the child to solve a series of increasingly difficult paper-and-pencil mazes).

Although all twelve WISC subtests were included in the standardization process, in the interest of shortening the time required to administer the battery only ten subtests (five from each scale) were used to establish the IQ tables. *Digit Span* from the Verbal Scale and *Mazes* from the Performance Scale were omitted on the basis of relatively low correlations with other subtests from their respective scales, and in the case of Mazes, because of its lengthy administration time (Wechsler, 1949b). Despite the two subtest omissions, the WISC manual encouraged clinicians to use all 12 subtests of the battery whenever possible due to the additional qualitative and diagnostic data gained by administering these tests (Wechsler, 1949b; Seashore, Wesman, & Doppelt, 1950; Littell, 1960).
Construct Validity

With respect to construct validity, the WISC manual reported composite and subtest score intercorrelations supporting the tripartite structure of the instrument. Correlations between the Verbal and Performance Scales were high enough to suggest considerable common variance (providing empirical support for the Full Scale), yet were low enough to indicate that unique abilities were being measured by each scale (providing support for the VIQ-PIQ division; Seashore et al., 1950). In terms of specific subtest correlation patterns, as anticipated, Verbal subtests correlated more highly with other Verbal subtests than with Performance subtests, and vice versa (Wechsler, 1949b).

Although the results of factor analytic studies were not reported in the original WISC manual, numerous independent investigations using this methodology were conducted after its publication (Littell, 1960). The presence of a general intellectual factor was consistently supported by this research, as was the Verbal-Performance dichotomy. Other factors frequently identified in the literature included a Verbal Comprehension factor (composed of Information, Comprehension, Similarities, and Vocabulary subtests - later termed the Verbal Comprehension Index; VCI), a Perceptual Organization factor (composed of Picture Completion, Picture Arrangement, Block Design, and Object Assembly subtests – later termed the Perceptual Organization Index; POI), and a third, less robust, factor composed of Arithmetic, Digit Span, and Coding (Cohen, 1959; Gault, 1954, Maxwell, 1959).

This latter factor was initially labeled *Freedom from Distractibility* (FDI) emphasizing the attention-concentration demands common to this subtest grouping.
(Bortner & Birch, 1969; Cohen, 1959). Subsequent interpretations of the third factor highlighted the mental sequencing abilities required to complete these tasks (Bannatyne, 1974), the susceptibility of these subtests to test anxiety (Lutey, 1977 cited in Saklofske, Weiss, Beal, & Coalson, 2003), the requirement for numerical or quantitative skills (Osborne & Lindsey, 1967 cited in Saklofske et al.), and the appreciable demands on short-term and working memory (Cohen, 1957; Wielkiewicz, 1990). While all of these interpretations held some merit (Saklofske et al.), it is apparent that, at the time, the most widely accepted interpretation of this factor involved the notion of attention-concentration, as the FDI label was carried forward for use with subsequent versions of the WISC.

WISC Patterns of Performance

During the early years of the WISC (1949), students with persistent academic difficulties were generally considered to represent a homogeneous entity such that an individual who struggled academically despite adequate educational opportunities was assumed to share many features with other children so afflicted (Rourke, 1989). Reflecting this, the bulk of research examining WISC performance in children with persistent academic difficulties attempted to identify a single profile representative of all struggling students (Rugel, 1974; Kender, 1972; Rourke). Studies employing this single syndrome theory were traditionally carried out by comparing the mean WISC profiles derived from samples of children with academic difficulties to the mean profiles obtained from groups of typically-achieving students or children with other challenges (McKinney, 1984; Rourke). This so called contrasting-groups approach led to a substantive body of literature indicating that no single pattern of WISC performance
characterized all children with persistent academic difficulties (Kavale & Forness, 1984; Rourke).

Specifically, depending on the sample employed, these children were found to be impaired on WISC subtests composing the Verbal Scale, the Performance Scale, or those subtests associated with the Freedom from Distractibility factor (Dudley-Marling, Kaufman, & Tarver, 1981; Huelsman, 1970). Moreover, many contrasting-group studies reported the ACID pattern in children with LD (depressed scores on Arithmetic; Coding, Information, and Digit Span; Rugel, 1974; Watkins, Kush, & Glutting, 1997); a robust profile which early on was thought to predict reading problems, and carried with it a particularly poor prognosis in all academic areas (Joschko & Rourke, 1985). Further demonstrating the heterogeneity of findings in early research, some studies reported relatively ‘flat’ mean WISC profiles in groups of struggling students (Miller, Stoneburner, & Brecht, 1978), with the majority of scores either mildly depressed, in the average range, or even slightly elevated (Kavale & Forness, 1984).

With the flurry of research in the area, it became increasingly clear that a single syndrome theory of academic difficulties could not account for the many cognitive processing patterns observed in this population (McKinney, 1984). Accordingly, a number of scholars recommended that the search for a pattern of cognitive deficits unique to children with academic limitations be abandoned in favor of a multiple syndrome paradigm which could accommodate the heterogeneous nature of academic problems and aid in differential programming efforts (McKinney). Reflecting this theoretical shift, a number of researchers began to explore various subgroups of children with academic difficulties. Presumably due to a lack of theoretical and technological sophistication,
these early studies relied exclusively on clinical-inferential approaches to classification in which children were sorted into subgroups on the basis of a priori theoretical criteria, rather than on the basis of empirical clustering techniques (McKinney).

In one such investigation, Rourke, Young, & Flewelling (1971) explored the heterogeneity of learning difficulties based on WISC patterns of performance. In this study, Rourke and his colleagues divided a sample of children with significant academic weaknesses into three groups on the basis of WISC VIQ/PIQ patterns. Group 1 consisted of children whose PIQ scores were at least 10 points greater than their VIQ scores (PIQ>VIQ Group); Group 2 was composed of children who achieved PIQ and VIQ scores that were within 4 points of each other (PIQ=VIQ Group); and Group 3 comprised children with PIQ scores at least 10 points lower than their VIQ scores (VIQ>PIQ). As predicted, patterns of neuropsychological and academic performance varied as a function of group membership. Children in the PIQ>VIQ group performed significantly better than the other two groups on neuropsychological measures of visual-perceptual functioning and showed a trend toward better Arithmetic skills relative to Reading and Spelling. Conversely, the VIQ>PIQ group performed better on neuropsychological tasks tapping verbal and auditory skills, and evidenced a pattern of academic performance characterized by better Reading and Spelling relative to Arithmetic. The PIQ=VIQ Group essentially performed intermediate to the other two groups on the majority of variables included in the study.

Subsequent investigations conducted by these researchers further demonstrated the external validity of this clinical-inferential typology by reporting marked differences between the three groups on complex motor/psychomotor tasks (Rourke, Dietrich, &
Young, 1973; Rourke & Telegdy, 1971) and patterns of personality functioning (Fuerst, Fisk, & Rourke, 1990). The results of these investigations led Rourke (1989) to conclude that valid subgroups characterized by unique patterns of neuropsychological, academic, and psychosocial functioning can be identified on the basis of WISC performance in children exhibiting significant academic difficulties.

**Summary of WISC Patterns of Performance.** Sundry results were obtained from contrasting-group studies involving children with persistent academic difficulties including relatively depressed scores on subtests composing the Verbal, Performance, and Freedom from Distractibility Indexes, low scores on the ACID tetrad, and relatively flat profiles reflecting globally depressed, average, or elevated subtest scores. These diverse findings eventually led to the abandonment of the *single syndrome theory* of academic difficulties in favor of a *multiple syndrome paradigm* which hypothesized the existence of distinct subtypes (McKinney, 1984). Clinical-inferential subtyping studies based on WISC VIQ/PIQ discrepancies supported the idea that children with persistent academic difficulties represent a heterogeneous population, and demonstrated that specific patterns of WISC performance could be linked to differential patterns of neuropsychological, psychosocial, and academic functioning. Although empirical clustering techniques were not used to identify patterns of performance using the original WISC, the clinical-inferential subtyping research described above laid the foundation for this type of research using subsequent versions of this instrument.
1.7b The WISC-R (1974)

Goals of Revision

In 1974, the WISC was replaced by the WISC-R. This revision was carried out primarily to address criticisms related to inadequate norms and inappropriate or insufficient item content (Anastasi, 1988; Wechsler 1974b). According to Wechsler (1974b), the goals of the WISC revision were twofold: (1) to preserve as much of the 1949 WISC as possible considering its ubiquitous use around the world; and (2) to add items to make the scale more suitable for young children, and to modify or eliminate test items that were considered to be outdated, ambiguous, or "differentially unfair to particular groups of children" (Anastasi, 1998; Sattler, 1988; Wechsler, p. 10).

Organization and Content

The organization of the WISC-R was nearly identical to that of its predecessor. Thus, the Verbal/Performance IQ dichotomy was retained, as was the FSIQ (Kaufman, 1979; Wechsler, 1974b). The revised WISC comprised the same twelve subtests that constituted the 1949 version, and in keeping with the same procedures, IQ scores were calculated on the basis of the same 10 subtests. Consistent with the original WISC, Digit Span on the Verbal Scale, and Mazes on the Performance Scale were considered supplementary tests and, therefore, were not included in the derivation of IQ scores. The sequence in which tests were administered was changed on the WISC-R such that Verbal and Performance tests were administered in alternating fashion (Wechsler).

With respect to item content, 72% of original WISC items were retained on the WISC-R, either unchanged (64%) or with modification (8%; Sattler, 1988). All items from the original Digit Span, Mazes, and Coding subtests were included in the revised
version; 58 new items were added, with Comprehension, Picture Completion, and Picture Arrangement subtests receiving the greatest proportion of new items (47%, 42%, and 41%, respectively; Wechsler, 1974b).

**Construct Validity**

Addressing the notion of construct validity, all IQ scales and subtests correlated at least minimally, thereby supporting measurement of a global factor (FSIQ). Supporting the Verbal/Performance dichotomy, Verbal subtests generally correlated more highly with other Verbal subtests than with Performance subtests, and vice versa (Wechsler, 1974b).

The results of factor analytic research on the WISC-R mirrored that of the WISC. Thus, there was strong support for the global composite scale (i.e., FSIQ), as well as the lower-order Verbal and Performance Scales (Sattler, 1988). Additionally, in keeping with research on the WISC, many factor analytic studies of the WISC-R identified Verbal Comprehension, Perceptual Organization, and Freedom from Distractibility factors; with the latter factor again described as more elusive than the first two factors (Kaufman, 1975; Prifitera, Weiss, & Saklofske, 1998; Sattler, 2001; Wechsler, 1991b).

Nevertheless, in addition to the normative sample, this three factor solution was reported in clinical populations comprising children with emotional, psychiatric, and medical conditions, as well as children evidencing mental retardation, delinquency, and learning difficulties (Blaha & Vance, 1979; Carlson, Reynolds, & Gutkin, 1983; Dean, 1980; DeHorn & Klinge, 1978; Groff & Hubble, 1982; Hodges, 1982; Hubble & Groff, 1981; Johnston & Bolen, 1984; Naglieri, 1981; Richards, Fowler, Berent, & Boll, 1980; Schooler, Beebe, & Koepke, 1978). Reflecting the increasing popularity of the WISC-R
three factor solution, norms for the Verbal Comprehension, Perceptual Organization, and Freedom from Distractibility factors were published by independent researchers who considered WISC-R interpretation at this level to be the most clinically meaningful (Gutkin, 1979; Sattler, 1988; Wechsler, 1991b).

WISC versus WISC-R

In the years following publication of the revised WISC, numerous studies were conducted to systematically compare the WISC and WISC-R in terms of level of performance obtained and the underlying constructs measured. Empirically supporting the need for a revision, studies comparing level of performance on the WISC and WISC-R universally reported the Flynn Effect (i.e., average FSIQ score increases of approximately 3 points per decade; Flynn & Weiss, 2007). With respect to the constructs measured by each of the versions, Swerdlik and Schweitzer (1978) investigated the comparability of the WISC and WISC-R factor structures by administering both of these instruments in counterbalanced order to the same group of clinically referred children. As expected, three factors were extracted from each administration yielding strikingly high congruence coefficients: .98 for Verbal Comprehension and Perceptual Organization factors, and .77 for the third factor. From these results, the authors concluded that the WISC and its successor were comparable in terms of the intellectual constructs being measured. Similar results were obtained in studies comparing WISC and WISC-R performances in typically-developing children (Kaufman, 1979), gifted children (Larrabee & Holroyd, 1976), children with academic difficulties (Fisk & Rourke, 1987), and children with mental retardation (Kaufman & Van Hagen, 1977).
Summary of Changes Apparent in the WISC-R

During development of the WISC-R, relatively minor changes were made to the original instrument including improved norms, more appropriate item content, an increased age range, and improved administration and scoring materials. Factor analytic studies supported the VIQ-PIQ division, and also suggested that the WISC-R measured three latent constructs; namely, Verbal Comprehension, Perceptual Organization, and Freedom from Distractibility. Taken together, studies investigating the similarity of the WISC to its successor led to the conclusion that the WISC and the WISC-R were in fact comparable measures (Kaufman, 1975; Quarttrocci & Sherrets, 1980).

WISC-R Patterns of Performance

The use of empirical clustering techniques to delineate homogenous subgroups of children with persistent academic difficulties burgeoned with the publication of the WISC-R (Bender & Golden, 1990). These studies, although disparate in terms of inclusion criteria and methodology, consistently identified a number of reliable and valid WISC-R patterns in children exhibiting persistent academic difficulties (Ward, Ward, Glutting, & Hatt, 1999).

Snow et al. (1985) identified subgroups of children with academic difficulties on the basis of WISC-R factor scores. In this investigation, 11 WISC-R subtest scores (Mazes omitted) of 106 children diagnosed with learning disabilities on the basis of discrepancy criteria were factor analyzed, resulting in the expected three factor solution. These factor scores were then subjected to hierarchical cluster analysis which resulted in six homogeneous subgroups. Group 1 demonstrated relatively low scores across all three Indexes. Group 2 was characterized by low VCI, high POI, and average FDI scores.
Group 3 was characterized by a depressed FDI score in relation to the other two factors. Group 4 evidenced a low VCI score within the context of average scores on the other two factors. Group 5 exhibited a relatively flat profile with all factor scores falling within the average range. Finally, Group 6 was characterized by a low POI score relative to the other two factor scores. Unfortunately, the authors of this study did not assess the reliability of this empirically derived taxonomy, nor did they make a systematic attempt to differentiate these subgroups on the basis of academic achievement, despite having access to Wide Range Achievement Test (WRAT) data for all participants (Ward et al., 1999).

Vance, Wallbrown, & Blaha (1978) identified subgroups of children diagnosed with Specific Reading Disability. In this empirical-clustering investigation, all 12 WISC-R subtest scores of 104 children were subjected to Q-factor analysis. Given that these researchers endeavored to examine WISC-R profile shape, exclusively, participants with IQ scores below 85 were eliminated from the sample as were students with insufficient subtest dispersion. Five factors were extracted from the dataset; a solution which classified 75% of the sample. The following subgroups were identified: Group 1, labeled Distractibility, was characterized by low scores on three of the four subtests in the ACID tetrad (Arithmetic, Information, and Digit Span). Group 2 was labeled Perceptual Organization, and was associated with low scores on Block Design, Object Assembly, and Mazes. Group 3 exhibited a mean profile reflecting weak Digit Span and Comprehension scores; which was labeled Language Disability-Automatic. Group 4 was characterized by low scores on Similarities, Arithmetic, Digit Span, and Information, and was labeled Language Disability – Pervasive. The final cluster was referred to as
Behavioral Comprehension and Coding with low scores on Picture Arrangement, Comprehension, and Coding. No systematic attempt was made to externally validate this typology.

Holcomb et al., (1987) introduced a new method for reliability assessment in their cluster analytic study of school-identified children. In this investigation, WISC-R scores from 119 children diagnosed with learning disabilities were randomly split into two groups: a criterion sample and a calibration sample. Scaled scores from the first group were subjected to hierarchical cluster analysis which generated a six-cluster solution. Using a profile analytic method for classification based on squared Euclidean distance, cases in the second group were then allocated to each of the clusters obtained on the criterion sample. On the basis of this procedure, it was determined that 93% of children in the calibration sample could be classified into the six cluster types generated by the criterion group. These findings were considered evidence in support of the internal validity of the derived cluster solution. The external validity of this taxonomy was also supported, as performance on two achievement measures (the Peabody Individual Achievement Test [PIAT] and the Comprehensive Test of Basic Skills [CTBS]) significantly differentiated all six subtypes. The identified subtypes were described as follows: Subtype 1 exhibited a mean WISC-R profile characterized by low scores on those subtests composing the Verbal Scale in addition to poor performance on the Coding subtest. Relative to the other subgroups, these children obtained the lowest scores on both of the achievement tests administered, with particularly pronounced deficits in Reading Comprehension and Math Computation. The authors speculated that this subtype likely reflects a pervasive reading disability. Subtype 2 exhibited a pattern of
higher scores on Verbal subtests compared to Performance subtests, within the context of a generally low average profile. Achievement scores were generally consistent with IQ scores and showed a pattern of particularly depressed performance on the PIAT General Information subtest. The profile associated with Subtype 3 involved depressed scores on subtests included in the ACID tetrad, and a pattern of better performance on individually-administered achievement tests relative to group administered tests; a finding which led the authors to hypothesize that this may be a subgroup of children with impaired attentional abilities. Subtype 4 was characterized by below average performance on all WISC-R subtests with no apparent strengths or weaknesses, and commensurate academic achievement. Subtype 5 exhibited low average WISC-R scores on all subtests and commensurate academic achievement. Finally, Subtype 6 was characterized by generally high average performance on all WISC subtests and above average performance on the majority of achievement tests administered. The investigators hypothesized that the latter subtype may include children with socioemotional or behavioral disorders rather than specific learning problems and/or children who live in environments that provide little incentive for academic success. In their discussion, the authors suggested specific intervention strategies for each of the derived subgroups, and strongly encouraged continued empirical clustering research using the WISC-R.

Hale and Saxe (1983) conducted a Q-factor analysis of the 10 WISC-R core subtest scores in children referred for psychoeducational assessment. Four distinct subgroups were identified. Cross validation of this solution demonstrated that the four representative WISC-R profiles were similar across two samples; two of these profiles were considered to be clinically interpretable. The first interpretable profile (Profile 1)
reflected relative and normative weaknesses on subtests believed to measure attention/sequencing abilities (Arithmetic and Coding); the second interpretable profile (Profile 3) was characterized by weaknesses on subtests composing the PIQ (Picture Completion, Picture Arrangement, Block Design, and Coding). Regression analyses indicated that the closer a child resembled Profile 1 (low attention/sequencing), the more poorly he/she performed on the WRAT-R Reading subtest. Conversely, the closer a child resembled Profile 3 (better scores on Verbal than Performance subtests) the higher his/her Reading achievement. These findings were considered to be consistent with factor analytic research of the WISC-R, as well as with models of cognitive processing. Although profile shape in this study was found to add statistically significant amounts of variance in the regression model, it should be noted that the predictive improvement of shape over that offered by profile elevation alone was of questionable practical utility (Hale & Saxe).

Several studies included additional variables in their taxonomic studies of the WISC-R in order to more fully understand subgroups of children evidencing persistent academic difficulties. One such investigation was conducted by Bender and Golden (1990) who included a number of additional variables believed to be important for educational programming. In this study, six WISC-R subtests (Similarities; Vocabulary; Comprehension; Block Design; Object Assembly; and Coding), two academic variables (Woodcock-Johnson: Letter/Word Recognition and Word Attack subtests), and five behavioral indices (Walker Problem Behavior Identification Checklist: Acting Out, Immaturity, and Shy/Withdrawn Indexes; and Piers-Harris Children’s Self-Concept Scale: Intellectual Status and Popularity Indexes) were included in a dataset that was
subjected to hierarchical cluster analysis. Participants comprised 57 children, grades 3 through 9, identified with persistent academic difficulties by their school districts. The following five subtypes emerged: Subtype 1 was described as a group of children with no observable deficits in cognitive (WISC-R performance), academic, self concept, or behavior. Subtype 2, which was labeled Visual Deficits, exhibited impaired performance on the three WISC-R Performance subtests and demonstrated prominent acting out behaviors. Subtype 3 was characterized by impaired performance on WISC-R Verbal subtests in combination with very low scores on self perception of intellectual status and popularity. Subtype 4 evidenced a profile characterized by average WISC-R scores, diminished reading achievement, and very high self concept. Finally, Subtype 5 included children with above average WISC-R and reading achievement scores within the context of significant self concept and behavioral problems. The internal validity of this cognitive-academic-behavioral typology was supported by a single-sample replication procedure in which a subset of the original sample was reclustered, rendering the same cluster solution. External validity was assessed using measures of Passage Comprehension (Woodcock-Johnson), Happiness (Piers-Harris), and Adaptive Functioning (Weller-Strawser Adaptive Behavior Scale). All groups were found to differ from the other clusters on at least one variable, with the exception of the no-deficit and visual deficit subtypes. Based on the findings of this study, the authors concluded that these subtypes were generally consistent with those subtypes identified in other taxonomic studies based on the WISC-R in children with persistent academic difficulties, with the added component of socioemotional and behavioral status, both of which were hypothesized to be important in the generation of targeted intervention programs.
In a similar investigation, McKinney (1982, cited in McKinney, 1984) cluster analyzed 59 school-identified first and second grade students with persistent academic difficulties using scores from the WISC-R, PIAT, and Classroom Behavior Inventory (CBI). From this data, four subtypes emerged: Subtype 1 exhibited a mean WISC-R profile characterized by average verbal skills within the context of impaired visual and sequencing abilities. Behaviorally, these children showed marked deficiencies in independence and task-orientation; academic weaknesses were present in the areas of word recognition and math. Subtype 2 was characterized by below average scores on Information, Arithmetic, and Picture Arrangement WISC-R subtests and high average performance on the Object Assembly subtest. Pervasive and severe behavioral problems were noted in this subtype, and these children appeared to be the most severely impaired group with respect to global academic functioning. The WISC-R profile from Subtype 3 was unremarkable in the sense that all subtests were at least average. With respect to behavior, children in this group were found to be extremely extroverted, and were rated as less considerate and more hostile by their classroom teachers than children in other groups. This subtype was also characterized by mildly impaired performance on the PIAT. The WISC-R profile representative of Subtype 4 was similar to that of Subtype 1 (average verbal skills; deficient spatial/sequencing abilities); however, unlike Subtype 1, behavioral concerns did not characterize this group of children and academic performance was found to be more impaired. Supporting the external validity of this typology, significant differences were found among groups on direct observation measures taken in the classroom and on the Pupil Rating Scale.
Summary of WISC-R Patterns of Performance. Despite considerable variability across WISC-R taxonomic studies in terms of methodology, definitional criteria, level of WISC-R interpretation, variables included in the analysis, description of WISC-R subtest groupings, and internal/external validation procedures, some consistencies emerged, which provided valuable information about the various intellectual profiles exhibited by children with persistent academic difficulties (Ward et al., 1999). First, from this research, it is clear that the WISC-R represented a useful instrument with respect to identifying subgroups of children with academic limitations (McKinney, 1984). Second, these studies generally indicated that either five or six WISC-R profiles best represented this population. Finally, common to the majority of these empirically derived typologies were patterns reflecting deficits in verbal, visual-spatial, or attention/sequencing (including the ACID pattern) domains, global cognitive deficits, or average ability with no observable deficits (Snow et al., 1985; Ward et al.). The reliability of these empirically-derived WISC-R subgroups was adequately demonstrated, and in some cases these subgroups were found to be externally valid in that they differed on measures not included in cluster formation. A summary of the WISC-R empirical clustering literature is presented in Appendix A.

1.7c The WISC-III (1991)

In 1991, the second revision of the WISC was released. Although notable changes were made to the instrument, the WISC-III was not considered a vast departure from its predecessor, reflecting the developers’ intentions to “maintain the basic structure and content of the WISC-R [to preserve] longitudinal consistency”, (Saklofske et al., 2003; Wechsler, 1991b, p.11).
Goals of Revision

According to the WISC-III manual, the changes made to the WISC-R were guided by three rather circumscribed goals (Wechsler, 1991b): First, primarily to offset the Flynn effect (Flynn, 1984; Flynn & Weiss, 2007), new norms were developed using a contemporary standardization sample. Second, test items and materials were modified to make the instrument more appealing to examinees, to improve measurement at both ends of the age continuum, and to reduce bias. Third, one supplementary subtest, Symbol Search, which involves rapidly determining the presence or absence of visual targets from a search group, was added to the instrument in an attempt to enhance the psychometric integrity of the Freedom from Distractibility factor.

Organization and Content

In keeping with the WISC tradition, the WISC-III retained the Verbal, Performance, and Full Scale IQ structure. Departing from previous versions, however, the WISC-III provided norms for four factor-based Index scores: VCI (calculated on the bases of Information, Similarities, Vocabulary, and Comprehension scaled scores); POI (calculated on the basis of Picture Completion, Picture Arrangement, Block Design, and Object Assembly scaled scores); FDI (calculated on the basis of Arithmetic and Digit Span scaled scores); and Processing Speed (PSI; calculated on the basis of Coding and Symbol Search scaled scores).

IQ scores for the WISC-III were derived on the basis of the same 10 subtests used in previous WISC versions, with supplementary subtests provided for each scale: Digit Span for the Verbal Scale and Mazes and the new Symbol Search test for the Performance Scale. The administration sequence of the WISC-III was similar to the
WISC-R in that Verbal and Performance subtests were administered in alternating fashion. Unlike its predecessor, however, WISC-III administration began with a Performance subtest (Picture Completion) which was considered to be a non-threatening task that would be helpful with respect to building rapport. Thus, with the exception of the addition of one supplementary subtest, the basic structure and organization of the WISC-III was consistent with that of previous WISC versions (Sattler, 2001; Wechsler, 1991b).

To address the mounting WISC-R literature indicating the presence of the Freedom from Distractibility factor in both normal and special populations, the WISC-III developers paid particular attention to this factor during the revision process. According to the WISC-III manual, literature reviews of the WISC-R suggested that the Freedom from Distractibility factor reflected a combination of attention, sequencing, short-term memory, and to a lesser degree numerical ability (Wechsler, 1991b). In an attempt to improve the clinical utility of this factor by enhancing the memory and attentional demands, the Symbol Search subtest was developed on the basis of memory scanning and controlled-attention research (Wechsler). Although the goal of adding the Symbol Search subtest to the WISC-III was to clarify the abilities measured by the third factor, the result was the splitting of the third factor into two separate factors; one retaining the original Freedom from Distractibility label (Arithmetic and Digit Span), and the other bearing the label Processing Speed (Coding and Symbol Search).

With respect to item content, a number of new items were added to the original 12 subtests (41 on the Verbal Scale and 64 on the Performance scale) primarily to extend accurate measurement at both ends of the age continuum. However, recognizing the
importance of maintaining continuity between WISC editions for clinical and research purposes, more than 73% of WISC-R items were retained in the updated version either in original or slightly altered form. Other changes evident in this revision included the addition of color to previously black-and-white test stimuli, deletion of items considered to be psychometrically unsuitable or outdated, and the addition of scoring procedures emphasizing speed of performance (Mahone et al., 2003; Sattler, 2001; Wechsler, 1991b).

Construct Validity

In keeping with the WISC and WISC-R, subtest, Index, and IQ correlations provided evidence for measurement of a global factor. Further, with the exception of Picture Completion at ages 14 – 16, “Verbal subtests generally correlate[d] more highly with each other than with Performance subtests…and Performance subtests generally correlate[d] more highly with each other than with Verbal subtests” (Wechsler, 1991b, pg. 186).

Factor analytic studies of the WISC-III provided strong empirical support for interpretation of the Verbal, Performance, and Full Scale composite scores, as well as the Verbal Comprehension and Perceptual Organization factors (Hynd et al., 1998). Although the new Symbol Search subtest was added to the WISC-III to strengthen the Freedom from Distractibility factor, as suggested previously, its inclusion resulted in the unexpected splitting of the FDI into two separate factors. Thus, based on exploratory and confirmatory factor analyses of the standardization data, Wechsler (1991b) concluded that a four-factor model, consisting of two major factors: Verbal Comprehension (Information, Similarities, Vocabulary and Comprehension) and Perceptual Organization
(Picture Completion, Block Design, Object Assembly, and Picture Arrangement), and
two smaller supplementary factors: Freedom from Distractibility (Arithmetic and Digit
Span) and Processing Speed (Coding and Symbol Search), best represented the WISC-III
structure (Cooper, 1995). This factor solution, which was found to be reliable over the
course of development in both normal and clinical populations (Watkins & Canivez,
2001), was supported by independent factor analytic studies conducted on the WISC-III
standardization sample (e.g., Kamphaus, Benson, Hutchinson, & Platt, 1994; Roid &
Worrall, 1997), the WISC-III/Wechsler Individual Achievement Test (WIAT) linking
sample (Roid, Prifitera, & Weiss, 1993), and various clinical groups including students
with persistent academic difficulties and those with attention-deficit disorders

Despite considerable support for the four factor solution espoused by the WISC-III
developers, numerous researchers challenged the adequacy of this model, particularly
with respect to the FDI and PSI (Burton et al., 2001; Carroll, 1993b; Daly & Nagel, 1996;
Sattler, 2001; Watkins & Kush, 2002). Illustrating this point, Little (1992) concluded
that “Processing Speed should be interpreted cautiously and Freedom from Distractibility
should be ignored” (pg. 153). Nevertheless, the majority of factor analytic research
suggested that four primary abilities were measured by the WISC-III (Roid et al., 1993).
Thus, preeminent assessment texts devoted considerable attention to interpretation of the
VCI, POI, FDI, and PSI (Sattler, 2001; Kaufman, 1994; Prifitera & Saklofske, 1998), and
the interpretive model outlined in the manual was generally accepted in clinical practice
**WISC-R versus WISC-III**

The majority of items from the WISC-R were retained in the WISC-III (Wechsler, 1991b). Thus, it is not surprising that these two WISC versions were considered comparable by prominent clinicians and researchers (Sattler, 2001). Demonstrating the similarity of these two instruments, the WISC-III manual reported findings from two concurrent validity studies which yielded correlation coefficients of the WISC-R and WISC-III of .86 and .90 for the Verbal Scale, .73 and .81 for the for the Performance scale, and .86 and .89 for the Full Scale. As would be expected, reflecting updated norms, the vast majority of studies conducted on both normative and clinical samples reported decreased IQ scores on the WISC-III relative to the WISC-R (Wechsler, 1991b). Sattler (2001) reviewed 33 studies that were conducted to compare level of performance on these two instruments, and concluded that the WISC-III yielded IQ scores approximately 6 points lower than the WISC-R, with median changes of -5.4, -6.06, and -5.97 points for the Verbal, Performance, and Full Scales, respectively. According to the WISC-III manual, WISC-R and WISC-III PIQ scores differed more than VIQ scores, presumably reflecting an increased emphasis on performance speed on the WISC-III subtests composing the Performance Scale (Mahone et al., 2003; Wechsler, 1991b).

**Summary of Changes Apparent in the WISC-III**

The basic structure and content of the WISC-III mirrored that of the WISC-R. The same three IQ scores were calculated on the basis of the same 10 subtests, and 73% of the items on the WISC-R were included on the WISC-III. Relatively minor changes in the content of the WISC-III included additions, omissions, and modifications to various items in an attempt to decrease bias, extend subtest floors and ceilings, and increase
examinee interest. Norms for the WISC-III were updated and improved, and the psychometric properties of this revised instrument were considered to be comparable to previous versions. The most dramatic change made to the WISC-III involved the addition of a supplementary subtest, Symbol Search, which changed the landscape of this instrument by adding a Processing Speed factor. Unlike its predecessors, the WISC-III manual provided a means for calculating four Index scores (VCI, POI, FDI, PSI) in addition to the traditional three IQ scores. Demonstrating the value placed by the authors on research and clinical continuity, the WISC-III “was not a vast departure from the WISC and WISC-R” (Saklofske et al., 2003 p. 8).

**WISC-III Patterns of Performance**

Interest in delineating subgroups of children based on WISC performance continued with the publication of the WISC-III. Empirical clustering studies using the WISC-III alone or in combination with measures of achievement (e.g., WIAT) were conducted to identify reliable and valid patterns of intellectual functioning in children from the general population as well as children exhibiting persistent academic weaknesses. The following discussion will begin with a review of the literature exploring empirically derived WISC-III typologies in the normative sample. Generally, the goal of this research was to identify core profiles (i.e., typically occurring profiles) that could be used to test the uniqueness of individual profiles believed to be clinically relevant (Glutting, McDermott, Prifitera, & McGrath, 1994). The cluster analytic literature on the WISC-III in clinical populations, focusing primarily on those studies involving children with academic difficulties, will then be reviewed.
To identify patterns of processing in the WISC-III normative sample, Donders (1996) conducted a multistage cluster analysis using the four Index scores that compose the FSIQ. From this procedure, five distinct profiles were identified, three of which were differentiated entirely by level (below average, average, and above average). The two remaining profiles were characterized by Index score patterns, with the defining feature of each being performance on the PSI. Specifically, while both clusters obtained average scores on VC, PO, and FD Indexes, one group demonstrated a relative weakness in Processing Speed, whereas the other exhibited a relative strength in this area. The reliability of this normative taxonomy was supported by the finding that the same cluster solution emerged on the basis of two distinct clustering algorithms (Ward’s Minimum Variance and Complete Linkage methods). Parental education was found to vary with WISC-III level of performance, thereby supporting the external validity of this cluster solution. That the derived subgroups did not differ significantly in age was considered evidence that this solution was applicable across the entire age span of the WISC-III standardization sample, and suggested that the observed PSI variations could not be attributed to reduced reaction times resulting from increased age. The author recommended that future research examine empirically-derived WISC-III profiles in specific diagnostic groups.

Konold et al. (1999) also identified reliable and valid WISC-III performance patterns in children composing the standardization sample. In contrast to the Donders (1996) study, which employed Index scores, these researchers based their taxonomy on the ten core subtests, and interpreted their clusters according to mean performance on the three IQ composites. Using multistage cluster analysis, eight distinct subgroups emerged.
Four of the representative profiles were characterized solely by level; the remaining four profiles were defined on the basis of both level and pattern. The clusters were labeled as follows: (1) high ability; (2) above average ability; (3) above average ability and VIQ > PIQ; (4) average ability and PIQ > VIQ; (5) average ability and VIQ > PIQ; (6) below average ability and PIQ > VIQ; (7) below average ability; (8) low ability. The reliability of this typology was supported by adequate within-cluster homogeneity and between-cluster dissimilarity, as well as cluster replication across the 11 age blocks. External validity was evaluated through the examination of demographic characteristics (i.e., age, sex, ethnicity, education placement, region, and parental education); a process which included contrasting actual demographic prevalence percentages for each cluster with expected prevalence rates. Significant differences between actual and expected prevalence rates on at least one demographic variable were observed in all eight clusters. The authors concluded by illustrating a multivariate procedure for evaluating the relative uniqueness of individual WISC-III profiles (i.e., subtest variation not consistent with one of the eight core profile types) and suggested that through the use of this procedure, “psychologists disposed to interpreting subtest score variation could begin to investigate patterns of dips and rises that may form the basis for hypothesis generation” (p. 44).

Using the procedures outlined by Konold et al. (1999), Borsuk et al. (2006) sorted a heterogeneous clinical sample into the eight core WISC-III subtypes listed above. Although the vast majority (90%) of participants matched one of the core profiles, only two of the nine groups investigated in this study (8 groups reflecting core profile types and one group exhibiting unusual profiles) were found to be stable over time. Specifically, test-retest results for children matching Profile 6 (below average ability and
PIQ>VIQ) and Profile 8 (low ability) remained consistent over a three year interval. Thus, according to the authors, only these two profiles were stable enough to warrant further clinical investigation.

To examine the effect of including supplemental subtests in the derivation of a normative taxonomy, Glutting et al. (1997) cluster analyzed the WISC-III standardization sample using all subtests with the exception of Mazes. The inclusion of these supplemental subtests resulted in the derivation of profiles characterized by variations associated with the FDI and PSI in addition to the more commonly reported variations associated with the FSIQ, VIQ, and PIQ. Specifically, on the basis of multistage cluster analyses, nine subtypes were identified: (1) High ability and depressed PSI; (2) Above-average ability; VIQ>PIQ and depressed PSI; (3) Above average ability; PIQ>VIQ and elevated PSI; (4) Average ability; (5) Average ability and elevated FDI; (6) Slightly below average ability; (7) Slightly below average ability; PIQ > VIQ and elevated PSI; (8) Below average ability and VIQ > PIQ; and (9) Low ability with elevated FDI and PSI. This nine-cluster typology was found to have adequate internal validity, and was deemed externally valid on the basis of significant differences between expected versus actual prevalence rates with respect to cluster demography. As with previous investigations conducted by this research group, a method for determining the exceptionality of individual WISC-III profiles was illustrated.

In the interest of “developing a taxonomy representative of the abilities and achievements typically evaluated during psychoeducational assessments”, Glutting et al. (1994) used cluster analysis to identify the most common ability and achievement profiles for children in the WISC-III – WIAT linking sample (p. 623). Using WISC-III
factor Index scores and WIAT composite scores (Reading, Mathematics, Language, and Writing) as clustering variables, six reliable and valid subgroups were identified. Mean profiles for the six extracted clusters were differentiated primarily on the basis of level, and were labeled according to WISC-III IQ scores and WIAT achievement levels. Supporting the internal and external validity of the typology, respectively, the resultant cluster solution met a number of predetermined reliability criteria, and the subgroups differed significantly on various demographic characteristics. The subtypes were labeled as follows: (1) High ability and VIQ > PIQ; (2) Above average ability with slightly above average achievement and PIQ>VIQ; (3) Average ability with underachievement in writing; (4) Average ability with over achievement in reading, mathematics, language, and writing; (5) Below average ability with below average achievement; and (6) Low ability with underachievement in reading, mathematics, writing, and PIQ>VIQ. Given that over half of the derived core profiles were characterized by significant ability-achievement discrepancies, the authors encouraged practitioners to exercise caution in diagnosing learning disorders on the basis of such a discrepancy, as it is occurs frequently in the general population. It was further recommended that future research address the extent to which multivariate typologies are useful with respect to predicting future outcomes and prescribing effective interventions for children with academic difficulties.

Recognizing the relative paucity of taxonomic research examining WISC-III patterns based on subtest variation in clinical populations, Saunders, Casey, & Jones (2001) applied cluster analysis to the 12 WISC-III subtest scores (Mazes omitted) of 343 children referred for neuropsychological evaluation. Based on the nature of the referrals, it was presumed by the investigators that this sample included a significant number of
children with persistent academic difficulties. Using cluster analytic methodology involving a combination of hierarchical and iterative partitioning techniques (described in detail later), six clusters emerged, with mean profile patterns reflecting the following: (1) broad based processing deficiencies (all subtest scores below average); (2) deficient language abilities (below average scores on all subtests contributing to the VIQ, and average scores on subtests included in the PIQ); (3) deficient nonverbal abilities (low average to below average scores on subtests contributing to the PIQ and roughly average scores on subtests included in the VIQ); (4) the ACID pattern; (5) deficient working memory (lowest scores on Digit Span and Coding subtests); and (6) deficits in tasks involving visual sequencing and language abilities (lowest scores on subtests contributing to the VCI, as well as low scores on Coding, Picture Arrangement, and Symbol Search subtests). The reliability of this typology was supported by the demonstration of significant similarity between initial and final cluster solutions. The fact that significant between-subgroup differences were found on the basis of WIAT Basic Reading, Spelling, Mathematics Reasoning, and Numerical Operations supported the external validity of these six subgroups. One limitation of this study, however, involves the researchers’ failure to report the specific nature of the WIAT patterns associated with each subgroup. Nevertheless, the authors concluded that clinically meaningful patterns of WISC-III strengths and weaknesses had been identified in a heterogeneous clinical sample, and that these empirically-derived patterns were generally consistent with factor analytic findings and models of cognitive processing. Further, the authors observed that while the majority of the clusters found in this study were characterized by well recognized WISC-III patterns (i.e., language based deficits, nonverbal deficits, ACID pattern, broad based
deficits), others were less frequently discussed in the literature (i.e., working memory deficits, language deficits in combination with sequencing weaknesses) indicating the need for continued research in this area.

In a follow up study, Waxman, Casey, and Fuerst (2003) tested the reliability of the cluster solution reported by Saunders et al. (2001) by applying Q-factor analysis to the same sample. Six subgroups were identified, four of which were markedly similar to those identified by Saunders et al. Specifically, like those groups found in the earlier study, this investigation yielded subgroups characterized by (1) verbal processing deficits; (2) the SCAD profile (lowest scores on subtests of the FDI and PSI; generally considered to be similar to the ACID profile with Symbol Search replacing Information in the list of lowest scores; Kaufman, 1994); (3) visual sequencing and language deficits; and (4) nonverbal processing deficits. Subtypes five and six, which did not resemble clusters found by Saunders et al., were characterized by general processing deficits with intact processing speed (below average scores on all subtests except those included in the PSI), and higher order processing deficits (lowest scores on Arithmetic, Digit Span, Picture Completion, Block Design, and Object Assembly; which, according to the authors, all require some degree of higher order reasoning with the exception of Digit Span).

In addition to subgroups derived solely on the basis of WISC-III patterns of performance, two empirical clustering studies used WISC-III and WIAT performances combined. Ward et al. (1999) cluster analyzed a group of 201 students with academic difficulties using the four WISC-III Index scores in addition to the four WIAT composite scores (Reading, Mathematics, Language, and Writing). To determine the extent to
which the profiles identified in this investigation matched those observed in the general population, the profiles of the participants in this study were compared to the six core profile types described by Glutting et al. (1994). Consistent with much of the empirical clustering literature, a five-group solution was selected, which was considered to be similar in number and pattern to those identified in previous research. Two clusters demonstrated low average ability with commensurate achievement (i.e., no significant differences) in all academic areas. One of these clusters was defined by a relative strength on the PSI. Another group exhibited below average ability and commensurate achievement. The remaining two clusters were characterized by significant ability-achievement discrepancies. Specifically, one profile exhibited low VCI and FDI scores relative to the other WISC-III Indexes and significant underachievement on WIAT Reading and Writing composites. The final cluster was characterized by average performance on all WISC-III Indexes and significant underachievement in writing. Of the 201 cases clustered, approximately 70% exhibited WISC-III/WIAT profiles resembling one of the six core profiles identified in the normative sample. Thus, it was concluded that many children with persistent academic difficulties exhibit ability and achievement profiles similar to those observed in the general population. According to these authors, the findings from this investigation do not imply that the presence of a typically-occurring profile precludes the need for intervention. It does, however, speak to the need for future research in this area, particularly with respect to the identification of other variables (e.g., neuropsychological functioning, behavior, adaptive functioning, personality, neurological features, and contextual factors) that may interact with specific
WISC-III/WIAT patterns to contribute to the manifestation of significant academic problems.

Focusing on subtest patterns rather than discrepancy analyses, Waxman and Casey (2006) used hierarchical and iterative partitioning cluster analyses to group WISC-III (all subtests excluding Mazes) and WIAT (Reading, Math Reasoning, Spelling, and Numerical Operations) subtests of 182 children referred for neuropsychological evaluation, the majority of whom were diagnosed with various learning difficulties. Consistent with other studies involving this population, five distinct subgroups were identified. Based on the most salient features of each profile, the subgroups were labeled: (1) Low Ability (predominantly below average scores across WISC-III and WIAT subtests); (2) Low Ability with Average Processing Speed (scores ranging from below average to low average on all WISC-III and WIAT measures with the exception of Coding and Symbol Search which were in the average range); (3) Low Visual Spatial/Processing Speed (broadly average abilities across WIAT and WISC-III subtests, with scores on nonverbal subtests consistently lower than scores on verbal subtests); (4) ACID pattern (depressed scores on Arithmetic, Digit Span, Coding, and Information, with low scores on all WIAT subtests, most notably Reading and Spelling); and (5) Verbal Processing Deficits (depressed WISC-III Verbal and WIAT achievement scores, particularly Reading and Spelling, within the context of average performance on nonverbal subtests). Demonstrating the internal validity of this empirically derived typology, all five clusters were well replicated across three hierarchical clustering methods, and all five clusters were similar to subgroups described in previous taxonomic research. Supporting the external validity of this solution, the five clusters exhibited
distinct patterns of performance across five neuropsychological domains. Specifically, the Low Ability group was found to have significantly depressed scores in all neuropsychological areas. The Low Ability with Average Processing Speed group evidenced below average language skills, low average attention and problem solving skills, and average motor skills. Unexpectedly, the Low Visual Spatial/Processing Speed group exhibited generally average to low average performance across all neuropsychological domains, with motor and language domains representing the areas of greatest difficulty. Although the ACID group was also characterized by average to low average neuropsychological test performance, this group exhibited the most difficulty on neuropsychological measures of language and attention. Finally, the Verbal Processing Deficits group was characterized by well below average performance on language tasks within the context of average performance on tasks related to all other neuropsychological domains. According to the authors, the findings from this study support the notion that children with learning difficulties represent a heterogeneous population with a variety of intellectual processing profiles. Further, that these cluster profiles overlapped with clinically meaningful neuropsychological ability constructs led the authors to conclude that these patterns may be of value in the development of individualized treatment recommendations for children with significant academic difficulties.

Summary of WISC-III Patterns of Performance. Interest in identifying WISC patterns of performance in children with academic difficulties continued with the publication of the WISC-III. Although research in this area was replete with methodological and interpretive variations, reliable and valid WISC-III typologies were
identified in the literature, supporting the utility of this type of research in understanding the processing patterns of both typically achieving children as well as children experiencing persistent academic difficulties. Reflecting the relatively modest changes made to this instrument, the majority of taxonomic studies based on WISC-III patterns of performance identified profiles similar to those reported in studies using the WISC-R (Daly & Nagle, 1996). With the inclusion of the Symbol Search subtest, however, some profiles were additionally characterized by variations in Processing Speed.

Empirical clustering investigations conducted on the WISC-III normative and WISC-III-WIAT linking sample generated ability profiles characterized primarily by level of performance with some variability noted in VIQ, PIQ, FDI, and PSI patterns. Conversely, taxonomic research conducted on children with persistent learning difficulties generally revealed WISC-III profiles characterized more by pattern than by elevation. In keeping with research on the WISC-R, empirical clustering studies investigating WISC-III performance in children with persistent academic difficulties consistently identified subgroups reflecting (1) verbal processing deficits; (2) nonverbal/visual spatial weaknesses; (3) attention/sequencing/working memory difficulties with and without processing speed weaknesses; (4) global processing deficits (with and without processing speed strengths); and (5) average performance with no observable cognitive weaknesses. These subgroups have been well replicated, and have been found to differ in terms of demographic characteristics, academic skill patterns, and neuropsychological test performance. Thus, the empirically-derived profiles identified using the WISC-III appear to be reliable and valid, are remarkably similar to those identified using the WISC-R, and are generally consistent with factor analytic findings,
cognitive processing theories, and neuropsychological constructs (Waxman et al., 2003). As such, these groups hold promise for the identification of effective intervention strategies for children with academic difficulties (Waxman et al.). Whereas the development and empirical testing of such interventions was the next logical step in empirical clustering research using the WISC-III, the WISC underwent another revision before this research was conducted. A summary of the WISC-III empirical clustering literature is depicted in Appendix B.

1.7d The WISC-IV (2003)

In 2003, the WISC-IV supplanted the WISC-III. Unlike all previous WISC editions, which were specifically designed to maintain consistency across revisions for the sake of clinical and research continuity, the WISC-IV represents a radical departure from its predecessors (Flanagan & Kaufman, 2004; Hale & Fiorello, 2004; Keith et al., 2006; Prifitera, Saklofske, Weiss, & Rolhus, 2005; Wechsler, 2003c). Based on consultation with experts in numerous fields of psychology, in addition to an exhaustive literature review in the areas of cognitive development, cognitive neuroscience, intelligence theory, and intellectual assessment, many changes were made to the overall structure and content of this instrument. These modifications have altered the clinical data obtained from intellectual assessment using the WISC, calling into question the applicability of the research base accrued on previous versions of this instrument (Baron, 2005; Kaufman et al., 2006). Following is an overview of the WISC-IV, and a review of the relatively modest research that has been conducted exploring patterns of performance using this revised scale in children exhibiting persistent academic difficulties.
Goals of Revision

According to the WISC-IV manual (2003c), five primary goals steered the latest revision of the WISC. First, responding to longstanding criticisms regarding the scale’s insufficient theoretical foundation and the abundance of literature espousing the role of fluid reasoning, working memory, and processing speed in the conceptualization of intelligence, significant changes were made to the instrument including the addition of several new subtests (Baron, 2005; Carroll, 1993a; 1997; Cattell & Horn, 1978; PsychCorp, 2004; Wechsler, 2003c). Thus, compared to its predecessors, the WISC-IV reflects a more modern approach to intellectual assessment by taking into account contemporary psychometric theories and neurocognitive models of information processing (Flanagan & Kaufman, 2009; Kain, 2006; Williams, Weiss, & Rolfhus, 2003b).

Second, toward the goal of enhancing the scale’s clinical utility, 16 special group studies were conducted by the WISC-IV developers, the results of which are presented in the WISC-IV manual (Wechsler, 2003c). Additionally, the WISC-IV was statistically linked to the updated version of the WIAT (WIAT-II; The Psychological Corporation, 2001), methods for deriving and comparing numerous process scores were added to the manual, and improved base rate data was provided to assist clinicians discriminate between statistically significant and clinically meaningful subtest and composite score differences (Baron, 2005; Wechsler).

Third, in an attempt to increase the developmental appropriateness of the new WISC, subtest instructions were simplified and/or reworded, the artwork was updated, and test items deemed outdated were revised, removed, or replaced. Teaching items,
queries, and prompts were improved to promote better understanding and retention of task instructions, and the scoring criteria for verbal responses were modified to better accommodate the limited vocabulary of younger test-takers (Wechsler, 2003c; PsychCorp, 2004).

The fourth goal, which involved improving the psychometric properties of the scale, was accomplished by standardizing the WISC-IV on a contemporary normative sample, conducting extensive reliability and validity studies, and extending a number of subtest floors and ceilings to ensure adequate coverage. Additionally, item bias was empirically addressed using contemporary statistical methodology, and items deemed problematic were omitted from the battery (Wechsler, 2003c).

The fifth and final goal of the WISC-IV revision involved improving user-friendliness. To reduce testing time, the developers reduced the number of subtests required to obtain all composite scores (Flanagan & Kaufman, 2004). Additionally, individual subtest administration times were reduced by 10 to 15 minutes. Two manuals rather than one accompany the WISC-IV. The Administration and Scoring Manual (Wechsler, 2003b) provides information necessary for administering, scoring, and recording the results of the WISC-IV; the Technical and Interpretive Manual provides information regarding the scale’s theoretical foundation, organization, psychometric properties, and interpretive procedures (Wechsler, 2003c). Finally, the Test Record Form was revised to facilitate the process of test administration and the provision of assessment results (Wechsler).
With respect to content, only 56% of items on the WISC-III were retained on the WISC-IV, and notable modifications were made to pre-existing subtests in terms of start points, querying procedures, scoring guidelines, and number of items. Speed of performance was deemphasized on the WISC-IV for all subtests not contributing to the PSI (Sattler & Dumont, 2004; Wechsler, 2003c). Five subtests were added to the battery (Picture Concepts, Letter Number Sequencing, Matrix Reasoning, Word Reasoning, and Cancellation), three subtests were omitted from the battery (Picture Arrangement, Object Assembly, and Mazes), three subtests were changed from core to supplemental subtests (Information, Arithmetic, Picture Completion), and two subtests were changed from supplemental to core subtests (Digit Span and Symbol Search; Wechsler, 2003c). Only 5 of the 15 subtests composing the WISC-IV were retained from the WISC-III in similar form (Similarities, Vocabulary, Comprehension, Coding, and Block Design).

In terms of organization, subtests on the WISC-IV continue to be divided into core and supplemental subtests, with core subtests used in the derivation of all composite scores and supplemental subtests used to provide additional clinical information or to act as substitutes for invalidated core subtests. The most salient structural differences between the new WISC and all previous versions of the instrument relate to the composition of the Full Scale and factor Indexes (Sattler & Dumont, 2004). Following is a review of the changes made to each of these scales.

The Full Scale. In keeping with the Wechsler tradition, the Full Scale was retained in the latest WISC revision. However, "the WISC-IV FS-IQ has changed so dramatically in content and concept that it barely resembles the FS-IQ of previous WISCs or of any
other Wechsler Scale...” (Kaufman et al., 2006, p. 280). Specifically, in contrast to earlier versions of the instrument, the WISC-IV Full Scale is divided into four composite scales rather than the traditional Verbal and Performance Intelligence subscales. Thus, representing one of the most radical changes made during the development of the WISC-IV, the longstanding Verbal – Performance dichotomy was abandoned in favor of the adoption of four individual Indexes, with each Index presumably measuring a relatively circumscribed cognitive domain (Baron, 2005; Burns & O’Leary, 2004).

**The Verbal Comprehension Index.** According to the WISC-IV developers, the new VCI is a “more refined, purer measure of verbal reasoning and conceptualization than the WISC-III VIQ” (Wechsler, 2003c, p. 103). Departing from the WISC-III VIQ, but in keeping with the WISC-III VCI, the Arithmetic subtest is not included on the WISC-IV VCI, as factor analytic research suggests that this subtest is a better reflection of working memory processes than of general verbal abilities. To reduce the influence of acquired knowledge, Information, a subtest that has contributed to the Verbal scales (both VIQ and VCI) in all previous WISC versions, was changed from a core to a supplemental subtest on the WISC-IV. Additionally, Word Reasoning, a supplemental subtest involving the identification of words and concepts based on a series of increasingly specific oral clues, was added to the scale to bolster measurement of fluid verbal abilities (i.e., verbal reasoning; Wechsler). Thus, the WISC-IV VCI comprises three core subtests (Similarities, Vocabulary, and Comprehension) and two supplemental subtests (Information and Word Reasoning). In contrast to its predecessors, this Index is intended to measure higher-order verbal reasoning skills, and, according to the test developers, is
less influenced by school achievement and crystallized knowledge (Kaufman, et al., 2006).

The Perceptual Reasoning Index. Of the four composite scales, the Perceptual Reasoning Index (PRI) represents the greatest departure from its predecessors (i.e., the PIQ and POI), with only one core subtest retained from previous WISC versions (Mayes & Calhoun, 2006; Sattler & Dumont, 2004). Consistent with the updated nomenclature, the WISC-IV PRI was specifically designed to emphasize nonverbal fluid reasoning and the ability to process novel, less crystallized, information (Wechsler, 2003c). Toward the goal of purifying the constructs measured, speed of processing was deliberately de-emphasized on this scale compared to previous WISC editions (Prifitera, Saklofske, Weiss, & Rolfhus, 2005; Wechsler).

With respect to the structural changes apparent in the PRI, Object Assembly, Picture Arrangement, and Mazes were eliminated from the Index on the basis of questionable psychometric properties, emphasis on speed, and lengthy administration times (Prifitera, Saklofske, Weiss, & Rolfhus 2005). Picture Completion was changed from a core to a supplemental subtest, and Coding was moved to the Processing Speed Index (which originally occurred with the development of the WISC-III POI). Two new subtests were incorporated into the scale to bolster representation of fluid reasoning skills: Picture Concepts (a subtest that involves the identification of pictures with a common characteristic) and Matrix Reasoning (a subtest that involves solving incomplete matrixes). The former subtest, which was adopted from the Wechsler Preschool and Primary Scale-Third Edition (WPPSI-III; Wechsler, 2002b), is described as a measure of categorical reasoning skill (Wechsler, 2003c). The latter subtest was adopted from both
the WPPSI-III and the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; Wechsler, 1997), and is considered to be a measure of visual spatial processing and abstract reasoning abilities (Wechsler). While the Matrix Reasoning subtest bears a close resemblance to longstanding matrix analogy tasks from other intelligence tests, and possesses a rich empirical foundation, the Picture Concepts subtest is a rather novel task that was only recently developed by the Psychological Corporation (Flanagan & Kaufman, 2004; Wechsler). As Kain (2006) points out, "little research exists regarding the cognitive processes and strategies underlying performance on [this subtest]" (p. 19).

In short, the PRI comprises three core subtests (Block Design, Picture Concepts, and Matrix Reasoning) and one supplemental subtest (Picture Completion). This composite scale represents a substantive theoretical and structural departure from its predecessors, with measurement of fluid reasoning abilities now representing the primary area of emphasis (Prifitera, Saklofske, Weiss, & Rolfhus, 2005).

*The Working Memory Index.* Reflecting an increased understanding of the constructs purportedly measured by the FDI, this scale was renamed the Working Memory Index (WMI) in the latest WISC revision (Keith et al., 2006; Prifitera, Saklofske, Weiss, & Rolfhus, 2005). According to Wechsler (2003c), working memory represents a critical component of higher-order cognitive processing, and requires the ability to attend, concentrate, remember, exert mental control, and reason. Thus, the WMI presumably comprises subtests that collectively measure these abilities.

In addition to a change in nomenclature, this scale underwent several structural modifications in the latest WISC revision. While the WISC-III FDI was composed of Arithmetic (a core subtest on the WISC-III) and Digit Span (a supplemental subtest on
the WISC-III, the WMI is composed of Digit Span (now a core subtest) and the new Letter-Number Sequencing subtest. This latter subtest, which was adapted from the WAIS-III, requires the examinee to recall letters and numbers in ascending order, and is described as a measure of "sequencing, mental manipulation, attention, short-term auditory memory, visuospatial imaging, and processing speed" (Wechsler, 2003c, p. 17).

The Arithmetic subtest was changed from a core to a supplemental subtest, and underwent considerable modification. The mathematical knowledge required to complete the task was substantially reduced, and the time bonuses associated with this subtest were eliminated (Wechsler; Yeates & Donders, 2005).

The Processing Speed Index. Processing Speed is an important cognitive process with respect to reading performance, fluid reasoning, and the proficient use of working memory (PsychCorp, 2004). To bolster measurement of this cognitive domain on the WISC-IV, the Symbol Search task was changed to a core subtest, and a new supplemental subtest, Cancellation, which involves identifying targets within random and structured visual arrays, was added (Wechsler, 2003c). Coding, in nearly its identical form, was retained from previous WISC editions as a core PSI subtest. Thus, the WISC-IV PSI is composed of two core subtests (Symbol Search and Coding) and one supplemental subtest (Cancellation).

Construct Validity

The construct validity of the WISC-IV was evaluated by the test developers through the examination of composite and subtest score intercorrelations, as well as exploratory and confirmatory factor analyses (Wechsler, 2003c). With respect to intercorrelation research, as predicted, all WISC-IV Indexes are significantly correlated, providing
empirical support for the FSIQ. Similarly, demonstrating the validity of the Verbal Comprehension, Working Memory, and Processing Speed Scales, subtests within each of these Indexes correlate more highly with each other than with subtests associated with other scales (Wechsler). In contrast, a number of subtests composing the PRI correlate almost as highly with subtests on the VCI as with subtests within the same scale, a finding that calls into question the precise constructs measured by the Perceptual Reasoning scale (Baron, 2005). Block Design exhibits the highest correlations with other PRI subtests, and shows the least extreme cross-correlation pattern. Conversely, Picture Completion and Picture Concepts evidence relatively high correlations with subtests on the VCI, ostensibly reflecting “children’s use of verbal mediation in problem solving and formulating responses to these subtests” (Wechsler, p. 50).

Exploratory and confirmatory factor-analytic studies conducted on normative, clinical, and referred samples support the four factor model espoused in the WISC-IV manual (Bodin, Pardini, Burns, & Stevens, 2009; Sattler & Dumont, 2004; Watkins, Wison, Kotz, Carbone, & Babula, 2006; Wechsler, 2003c.) Mirroring the intercorrelation patterns described above, however, Picture Completion exhibits a significant secondary loading on the VCI, and in younger children the factor loading for Picture Concepts is evenly divided between Perceptual Reasoning and Verbal Comprehension Indexes (Kaufman et al., 2006; Keith et al., 2006; Williams, e al., 2003b; Wechsler, 2003c).

Further calling into question the constructs measured by the PRI, findings from an independent confirmatory factor analysis of the standardization sample led Keith et al. (2006) to adopt a five-factor model of the WISC-IV which includes a novel grouping of subtests included on the Perceptual Reasoning scale. Interpreting their findings through
the lens of the Cattell-Horn-Carroll (CHC) theory of cognitive abilities (see McGrew, 2005), the authors suggested that the core subtests contributing to the VCI, WMI, and PSI could be confidently interpreted as measures of Crystallized Intelligence (Gc), Short-term/Working Memory (Gsm), and Processing Speed (Gs), respectively. Conversely, the PRI, according to these researchers, should not be interpreted as a measure of perceptual reasoning, but rather a combination of two cognitive abilities; namely, fluid reasoning (Gf) and visual processing (Gv). Thus, based on their confirmatory factor loadings, Keith et al. recommended splitting the PRI into two separate scales, with Matrix Reasoning, Picture Concepts, and Arithmetic representing measures of Gf, and Block Design and Picture Completion representing measures of Gv. It should be noted that, in contrast to the studies supporting the four factor model, this factor structure is dependent on the inclusion of two supplemental subtests: Picture Completion and Arithmetic.

**WISC-III versus WISC-IV**

As anticipated, findings from studies examining the relationship between WISC-III and WISC-IV performances generally support the Flynn effect (Flynn & Weiss, 2007; Wechsler, 2003c; Williams et al., 2003b). On average, children’s FSIQ scores on the WISC-IV are approximately 2.5 points lower than FSIQ scores obtained on the WISC-III (Williams et al.), a discrepancy that is slightly lower than observed in previous revisions, likely reflecting, at least in part, the substantial differences in the subtests contributing to the FSIQ (Flynn & Weiss, 2007). Correlation coefficients of the Full Scale and Index scores for the two instruments range from .72 (WMI-FDI and PRI-POI) to .89 (FSIQ). According to the test developers, the low correlations associated with the former Indexes
reflect the extensive structural changes made to these particular scales during the revision process (Wechsler).

Despite consensus that the WISC has been dramatically altered during its most recent revision, claims have been made that WISC-IV interpretation should proceed on the vast literature base available for the WISC-III (e.g., Gabel & Shaughnessy, 2006; Prifitera, Saklofske, & Weiss, 2005; Wechsler, 2003c; Williams et al., 2003a). These suggestions, however, should be exercised with due caution considering that the constructs measured by the WISC-IV have been deliberately altered, and to some extent have yet to be elucidated (Baron, 2005; Kaufman et al., 2006; Needleman, Schnoes, & Ellis, 2006). Supporting this claim, the few studies that have compared WISC-III and WISC-IV patterns in known clinical groups (e.g., ADHD; TBI) suggest that these measures perform quite differently (e.g., Donders & Janke, 2008; Mayes & Calhoun, 2006, 2008; Allen, Thaler, Donohue, & Mayfield, 2010). The latter studies will be reviewed in a subsequent section of this manuscript.

Summary of Changes Apparent in the WISC-IV

On the basis of theoretical and empirical advances in the field of intellectual assessment, manifold changes are apparent in the WISC-IV. Among the most salient changes made to the instrument was the supplantation of the Verbal-Performance dichotomy by the Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed Indexes. Relative to previous versions of the WISC, the WISC-IV places greater emphases on fluid, non-crystallized, intellectual abilities. Reflecting this perspective, several new subtests have been added to the battery, a number of long-standing subtests have been omitted, and the retained subtests have undergone
considerable modifications. The psychometric properties of the instrument remain adequate, and the clinical utility and user-friendliness of the scale have ostensibly been improved. Although there is certainly evidence to support the four-factor model upon which the WISC-IV scoring system is based, there is research to suggest that a five-factor model fits the data better when Arithmetic and Picture Completion are included in the analysis. The fifth factor results from the splitting of the PRI into two factors, ostensibly reflecting visual processing (Block Design and Picture Completion) and fluid reasoning (Picture Concepts, Matrix Reasoning, and Arithmetic) skills. Thus, some members of the psychological community have questioned the constructs underlying the PRI (Baron, 2005; Keith et al., 2006), and have challenged the decision to add the Picture Concepts subtest to the WISC-IV (Baron). Despite the substantial changes apparent in the new WISC, it has been suggested that interpretation of this measure proceed on the basis of empirical data generated on earlier WISC editions. This practice does not appear to be warranted considering its limited research base in children with exceptionalities.

Following is a review of the extant research exploring WISC-IV performance patterns in exceptional populations including children with persistent academic difficulties.

WISC-IV Patterns of Performance

To date, no published study has empirically delineated subgroups of children based on WISC-IV profiles. Thus, the existing literature has taken a clinical-inferential approach to studying WISC-IV performance in groups of exceptional children.

Donders and Janke (2008) compared WISC-IV performance of children with traumatic brain injuries (TBI) to that of demographically matched controls. This study was predicated on a considerable body of literature demonstrating particular sensitivity of
the WISC-III Perceptual Organization and Processing Speed Indexes to TBI severity. Given the substantial differences between the WISC-III POI and the WISC-IV PRI, the authors questioned the extent to which the latter index would also be sensitive to injury severity in children with TBI. Contrary to research involving the WISC-III, the results of this study indicated that only the PSI differed significantly between the groups and yielded a negative correlation with length of coma. The authors speculated that the lack of sensitivity of the PRI in this population likely reflects decreased emphasis on speeded performance relative to the WISC-III POI. In conclusion, the authors cautioned practitioners working with pediatric TBI populations not to expect WISC-IV index or subtest profiles similar to those found on the WISC-III. Additionally, it was recommended that attempts be made to determine if, in this population, the WISC-IV PRI subtests would be better interpreted along the two dimensions suggested by Keith et al. (2006; i.e., Gv and Gf).

Remarkably similar results were reported in a study that compared WISC-IV performance in a pediatric TBI sample to that of the standardization sample (Allen, Thaler, Donohue, & Mayfield, 2010). In addition to concluding that substantive differences were found between performance on the WISC-III POI and the WISC-IV PRI, these researchers acknowledged that WISC-IV patterns of performance likely vary between children with TBI. Based on the assumption that this variability would be useful in terms of identifying homogeneous subgroups of children with TBI, Allen et al. encouraged the use of cluster analysis to explore WISC-IV patterns in this population. According to these authors, this type of research would help identify the relationship
between various WISC-IV patterns and meaningful variables such as injury type, demographics, clinical presentation, or educational/functional outcomes.

Mayes and Calhoun (2006) studied WISC-III and WISC-IV patterns of performance in children with ADHD. According to these authors, studies using the WISC-III have consistently demonstrated a pattern of significantly lower FDI and PSI scores relative to VCI and POI scores in this population. To determine the extent to which a similar pattern would emerge using the WISC-IV, the authors compared WISC-III and WISC-IV profiles of two groups of children with ADHD (one group was administered the WISC-III and the other was administered the WISC-IV). Although the pattern of VCI and POI/PRI greater than FDI/WMI and PSI was found using both measures, a notable distinction was reported with respect to the magnitude of these differences. Specifically, in the WISC-IV sample, the VCI and PRI were significantly more discrepant from the WMI and FDI than were the analogous indexes on the WISC-III. Consistent with the studies reviewed above, these researchers speculated that the observed differences likely reflected reduced motor and speed demands associated with the PRI. Supporting this claim, the WISC-IV PRI was significantly higher than the WISC-III POI in this study, primarily reflecting higher scores on the Matrix Reasoning and Picture Concepts subtests relative to the Picture Arrangement and Object Assembly subtests. Additionally, the VCI in this study was significantly higher in the WISC-IV group, likely reflecting the elimination of the Information subtest in the calculation of the WISC-IV VCI (i.e., Information was the lowest VCI score in both groups). Based on these collective findings, the authors concluded that the WISC-IV may have an advantage over the WISC-III for the assessment of children with ADHD due to the reduction of
motor and speed demands on the PRI and crystallized processing demands on the VCI. Similar conclusions were drawn from the results of a study conducted by the same authors exploring WISC-IV performance in children with high-functioning autism (Mayes & Calhoun, 2007c).

In a subsequent study of children with ADHD, Mayes and Calhoun (2007b) sought to determine the WISC-IV correlates of academic achievement in reading, math, and written expression, and to compare the results derived from the WISC-IV to those derived from the WISC-III. The results of this investigation indicated that for both the WISC-III and the WISC-IV, the best predictor of a learning disability (based on discrepancy criteria) was poor performance on WMI/FDI and PSI subtests. According to these authors, regardless of the WISC version employed, 67% of children were accurately identified with or without a learning disability on the basis of these subtest scores.

Further exploring WISC-IV WMI and PSI performances in children with persistent academic difficulties, Weiss and Gabel (2008) conducted a study using the WISC-IV Cognitive Proficiency Index (CPI). The CPI combines the WISC-IV WMI and PSI into a single score that reflects "a set of functions whose common element is the proficiency with which a person processes certain types of cognitive information" (Weiss & Gabel, 2008 pg. 1). The CPI is considered to be in stark contrast to the General Abilities Index (GAI), which represents the combination of the VCI and PRI, providing a summary score that reduces the influence of working memory and processing speed (Raiford, Weiss, Rolfhus, & Coalson, 2005). As part of their study, Weiss and Gable (2008) examined sensitivity and specificity rates for classifying children into various clinical groups based on the presence of a CPI < GAI pattern. Of the 12 clinical groups studied, acceptable
rates were found for only four: closed-head TBI; open-head TBI; Asperger’s Disorder; and combined reading and writing disabilities. Of particular interest in relation to the present investigation is that this pattern did not yield acceptable sensitivity/specificity rates for groups consisting of children with isolated reading or math difficulties suggesting that this pattern may be somewhat more characteristic of children with combined reading and writing disorders.

Other attempts to identify WISC-IV profile patterns in exceptional children were conducted by the WISC-IV developers, and are presented in the *Special Group Studies* section of the *Technical and Interpretive Manual* (Wechsler, 2003c). In these studies, small groups of children diagnosed with various conditions were compared to matched control groups on the basis of mean WISC-IV performances. The studies of interest within the context of the current literature review involve children diagnosed with disorders of (1) reading; (2) reading and written expression; (3) mathematics; and (4) reading, written expression, and mathematics. These studies are reviewed below.

The first special group study of interest investigated children diagnosed with Reading Disability (RD) on the basis of DSM-IV criteria. Relative to control participants, children with RD exhibited significantly lower scores on all WISC-IV composites, with large effect sizes apparent for the FSIQ, VCI and WMI. With respect to individual subtest differences, large effect sizes were obtained for Vocabulary, Letter-Number Sequencing, Information, and Arithmetic. According to the authors, these results are consistent with contemporary research that has established a connection between working memory and reading performance, and demonstrates the negative impact of RD on the acquisition of general verbal information (Wechsler, 2003c).
The second special group study examined children with combined Reading and Written Expression Disorders (RWD). In comparison to the control group, the RWD group exhibited significantly lower mean scores on all composite scores with the exception of the PRI. According to the researchers, both RD and RWD groups exhibit moderate to large effect sizes for the WMI; but “the RWD group is distinguished from the RD group by the PSI, which seems to play a larger role in the RWD group” (Wechsler, 2003c).

The third special group study examined children with Mathematics Disorder (MD). All mean composite scores, with the exception of the PSI, were lower in the MD group relative to the control sample. The largest effect size at the Index level involved the PRI, and the largest subtest effect sizes were observed on the Arithmetic subtest followed by Picture Concepts. These results support previous research indicating a relationship between MD and deficits in visual spatial processing (e.g., Clifford, 2009; Hale, Fiorello et al., 2008; Kulp et al., 2004; Proctor, Floyd, & Shaver, 2005; Rourke, 1994) as well as nonverbal problem solving and reasoning (e.g., Fuchs et al., 2005; Hale, Fiorello et al., 2008; Hale, DeLuca, & Casey, 2008; Taub, Keith, Floyd, & McGrew, 2008). That being said, a recent review of the literature on MD concluded there is little evidence to support the notion that poor visual-spatial processing directly affects mathematical computation skills (Barnes, Fuchs, & Ewing-Cobbs, 2010). Rather, visual-spatial skills appear to affect performance on higher order mathematical tasks (e.g., algebra, word problems, trigonometry and geometry), and may impact the development of a number of early math skills such as quantity matching and cardinality (i.e., the understanding that the last count in a set reflects the set’s quantity; Barnes et al.). Further, weaknesses in nonverbal and
fluid reasoning appear to be related more to math reasoning abilities than to basic arithmetic skills (Morris & Mather, 2008).

The fourth special group study presented in the manual explored WISC-IV profiles of children with combined Reading, Written Expression, and Mathematics Disorders (RWMD). In relation to the matched control group, the RWMD group obtained significantly lower scores on all WISC-IV composites and, unlike the other groups, demonstrated no discernable trends across cognitive domains (Wechsler, 2003c). This finding underscores "the importance of [identifying] homogeneous clinical groups, or grouping disorders according to common underlying neuropsychological processes when researching learning disorders" (Wechsler, p. 86).

**Summary of WISC-IV Patterns of Performance.** Studies comparing WISC-III and WISC-IV performance in groups of exceptional children indicate that these measures perform differently. The findings from this research suggest that the majority of the observed differences appear to relate to the reduction of motor and speed demands on the PRI, and have led to the recommendation that practitioners resist assuming that profiles found on the WISC-III will emerge in the same fashion on the WISC-IV. Other studies using clinical-inferential methodology to explore patterns of WISC-IV performance indicate that specific cognitive and academic weaknesses may be associated with distinct WISC-IV profiles. Specifically, relative to typically-achieving students, children with reading disorders perform most poorly on subtests that contribute to the VCI and WMI. Children exhibiting difficulties in both reading and written expression perform most poorly on the VCI, WMI, and PSI subtests. Children exhibiting problems in mathematics perform most poorly on PRI subtests, and children exhibiting problems in reading,
written expression, and mathematics exhibit poor performance on all WISC-IV subtests with no discernable pattern. Although certainly informative, the studies reviewed above suffer from a number of inherent limitations. These limitations are discussed below.

*Limitations of Existing WISC-IV Research.* A fundamental limitation of the existing research exploring WISC-IV performance in children with persistent academic difficulties relates to potential threats to generalizability due to the nature of the samples employed. Specifically, Mayes and Calhoun (2007b) included only children with ADHD in their investigation. As such, the degree to which these findings apply to children with academic difficulties in the absence of ADHD is unclear. Similarly, caution must be exercised when generalizing the results of the Wechsler (2003c) and Weiss and Gabel (2008) special group studies considering that the samples were small ($N = 32$ to $56$), participants represented only a portion of the WISC-IV age range (8 – 13 years), and the data was gathered from a variety of independent clinical settings that did not guarantee consistency of diagnostic criteria and procedures (Hebben, 2004; Wechsler). To address these limitations, future studies should employ large representative samples of children recruited from the same setting.

A second limitation of the existing literature relates to the methods used to define groups. In the Mayes and Calhoun (2007b) investigation, children with learning disabilities were identified on the basis of a significant discrepancy between WISC-IV and WIAT-II performances. Similarly, the learning disability groups investigated by Wechsler (2003c) and Weiss and Gabel (2008) were formed on the basis of DSM-IV-TR criteria, which also require the presence of a discrepancy between psychometric intelligence and academic achievement. A fundamental issue here is that the validity of
the WISC-IV patterns reported in these studies hinges on the accuracy of the theoretical model upon which the groups are formed. Given the substantive problems associated with the use of the discrepancy model, the WISC-IV profiles described in these studies are of questionable utility. Considering unresolved issues in the literature regarding the definition of learning disorders (Fletcher et al., 2007), research exploring WISC-IV patterns in children evidencing persistent academic difficulties should attempt to empirically, rather than clinically, define groups. To this end, use of a referred sample would be advisable, as this would make possible the identification of homogeneous subgroups within a heterogeneous population of struggling students.

The final and arguably most compelling limitation of the existing literature involves the use of the group-means method for identifying WISC-IV profiles in children with academic difficulties (i.e., reporting variable means in a clinically-defined sample). As discussed previously, even in well-defined samples, group-mean data obscures performance heterogeneity, thereby providing only minimal meaningful information regarding patterns of performance (Lange, 2007). Based on this limitation, the need for empirical clustering research exploring WISC-IV patterns in exceptional children has been clearly articulated in the scientific literature (e.g., Allen et al., 2010; Wechsler, 2003c).

1.8 Summary of Literature Review and Rationale for the Current Study

Procedures for evaluating children with persistent academic difficulties have changed dramatically in recent years. The guidelines specified in IDEA (2004) have enabled schools to abandon the use of the IQ-Achievement discrepancy model of assessment in favor of a more empirically defensible method that involves delineating the
psychological profiles of children who fail to respond to evidence-based academic interventions. Within this context, measures of intellectual functioning continue to play an important role. Many psychologists believe that, when integrated with other sources of information, these tests provide a means for generating hypotheses regarding cognitive strengths and weaknesses, with the goal of informing educational programming.

The WISC is the most widely used individually-administered intelligence test worldwide. Since its initial publication in 1949, this instrument has been revised three times. The latest edition of this measure, the WISC-IV, is structurally distinct from its predecessors and represents the most dramatic change ever made to this battery. Despite the widespread use of this instrument, few studies have investigated WISC-IV performance in exceptional populations. Within the context of this limited research base, it has been recommended that WISC-IV interpretation proceed on the basis of empirical data amassed on previous versions of this instrument. This, however, is a questionable practice considering the extent to which the WISC has been modified. According to the Standards for Educational and Psychological Testing, the substantial changes made in the latest WISC revision necessitate research exploring the use of this instrument in populations with which it is most commonly employed (American Educational Research Association et al., 1999). Contrary to these Standards, however, the WISC-IV has not been adequately studied in children exhibiting persistent academic difficulties (Hale & Fiorello, 2004; Kaufman; Mayes & Calhoun, 2007b). The present study addresses this need.

The rationale for the current investigation is predicated, in part, on a substantial body of literature demonstrating the utility of previous versions of the WISC in
identifying subgroups of children based on patterns of cognitive strengths and weaknesses. Numerous studies employing empirical clustering methodology (e.g., cluster analysis) have reported distinct performance patterns in children with persistent academic difficulties using the WISC-R and the WISC-III. Despite considerable methodological and interpretive variability across investigations, some consistencies have emerged, which provide valuable information about the various intellectual profiles demonstrated by these children. Common to many of these studies are WISC profiles reflecting (1) verbal processing weaknesses; (2) nonverbal/perceptual organization weaknesses; (3) attention, sequencing, or working memory weaknesses (with or without processing speed weaknesses); (4) global processing weaknesses (with or without processing speed strengths); and (5) no observable weaknesses. These empirically derived profiles, which are generally consistent with factor analytic findings, have been adequately replicated across samples and procedures, and are valid in the sense that they relate to differences in neuropsychological functioning, socioemotional status, behavior, and academic achievement.

In contrast to the numerous cluster analytic studies using the WISC-R and WISC-III, no study has employed this methodology to delineate patterns of performance using the WISC-IV. The few clinical-inferential studies that have attempted to identify WISC-IV patterns in children with academic difficulties suffer from methodological problems that limit the conclusions that can be reasonably drawn from their results. As such, further research is needed to explore WISC-IV patterns of performance in children exhibiting persistent academic difficulties. To address the limitations apparent in the existing literature, research is needed that: (1) employs a large representative sample of
children obtained from the same setting; (2) focuses on a referred rather than clinically-defined sample; and (3) takes an empirical-clustering rather than group-means approach to identifying WISC-IV patterns in children with persistent academic difficulties.

1.9 Overview of the Current Study

The current investigation was designed to fill a gap in the scientific literature by addressing the limitations described above. Toward this goal, two cluster analytic studies were conducted on the basis of WISC-IV subtest scores obtained from a large heterogeneous sample of children. Participants were all referred for psychoeducational assessment through the same educational system due to persistent academic difficulties.

Toward the goal of developing a typology that would be relevant to the majority of WISC-IV users, Study 1 was conducted using only the ten core WISC-IV subtests. In contrast, Study 2 was conducted using ten core and three supplemental WISC-IV subtests. Information, Picture Completion, and Arithmetic subtests were selected for inclusion in the second study for two reasons. First, in previous WISC editions, these subtests represented part of the core battery. Thus, it was anticipated that their inclusion would facilitate comparisons between the representative profiles found in the current investigation and those derived on the basis of previous WISC editions. Second, in light of the ongoing debate regarding the constructs tapped by the PRI (i.e., perceptual reasoning skills versus fluid reasoning and visual processing skills), the inclusion of Picture Completion and Arithmetic provided the opportunity to determine if profiles reflecting subtest-variations consistent with the latter orientation would emerge.
The derived typologies were assessed for reliability via multiple-method and split-half techniques. External validity was explored by determining the extent to which WIAT-II achievement scores varied as a function of WISC-IV subgroup membership.

It was anticipated that reliable and valid WISC-IV subtest patterns would emerge in this investigation. A number of these patterns were expected to resemble profiles found in previous empirical-clustering studies (Study 1), while others were expected to reflect patterns unique to this new measure (Study 2). It was further predicted that some of the derived clusters would differ on measures of academic achievement. To date, no study of this kind has been published.
CHAPTER 2

METHOD

2.1 Study 1

2.1a Participants

Participants were drawn from a population of students referred for psychological assessment through the Greater Essex County District School Board (GECDSB; Windsor, Ontario). Referrals for psychological assessment at the GECDSB typically involve the following steps: (1) A child is identified by his or her teacher as a student who is experiencing significant difficulties at school; (2) The classroom teacher consults with the Learning Support Teacher (LST) who provides recommendations for in-class interventions based on the apparent needs of the student; (3) If the student does not show marked improvement in response to these interventions, a School Team Meeting (STM) is convened. This meeting generally includes the classroom teacher, LST, educational coordinator, school administrator, parent(s)/guardian(s), and possibly the psychologist. During the STM, additional in-class interventions are determined, and decisions are made regarding the need for an educational assessment and/or the need to pursue other avenues such as sensory testing or speech and language evaluation; (4) If the student continues to have difficulties after implementing the recommendations arising from the STM, a consultation with Psychological Services is requested; (5) In response to such a request, Psychological Services personnel review the students school record, which includes all previous assessment findings, report cards, and related documentation. (6) If, from this review, it is determined that a psychological evaluation may benefit the child, informed consent is sought from the parent or guardian and the child is assessed.
Full Sample

Although the sample employed in Study 1 originally consisted of 477 students, five children were deemed outliers on the basis of criteria that will be described later and as a result were excluded from the study. Thus, the final sample comprised 472 participants (169 girls and 303 boys) ranging in age from 8 years, 0 months to 16 years, 11 months ($M = 10.38; SD = 2.17$). Full Scale IQ scores ranged from 70 to 130 with a mean of 85.29 and a standard deviation of 8.05. The majority of students were referred for assessment due to academic concerns in isolation ($N = 416; 88.1\%$); the remainder were referred due to a combination of academic and behavioural concerns ($N = 56; 11.9\%$). All participants spoke English as their primary language, were not taking medication at the time of testing, and had valid scores for all tests pertinent to this investigation (i.e., WISC-IV core subtests; WIAT-II Word Reading, Spelling, and Numerical Operations). All participants were free of uncorrected sensory impairments, significant medical histories (e.g., head injury, cerebral palsy, and epilepsy), psychiatric conditions (e.g., bipolar disorder, obsessive compulsive disorder), and pervasive developmental disorders. Children previously diagnosed with ADHD were included in the study ($n = 17$).

Split Samples

To assess the reliability of the cluster solution derived on the basis of the full sample, following the initial cluster analysis, the sample was randomly split in half. Split-sample 1 comprised 236 students (80 girls and 156 boys) with a mean FSIQ of 85.03 ($SD = 7.63$) and a mean age of 10.41 years ($SD = 2.23$). Of the children included in this sample, 207 (87.7\%) were referred for assessment due to academic difficulties
exclusively; 29 (12.3%) were referred due to academic and behavioural concerns combined. Similarly, split-sample 2 consisted of 236 students (89 girls and 147 boys) with a mean FSIQ of 85.57 ($SD = 8.47$) and a mean age of 10.35 years ($SD = 2.12$). With respect to reason for referral, 209 students (88.6%) were referred exclusively due to academic concerns; 27 students (11.4%) were referred due to concerns regarding academic and behavioural functioning combined. The sub-samples did not differ significantly in terms of age ($t \ [470] = .32, p = .29$) or FSIQ ($t \ [470] = -.720, p = .19$).

2.1b Measures

The measures employed in this investigation were administered and scored by trained examiners at the GECDSB under the direction of the Supervisor for Psychological Services between the years of 2004 and 2008. All scores were obtained within the context of a psychoeducational assessment, which in addition to the WISC-IV and WIAT-II sometimes included measures of phonological processing, memory, executive functioning, behavior, socioemotional status, or adaptive functioning. Informed consent for each assessment was obtained from parents or other legal guardians prior to test administration. All scaled scores were generated according to age-based Canadian norms.

WISC-IV

The relationship between the WISC-IV subtests, Indexes, and FSIQ is depicted in Appendix C. A description of each the WISC-IV subtests included in this investigation is presented in Appendix D. Scaled scores from the ten core WISC-IV subtests were used in this study: Block Design (BD), Similarities (SI), Digit Span (DS), Picture Concepts (PCn), Coding (CD), Vocabulary (VC), Letter-Number Sequencing (LN), Matrix
Reasoning (MR), Comprehension (CO), and Symbol Search (SS). The five supplemental WISC-IV subtests (i.e., Picture Completion [PCm], Cancellation [CA], Word Reasoning [WR], Information [IN], and Arithmetic [AR]) were not included in this study due to the assumption that a taxonomy derived on the basis of the core battery would be relevant to more psychologists than a taxonomy based on all 15 subtests given the inconsistent use of the latter measures (Konold, Glutting, McDermott, Kush, & Watkins, 1999).

**WIAT-II (Canadian Adaptation)**

The WIAT-II is a comprehensive individually-administered battery used to assess achievement in a variety of academic domains. This measure comprises nine subtests (Word Reading, Reading Comprehension, Pseudoword Decoding, Numerical Operations, Math Reasoning, Spelling, Written Expression, Listening Comprehension, and Oral Expression), which are organized into five primary achievement domains (Reading, Mathematics, Written Language, Oral Language, Total Achievement). The reliability and validity of the WIAT-II is well documented and indicates that this instrument is a stable and useful tool for academic assessment (PsychCorp, 2003). Age-based split-half correlations for all three subtests included in the current investigation meet or exceed .90 (Word Reading .96, Numerical Operations .90, and Spelling .93) and correlation coefficients between apposite scores on the WIAT-II, the original WIAT, and the Wide Range Achievement Test-Third Edition (WRAT-3, Wilkinson, 1993) all exceed .72 (Wechsler, 2002a).

The WIAT-II is the only achievement battery empirically linked with the WISC-IV and is widely used to assess academic functioning in children referred for psychoeducational assessment (Wechsler, 2002a). According to the **WISC-IV Canadian**
Manual (Wechsler, 2003c), the WISC-IV correlates moderately to highly with the WIAT-II. As would be expected, the WIAT-II Total Achievement Composite correlates most highly with the WISC-IV FSIQ \( r = .80 \) and least highly with the PSI \( r = .18 \). At the Index/Composite level, the VCI and WMI correlate most highly with the Reading Composite \( r = .76 \) and \(.63\), respectively) and the PRI and PSI correlate most highly with the Mathematics Composite \( r = .75 \) and \(.39\), respectively).

The WIAT-II was selected as the external validation measure in the current investigation due to the obvious clinical relevance of achievement patterns in children evidencing academic difficulties, and the frequent use of this instrument in psychoeducational evaluations (Wechsler, 2002a). The same WIAT-II subtests were used in both studies; namely, Word Reading, Spelling, and Numerical Operations. These subtests provide a means for identifying achievement patterns in three fundamental academic domains (i.e., single-word reading, spelling, and arithmetic; Wechsler) and as evidenced by their inclusion in the WIAT-II – Abbreviated battery (Wechsler, 2001) are assumed to be among the most frequently used WIAT-II subtests.

For a description of the WIAT-II subtests selected for inclusion in the proposed study refer to Appendix E. All analyses were based on WIAT-II subtest standard scores \( M = 100; SD = 15 \).

2.1c Procedures

Data Screening

Given the heterogeneous nature of the sample employed in this investigation, univariate outliers, defined as participants with WISC-IV subtest Z-scores greater than 3.29 \( p < .001 \); Field, 2005), were considered part of the target population and as a result
were retained for analyses. In contrast, multivariate outliers, conservatively identified on the basis of Mahalanobis distance \((p < .001; \text{Tabachnick \\& Fidell, 2001})\), were excluded from the study, as they have the potential to significantly distort the results of a cluster analysis (Hair \\& Black, 1998). The WISC-IV subtest profiles characterizing all omitted outliers were plotted and reviewed, as these individuals may represent an undersampling of naturally-occurring groups rather than truly aberrant observations (Hair \\& Black).

Initial Cluster Analysis

Considering that data input order can significantly affect the results of hierarchical cluster analyses due to potential ties in similarity coefficients (Podani, 1997), prior to all cluster analyses, the data were arranged to reflect optimal input order. Optimal input order was determined by running hierarchical cluster analyses under 1,000 random permutations of the data using PermuCLUSTER software (van der Kloot, Spaans, \\& Heiser, 2005). PermuCLUSTER is an SPSS add-on that computes a ‘goodness-of-fit’ index for each solution derived on the basis of each random permutation. Based on these ‘goodness-of-fit’ indexes, the optimal input order was identified. Once the data had been re-ordered to reflect the latter, WISC-IV scaled subtest scores were subjected to a two-stage process involving hierarchical and \(k\)-means iterative partitioning cluster analyses.

First, to estimate the number of clusters present in the data, a hierarchical agglomerative cluster analysis was conducted using Ward’s minimum variance method with squared Euclidean distance as the measure of similarity. Ward’s method was selected for several reasons: (1) it produces subgroups in which within-cluster variance is minimized (Milligan \\& Cooper, 1987); (2) it has been shown empirically to be among the
best performing hierarchical clustering algorithms in terms of recovering known typological structure (Blashfield, 1976; Milligan & Cooper; Overall, Gibson, & Novy, 1993); and (3) it has been successfully employed in previous empirical clustering research using the WISC-R (e.g., Holcomb et al., 1987; Snow et al., 1985) and the WISC-III (e.g., Ward et al., 1999; Saunders et al., 2001; Waxman & Casey, 2006). Squared Euclidean distance was chosen as the similarity measure as it is the recommended method for use with Ward’s clustering algorithm (Hair & Black, 2000), and because it is sensitive to profile shape, elevation, and scatter (Morris & Fletcher, 1988). Potential cluster solutions were identified based on a review of the agglomeration schedule (i.e., a display of the entities combined at each stage of the analysis and the resultant within-cluster variability) and visual inspection of the dendrogram (i.e., a graphical representation of the results of the hierarchical analysis). With respect to the latter, a sudden increase in within-cluster variability, as measured by percentage change, suggests that two clusters were combined to form a heterogeneous entity (i.e., dissimilar clusters were being combined). Thus, cluster solutions immediately preceding such mergers were considered potential solutions (Hair & Black, 2000). Refer to Appendix F for an example of an agglomeration schedule and a dendrogram.

The second stage of the analysis involved a procedure recommended by DeLuca et al., (1991) for refining the tentative cluster solutions derived from the first stage hierarchical analysis. Specifically, to circumvent one of the greatest shortcomings of hierarchical techniques (i.e., no provision for the relocation of cases that have been improperly classified), for each solution, $k$-means iterative partitioning methods were employed to correct potential fusion errors, allowing participants to be reassigned to
more appropriate clusters if necessary. This was accomplished by using cluster centroids from each of the suggested first-stage hierarchical solutions as seeds for determining the final cluster centers. Mean subtest scores were calculated and plotted for each derived cluster to generate representative WISC-IV profiles. The resultant profiles were then visually inspected to determine the extent to which they could be clinically interpreted. Although subjective, clinical interpretability is an important consideration in empirical clustering research (Kamphaus, Distefano, & Lease, 2003; Morris et al., 1981). Those cluster solutions characterized by interpretable profiles were forwarded for stability analysis, the first step of the internal validation process described below.

Internal Validation

The internal validation procedures for Study 1 are depicted in Figure 1.

Hierarchical and K-Means Stability. Morris et al. (1981) contended that the number of cases changing cluster membership from the first stage (hierarchical) solution to the second stage (k-means) solution reflects an index of cluster stability. Thus, all prospective solutions were compared on the basis of association between the results from the hierarchical analysis and the results from the k-means relocation pass. Cohen's kappa and one-way intraclass correlation coefficients (ICCs) were used to compare these solutions.

Cohen's kappa measures agreement between two cluster solutions by evaluating the extent to which subjects who are grouped together on one occasion are also grouped together on another. Kappa values range from -1 to 1, reflecting less than chance agreement to perfect agreement, respectively (Cohen, 1960). Of the available agreement
Hierarchical – K-means Stability Analysis

Multiple-Method Reliability Analysis

Split-Sample Reliability Analysis

Figure 1. Internal validation procedures for Study 1.
Note: Alb = Average Linkage Between Groups; ALw = Average Linkage Within Groups; CL = Complete Linkage; ICC = Intraclass Correlation Coefficient
measures, Cohen's kappa was selected due to its frequent use in psychological research, consideration of chance agreement, and demonstrated equivalence to alternative measures of association (e.g., Adjusted Rand Index; Warrens, 2008). Kappa values in this investigation were interpreted based on a system suggested by Landis and Koch (1977). This system is shown in Appendix G.

The one-way ICC (Haggard, 1958) was used to evaluate similarity between the mean profiles generated from the initial Ward's analysis and those derived on the basis of the k-means analysis. The ICC is based on an analysis of variance (ANOVA) model, providing an index of the proportion of variance shared by different profiles (Haggard). The ICC was chosen over more traditional correlation coefficients (e.g., Pearson Product Correlation - r) because of its ability to simultaneously evaluate more than two profiles, sensitivity to both profile shape and elevation, good performance in studies comparing alternative methods, and successful use in similar empirical-clustering investigations (Curry & Thomson, 1982; Beg, Casey, & Saunders, 2007; Edelbrock & McLaughlin, 1980). On the basis of the above described analyses, all interpretable profiles characterized by stable solutions were forwarded for multiple-method reliability analyses.

Multiple-Method Reliability. Numerous reviews of cluster analytic methodology suggest that a high degree of similarity between cluster solutions derived from the same sample using different hierarchical algorithms provides an index of cluster reliability (e.g., DeLuca et al., 1991; Jones, Drummond, Saunders, & Strang, 2006; Lange, Iverson, Senior, & Chelune, 2002; Speece, 2003). The logic underlying this notion is that true structure in a dataset should emerge even when the rule for joining cases changes. Thus, a reliable cluster solution should not be dependent on a particular algorithm for its
derivation (Speece). In the current investigation, multiple-method reliability of all potential cluster solutions was assessed by applying three new hierarchical agglomerative algorithms to the data (Complete Linkage, Average Linkage Between Groups, and Average Linkage Within Groups) and specifying the number of clusters to be recovered. A k-means relocation pass through the data followed using the cluster centroids from each of the hierarchical analyses as starting points. Cohen’s kappa was used to assess membership agreement between each new cluster solution and the analogous solution generated on the basis of the two-stage Ward’s analysis. ICCs were used to compare the solutions in terms of profile similarity. All solutions that were well replicated across the four agglomerative methods were sent on for split-half reliability analyses.

*Split-Half Reliability.* To determine the extent to which the cluster solutions derived in Study 1 could be replicated in different samples, the initial sample was randomly split in half and each subsample was subjected to a two-stage Ward’s analysis specifying the number of clusters to be recovered. The mean cluster profiles derived from these split-half samples were visually inspected and assessed for similarity via ICC. The cluster solution replicated across split-half samples was identified as the final solution.

Once the final cluster solution was determined, mean subtest patterns, based on the initial Ward’s analysis, were labeled according to their most salient features. Cluster demographics were then calculated, and the solution was assessed for external validity.

*External Validation*

The external validity of a cluster solution addresses the degree to which empirically-derived subgroups can be distinguished on the basis of theoretically
important variables not used in the cluster analysis (Fletcher, 1985). External validity in the proposed study was assessed by comparing the derived WISC-IV subgroups on the basis of mean performance (i.e., standard scores) on WIAT-II Word Reading, Spelling, and Numerical Operations subtests. Multivariate analysis of variance (MANOVA) was conducted to determine if the derived subgroups differed on these external variables. In response to significant MANOVA findings, univariate analyses (ANOVA) with subsequent post hoc comparisons (Games-Howell procedure) were conducted to determine which WISC-IV clusters differed on the basis of which WIAT-II subtests. The Games-Howell Test, which provides a correction for multiple comparisons, was selected over other post hoc procedures (e.g., Tukey’s HSD, Bonferroni-corrected t-tests) because it represents the most appropriate form of post-hoc analyses for unbalanced designs (i.e., unequal samples sizes; Field, 2005). Effect sizes were calculated using Cohen’s $d$, and were interpreted on the basis of suggestions made by Cohen (1988): $d = .20$ (small effect); $d = .50$ (medium effect); $d = .80$ (large effect).

Due to the limited range of WIAT-II subtests employed, and the fact that many children with persistent academic difficulties exhibit difficulty in all three achievement areas included in this study (Fletcher et al., 2007), it was expected that some, but not all, of the derived subgroups would differ on the basis of these measures.

2.1d Hypotheses

Based on cluster analytic research conducted on previous WISC editions, and in keeping with WISC-IV research findings, the following predictions were made:
Initial Cluster Analysis

1. It was hypothesized that this cluster analytic investigation would yield a reliable typology that would include clusters characterized by the following mean WISC-IV subtest patterns:
   
a. relative weaknesses on subtests associated with the VCI (i.e., Vocabulary, Comprehension, Similarities);
   
b. relative weaknesses on subtests associated with the VCI and WMI (i.e., Vocabulary, Comprehension, Similarities, Digit Span, and Letter-Number Sequencing – similar to the VIQ in previous WISC editions);
   
c. relative weaknesses on subtests associated with the WMI and PSI (i.e., Digit Span, Letter Number Sequencing, Coding, and Symbol Search);
   
d. relative weaknesses on subtests associated with the PRI (i.e., Block Design, Picture Concepts, and Matrix Reasoning);
   
e. relative weaknesses on subtests associated with the PRI and PSI (i.e., Block Design, Picture Concepts, Matrix Reasoning, Coding, Symbol Search – similar to the PIQ in previous WISC editions);
   
f. relatively low scores on all subtests included in the analyses (all subtest scaled scores lower than 8);

Internal Validation

2. It was hypothesized that the derived typology would exhibit adequate internal validity as evidenced by:
   
a. a statistically significant level of association in case membership and profile similarity between the clusters generated during the initial hierarchical cluster analysis and the final k-means iterative partitioning cluster analysis;
   
b. a statistically significant level of association in case membership and profile similarity between the final clusters generated using Ward’s method and those generated using three alternative hierarchical algorithms;
   
c. a statistically significant level of profile similarity between the final Ward’s clusters derived in each of the split-half samples.
**External Validation**

3. Based broadly on the available literature, it was hypothesized that WIAT-II Reading, Spelling, and Numerical Operations mean subtest standard scores would differ significantly as a function of WISC-IV subgroup membership such that:

a. a cluster characterized by low scores across all subtests (i.e., globally low) would be expected to perform poorly relative to the other clusters on all three WIAT-II subtests;

b. clusters characterized by low scores on subtests contributing to the VCI and WMI (i.e., low VCI; low VCI & WMI; low WMI & PSI) would be expected to demonstrate lower mean scores on Word Reading and Spelling subtests relative to clusters characterized by low scores on the PRI and PSI subtests (i.e., low PRI; and low PRI & PSI clusters);

c. clusters characterized by low scores on subtests contributing to the PRI and PSI (i.e., low PRI; and low PRI and PSI) would be expected to exhibit lower mean scores on the Numerical Operations subtest compared to those clusters characterized by low scores on VCI and WMI subtests.

**2.2 Study 2**

**2.2a Participants**

The participants in Study 2 represented a subset of the sample employed in Study 1. Children included in this study were selected due to the availability of scores for the Information, Picture Completion, and Arithmetic subtests in addition to the subtests used in Study 1. The total sample for Study 2 \((N = 115)\) consisted of 40 girls and 75 boys. No students were identified as multivariate outliers; as such, data from all students were retained for analyses. The mean age for this sample was 10.23 years \((SD = 1.97)\), and the mean FSIQ was 84.44 \((SD = 7.26)\). A total of 102 children \((88.7\%)\) were referred for assessment due to academic concerns exclusively; 13 students \((11.3\%)\) were referred due
to a combination of academic and behavioural concerns. Three children included in this study were diagnosed with ADHD prior to assessment.

2.2b Measures

WISC-IV

To explore the effects of incorporating specific supplemental subtests into the cluster analysis, in addition to the WISC-IV core battery, three additional subtests were used to form clusters in this study; namely, Picture Completion (PCm), Information (IN), and Arithmetic (AR).

WIAT-II (Canadian Adaptation)

The WIAT-II subtests used in this study mirrored those of Study 1. Thus, Word Reading, Spelling, and Numerical Operations represented the external validation measures.

2.2c Procedures

The procedures employed in this study were identical to those of Study 1, with the exception of split-half reliability analysis, which was not possible due to a reduced sample size.

2.2d Hypotheses

Initial Cluster Analysis

1. In keeping with factor analytic research conducted on the WISC-IV, it was hypothesized that the Information subtest would vary with the other VCI subtests. Thus, those clusters characterized by relative weaknesses on core subtests associated with the VCI would also exhibit a relative weakness on the Information subtest.
2. Reflecting the factor analytic findings of Keith et al. (2006), it was hypothesized that in addition to the cluster mean profiles found in Study 1, subgroups characterized by the following WISC-IV patterns would emerge:

   a. relative weaknesses on those subtests associated with Gv (Block Design and Picture Completion);
   b. relative weaknesses on those subtests associated with Gf (Picture Concepts, Matrix Reasoning, and Arithmetic).

*Internal Validation*

3. It was hypothesized that the derived typology would exhibit adequate internal validity as evidenced by:

   a. a statistically significant level of association in case membership and profile similarity between the clusters generated during the initial hierarchical cluster analysis and the final k-means iterative partitioning cluster analysis;
   b. a statistically significant level of association in case membership and profile similarity between the final clusters generated using Ward’s method and those generated using three alternative hierarchical algorithms.

*External Validation*

4. It was hypothesized that WIAT-II Word Reading, Spelling, and Numerical Operations mean subtest standard scores would differ significantly as a function of WISC-IV subgroup membership such that clusters characterized by relatively low scores on subtests associated with Gv and Gf would demonstrate lower mean scores on the Numerical Operations subtest relative to those clusters characterized by weaknesses on subtests associated with the VCI and WMI.
CHAPTER 3

RESULTS

3.1 Study 1

3.1a Data Screening

Prior to conducting the initial cluster analysis, the dataset was screened for input accuracy, missing data, and multivariate outliers. Based on Mahalanobis distance ($p < .001$), five participants were deemed multivariate outliers and were removed from further analyses. The WISC-IV profiles of these participants are presented in Appendix H. Following removal of all multivariate outliers, two univariate outliers were identified. One outlier exhibited an exceedingly high score on the Similarities subtest ($Z = 3.49, p < .001$), while the other evidenced an inordinately high score on the Comprehension subtest ($Z = 3.53, p < .001$). As previously indicated, these participants were considered to be part of the target population and were, therefore, retained for further analyses.

3.1b Initial Cluster Analysis

To estimate the number of clusters that would best represent the data, a Ward’s hierarchical cluster analysis was conducted on the basis of WISC-IV core subtest scores (Stage 1). Review of the dendrogram and agglomeration schedule supported two-through eight-clusters. Following $k$-means analyses of these solutions (Stage 2), it was determined via visual inspection of the resultant mean profiles that two-, three-, and five-cluster solutions were the most interpretable. The two-cluster solution, which was characterized exclusively by low average and average performance, was excluded from further analyses due to limited subtest scatter. Although certainly meaningful from a
clinical perspective, the two-cluster solution was not consistent with the goal of the current investigation; namely, to examine profiles characterized by variability in both elevation and shape. The remaining interpretable cluster solutions were examined for internal validity.

3.1c Internal Validation

Hierarchical and K-means Stability

The stability of the three- and five-cluster solutions was evaluated by comparing the clusters generated on the basis of the first-stage Ward’s analysis to those derived from the second stage $k$-means analysis. Cluster membership concordance between the two stages was assessed via Cohen’s kappa. In both solutions, kappa values were statistically significant ($p < .001$). This suggests that, in general, the same participants were grouped together across the two stages of the initial cluster analysis.

Similarity between the mean profiles generated during each of the two stages was assessed via ICCs. All profiles in the three-cluster solution yielded ICCs that were significant at the $p < .01$ level. All profiles in the five-cluster solution yielded ICCs that were significant at the $p < .05$ level. These findings suggest that markedly similar profiles were derived during hierarchical and $k$-means stages of the analysis in both the three- and the five-cluster solution. Kappa values and ICCs for each solution are presented in Table 1. On the basis of these results, both solutions were deemed adequately stable, and were sent on for multiple-method reliability assessment.
Table 1

*Kappa Values and ICCs comparing prospective solutions generated during Stages 1 and 2 of the Initial Ward's Cluster Analysis in Study 1*

<table>
<thead>
<tr>
<th>Ward's Cluster Solution</th>
<th>Kappa</th>
<th>Cluster Number</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Cluster</td>
<td>.541***</td>
<td>1</td>
<td>.954***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>.877**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.924***</td>
</tr>
<tr>
<td>5-Cluster</td>
<td>.545***</td>
<td>1</td>
<td>.629*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>.867**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.872**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>.904***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>.979***</td>
</tr>
</tbody>
</table>

Note. ICC = Intraclass Correlation Coefficient.

* p < .05. ** p < .01. *** p < .001.
Multiple-Method Reliability

To assess the extent to which each of the prospective $k$-corrected Ward’s solutions would emerge using alternative methods, the dataset was subjected to three additional hierarchical clustering algorithms (i.e., Average Linkage Between Groups; Average Linkage Within Groups; and Complete Linkage) specifying recovery of three and five clusters. This procedure was followed by $k$-means analyses using hierarchical cluster centroids as initial seeds. Similarity of the cluster solutions derived using each hierarchical method was assessed via Cohen’s kappa and ICCs. Good agreement was obtained for both solutions. Kappa values suggested Fair to Almost Perfect agreement in the three-cluster solution, and Moderate agreement in the five-cluster solution. All kappa values were significant at the $p < .001$ level.

To assess profile similarity between the various solutions, ICCs were calculated for the mean WISC-IV profiles generated on the basis of Ward’s method and those derived on the basis of the other three hierarchical algorithms. All ICCs obtained from profiles in the three-cluster solution were significant ($p < .01$); conversely, a non-significant ICC was found for one profile within the five-cluster solution. Kappa values and ICCs are presented in Table 2. On the basis of these analyses, both solutions were deemed adequately replicable and were sent on for split-half reliability analyses.

Split-half Reliability

As a final measure of internal validity, the sample was randomly split in half, and each subsample was subjected to a two-stage Ward’s analysis specifying recovery of three and five clusters. The split-half profiles associated with the three-cluster solution had good visual agreement, and all ICCs were significant ($p < .05$). In contrast, the split-
Table 2

*Kappa Values and ICCs comparing Ward's solutions to solutions derived using three alternative hierarchical clustering techniques in Study 1*

<table>
<thead>
<tr>
<th>Ward's Cluster Solution</th>
<th>Kappa</th>
<th>Cluster Number</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AbG</td>
<td>AwG</td>
<td>CoL</td>
</tr>
<tr>
<td>3-Cluster</td>
<td>.975***</td>
<td>.699***</td>
<td>.263***</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.852***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.945***</td>
<td></td>
</tr>
<tr>
<td>5-Cluster</td>
<td>.590***</td>
<td>.589***</td>
<td>.494***</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.971**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.237</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.978***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.952***</td>
<td></td>
</tr>
</tbody>
</table>

half profiles from the five-cluster solution were difficult to match, and two of the ICCs were non-significant. This suggested that the same profiles had not emerged in both split samples, calling into question the reliability of the five-cluster solution. Split-half ICCs for both solutions are presented in Table 3. Based on these findings, the three-cluster solution was considered representative of the data, and was selected as the final solution.

3.1d Description of Clusters

The three clusters generated on the basis of the initial two-stage Ward's analysis were assigned descriptive labels reflecting the most salient features of each mean WISC-IV profile. These representative profiles are illustrated in Figures 2 through 4. Figure 5 depicts the similarities and differences among all three cluster profiles. Descriptions of the clusters were based on the *Three-Category Approach to Describing WISC-IV Subtest Scaled Scores* developed by Sattler and Dumont (2004). According to this system, scaled subtest scores between 8 and 12 are considered average; scores between 1 and 7 are considered below average; and scores between 13 and 19 are considered above average.

Cluster 1 (n = 180) was labeled Globally Low (GL) to reflect below average scores on all subtests with the exception of Picture Concepts. Cluster 2 (n = 166) was characterized by below average scores on subtests associated with the VCI. As such, this cluster was designated Low VCI (LVCI). Cluster 3 (n = 126) was labeled Lower WMI and PSI (LWMPIS) to reflect a generally average profile characterized by relatively low scores on subtests contributing to the WMI and PSI and a relatively high score on the Picture Concepts subtest. It is interesting to note that the latter subtest represented the highest score in all three clusters. WISC-IV subtest- and index-score means and
Table 3

*ICCs comparing mean WISC-IV profiles generated in each split-half sample in Study 1*

<table>
<thead>
<tr>
<th>Cluster Solution</th>
<th>Cluster Number</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Cluster</td>
<td>1</td>
<td>.868**</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.738*</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.908***</td>
</tr>
<tr>
<td>5-Cluster</td>
<td>1</td>
<td>.481</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.891***</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.940***</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.244</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.921***</td>
</tr>
</tbody>
</table>

Note. ICC = Intraclass Correlation Coefficient.

* *p < .05. ** p < .01. *** p < .001.
Figure 2. Mean WISC-IV profile for Cluster 1 in Study 1 (Globally Low).

Note: BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; VOC = Vocabulary; COM = Comprehension; SIM = Similarities; DS = Digit Span; LNS = Letter Number Sequencing; COD = Coding; SS = Symbol Search.
Figure 3. Mean WISC-IV profile for Cluster 2 in Study 1 (Low VCI).

Note: BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; VOC = Vocabulary; COM = Comprehension; SIM = Similarities; DS = Digit Span; LNS = Letter Number Sequencing; COD = Coding; SS = Symbol Search.
Figure 4. Mean WISC-IV profile for Cluster 3 in Study 1 (Lower WMI and PSI).

Note: BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; VOC = Vocabulary; COM = Comprehension; SIM = Similarities; DS = Digit Span; LNS = Letter Number Sequencing; COD = Coding; SS = Symbol Search.
Figure 5. Mean WISC-IV profiles for all clusters in Study 1.

Note: GL = Globally Low; LVCI = Low Verbal Comprehension Index; LWMPS = Lower Working Memory Index and Processing Speed Index; BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; VOC = Vocabulary; COM = Comprehension; SIM = Similarities; DS = Digit Span; LNS = Letter Number Sequencing; COD = Coding; SS = Symbol Search.
### Table 4

*WISC-IV Means and standard deviations for the final 3-Cluster solution in Study 1*

<table>
<thead>
<tr>
<th>WISC-IV Subtest/Index</th>
<th>Cluster 1 GL (n = 180)</th>
<th>Cluster 2 LVCI (n = 166)</th>
<th>Cluster 3 LWMPS (n = 126)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$ (SD)</td>
<td>$M$ (SD)</td>
<td>$M$ (SD)</td>
</tr>
<tr>
<td>Block Design</td>
<td>6.46 (1.86)</td>
<td>8.48 (2.23)</td>
<td>9.92 (2.53)</td>
</tr>
<tr>
<td>Picture Concepts</td>
<td>8.18 (2.56)</td>
<td>9.54 (2.22)</td>
<td>11.69 (2.38)</td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>6.78 (2.17)</td>
<td>9.33 (1.97)</td>
<td>10.33 (2.24)</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>6.82 (1.75)</td>
<td>6.81 (1.44)</td>
<td>9.34 (1.69)</td>
</tr>
<tr>
<td>Comprehension</td>
<td>7.32 (1.93)</td>
<td>7.61 (1.84)</td>
<td>9.05 (1.85)</td>
</tr>
<tr>
<td>Similarities</td>
<td>6.65 (2.06)</td>
<td>6.36 (1.71)</td>
<td>9.95 (1.95)</td>
</tr>
<tr>
<td>Digit Span</td>
<td>7.08 (1.94)</td>
<td>8.21 (2.27)</td>
<td>8.05 (2.17)</td>
</tr>
<tr>
<td>LN Sequencing</td>
<td>6.38 (2.30)</td>
<td>8.82 (1.91)</td>
<td>8.46 (1.99)</td>
</tr>
<tr>
<td>Coding</td>
<td>6.76 (1.89)</td>
<td>8.81 (2.14)</td>
<td>8.05 (2.15)</td>
</tr>
<tr>
<td>Symbol Search</td>
<td>6.90 (1.93)</td>
<td>9.06 (1.86)</td>
<td>8.67 (2.24)</td>
</tr>
<tr>
<td>PRI</td>
<td>81.78 (8.55)</td>
<td>93.69 (8.79)</td>
<td>104.24 (10.27)</td>
</tr>
<tr>
<td>VCI</td>
<td>82.52 (8.88)</td>
<td>82.56 (7.02)</td>
<td>97.03 (7.66)</td>
</tr>
<tr>
<td>WMI</td>
<td>81.43 (9.59)</td>
<td>91.31 (9.50)</td>
<td>89.94 (8.64)</td>
</tr>
<tr>
<td>PSI</td>
<td>82.21 (7.67)</td>
<td>93.85 (9.11)</td>
<td>90.42 (10.80)</td>
</tr>
<tr>
<td>FSIQ</td>
<td>78.03 (4.34)</td>
<td>86.68 (4.27)</td>
<td>93.84 (6.42)</td>
</tr>
</tbody>
</table>

*Note.* GL = Globally Low; LVCI = Low VCI; LWMPS = Lower WMI and PSI with Higher Picture Concepts
standard deviations for each of the three clusters are presented in Table 4. Age and
gender distributions for each cluster are presented in Table 5. There were no significant
differences in mean age as a function of cluster membership ($F[2, 469] = .111, p = .895$).

3.1e External Validation

The external validity of the final solution was assessed by determining the extent
to which the derived clusters could be differentiated on measures not used during group
formation. To this end, a MANOVA was employed to compare the three WISC-IV
clusters on the basis of mean performance on WIAT-II Word Reading, Spelling, and
Numerical Operations subtests. The MANOVA was found to be significant on the basis
of Pillai’s Trace ($F[6, 936] = 14.74, p < .001$), Wilks’ Lambda ($F[6, 934] = 15.39, p <
.001$), Hotelling’s Trace ($F[6, 932] = 16.049, p < .01$), and Roy’s Largest Root ($F[3,
468] = 31.481, p < .001$).

Three univariate ANOVAs were computed to assess the significance of each
WIAT-II variable. All ANOVAs were significant at the $p < .001$ level, indicating that the
empirically-derived WISC-IV clusters differed on all three WIAT-II subtests. A listing
of univariate F-scores, along with means and standard deviations for the WIAT-II
subtests, can be found in Table 6. Figure 6 illustrates mean WIAT-II subtest performance
for each of the WISC-IV clusters.

To determine which clusters were differentiated on the basis of each WIAT-II
subtest, post-hoc multiple comparisons were conducted using the Games-Howell Test.
This was followed by effect size calculations (Cohen’s $d$) to estimate the strength and
practical significance of the differences between means. Post-hoc results and effect size
Table 5

*Age and gender distribution for the final 3-cluster solution in Study 1*

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>M</th>
<th>F</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>108</td>
<td>72</td>
<td>50</td>
<td>34</td>
<td>23</td>
<td>29</td>
<td>10</td>
<td>14</td>
<td>10</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>10.32 (2.16)</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>40</td>
<td>28</td>
<td>19</td>
<td>13</td>
<td>16</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>LVCI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>105</td>
<td>61</td>
<td>38</td>
<td>34</td>
<td>24</td>
<td>21</td>
<td>12</td>
<td>23</td>
<td>11</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>10.43 (2.05)</td>
</tr>
<tr>
<td></td>
<td>63</td>
<td>37</td>
<td>23</td>
<td>20</td>
<td>14</td>
<td>13</td>
<td>7</td>
<td>14</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>LWMPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>90</td>
<td>36</td>
<td>41</td>
<td>22</td>
<td>13</td>
<td>8</td>
<td>14</td>
<td>17</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>10.37 (2.35)</td>
</tr>
<tr>
<td></td>
<td>71</td>
<td>29</td>
<td>33</td>
<td>17</td>
<td>10</td>
<td>6</td>
<td>11</td>
<td>13</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>303</td>
<td>169</td>
<td>129</td>
<td>90</td>
<td>60</td>
<td>58</td>
<td>36</td>
<td>54</td>
<td>25</td>
<td>12</td>
<td>8</td>
<td>2</td>
<td>10.38 (2.12)</td>
</tr>
</tbody>
</table>

Note. GL = Globally Low; LVCI = Low VCI; LWMPS = Lower WMI and PSI. All percentages are rounded to the nearest whole number.
Table 6

*Means, standard deviations, and univariate F-scores for differences on WIAT-II subtests (standard scores) based on cluster membership in Study 1*

<table>
<thead>
<tr>
<th>WIAT-II Subtests</th>
<th>WISC-IV Cluster</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GL</td>
<td>LVCI</td>
<td>LWMPS</td>
<td>F</td>
<td>p</td>
</tr>
<tr>
<td>Word Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>72.13</td>
<td>76.10</td>
<td>79.20</td>
<td>10.13</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>(SD)</td>
<td>(13.95)</td>
<td>(13.63)</td>
<td>(13.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spelling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>74.42</td>
<td>78.24</td>
<td>79.35</td>
<td>7.405</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>(SD)</td>
<td>(12.27)</td>
<td>(12.56)</td>
<td>(10.98)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numerical Operations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>71.07</td>
<td>80.50</td>
<td>82.91</td>
<td>42.58</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>(SD)</td>
<td>(11.40)</td>
<td>(12.36)</td>
<td>(12.96)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. GL = Globally Low; LVCI = Low VCI; LWMPS = Lower WMI and PSI.
Figure 6. Mean WIAT-II profiles for the WISC-IV clusters derived in Study 1.

Note: GL = Globally Low; LVI = Low Verbal Comprehension Index; LWMPS = Lower Working Memory Index and Processing Speed Index; Num Ops = Numerical Operations.

*Significantly different than the other two clusters on all WIAT-II subtests (p < .05).
estimates are presented in Table 7. Overall, Cluster 1 (GL) demonstrated the lowest mean performance across the WIAT-II subtests, differing significantly from the other two clusters on all three achievement variables. In contrast, Clusters 2 (LVCI) and 3 (LWMPS) did not differ significantly on any of the WIAT-II variables. All effect size estimates for between group differences were considered small with the exception of those calculated for the Numerical Operations subtest. In the latter case, large effect sizes were obtained for differences between Clusters 1 (GL) and 2 (LVCI; \( d = .80 \)) and between Clusters 1 and 3 (LWMPS; \( d = .97 \)). These results suggest that while the GL cluster performed poorly relative to the LVCI and LWMPS clusters on all WIAT-II subtests, the magnitude of the differences between groups is likely to have practical significance only in the case of Numerical Operations. Supporting this finding, the results of a discriminant function analysis indicated that Numerical Operations played the largest role in separating the GL cluster from the other two clusters (Wilks' Lamda = .828, \( p < .001 \)), with this model accounting for approximately 17% of the variance between groups.

3.2 Study 2

3.2a Data Screening

Prior to conducting the initial cluster analysis, the dataset was screened for input accuracy, missing data, and multivariate outliers. No multivariate outliers were identified on the basis of Mahalanobis distance (\( p < .001 \)) criteria. Conversely, three univariate outliers were identified. One outlier exhibited an exceedingly high score on the Similarities subtest (\( Z = 3.31, p < .001 \)), one exhibited an inordinately high score on the Arithmetic subtest (\( Z = 3.61, p < .001 \)), and one exhibited an exceedingly low score on the
Table 7

Post-hoc analyses (Games-Howell Test) and effect size estimates (Cohen's d) for mean differences between WISC-IV clusters on WIAT-II subtests in Study 1

<table>
<thead>
<tr>
<th>Cluster</th>
<th>GL</th>
<th>LVCI</th>
<th>LWMPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Difference</td>
<td></td>
<td>3.97*</td>
<td>7.07***</td>
</tr>
<tr>
<td>d</td>
<td>.28</td>
<td>.51</td>
<td></td>
</tr>
<tr>
<td>LVCI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Difference</td>
<td></td>
<td>-</td>
<td>3.09</td>
</tr>
<tr>
<td>d</td>
<td></td>
<td>.22</td>
<td></td>
</tr>
</tbody>
</table>

WIAT-II Word Reading

<table>
<thead>
<tr>
<th>Cluster</th>
<th>GL</th>
<th>LVCI</th>
<th>LWMPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Difference</td>
<td></td>
<td>3.82*</td>
<td>4.93**</td>
</tr>
<tr>
<td>d</td>
<td>.31</td>
<td>.42</td>
<td></td>
</tr>
<tr>
<td>LVCI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Difference</td>
<td></td>
<td>-</td>
<td>1.11</td>
</tr>
<tr>
<td>d</td>
<td></td>
<td>.09</td>
<td></td>
</tr>
</tbody>
</table>

WIAT-II Spelling

<table>
<thead>
<tr>
<th>Cluster</th>
<th>GL</th>
<th>LVCI</th>
<th>LWMPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Difference</td>
<td></td>
<td>9.43***</td>
<td>11.84***</td>
</tr>
<tr>
<td>d</td>
<td>.80</td>
<td>.97</td>
<td></td>
</tr>
<tr>
<td>LVCI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Difference</td>
<td></td>
<td>-</td>
<td>-2.40</td>
</tr>
<tr>
<td>d</td>
<td></td>
<td>.09</td>
<td></td>
</tr>
</tbody>
</table>

WIAT-II Numerical Operations

Note. GL = Globally Low; LVCI = Low VCI; LWMPS = Lower WMI and PSI.
* p < .05. ** p < .01. ***p < .001. Cohen’s d values in bold denote large effect sizes.
Information subtest \((Z = -3.29)\). As in study 1, these participants were retained for analyses based on the rationale that they represented part of the target population. Given that no participants were removed from the study, all analyses were conducted on a sample of \(N = 115\).

3.2b Initial Cluster Analysis

After re-ordering the data to reflect optimal input order, a Ward’s hierarchical cluster analysis was conducted to estimate the number of clusters that would best represent the data (Stage 1). A review of the dendrogram and agglomeration schedule supported two- through five-clusters. Following \(k\)-means analyses of these solutions specifying recovery of two through five clusters (Stage 2), it was determined via visual inspection of the resultant mean profiles that two-, three-, and four-cluster solutions were the most interpretable. In keeping with the procedures of Study 1, the two-cluster solution was excluded from further analyses as the clusters were defined almost exclusively by level of performance (i.e., relatively ‘flat’ profiles characterized by either low or average scores). The three- and four-cluster solutions were forwarded for internal validity assessment.

3.2c Internal Validation

Hierarchical and \(K\)-means Stability

Cluster membership concordance between the two stages of the initial hierarchical analysis was assessed via Cohen’s kappa. In both solutions, the results indicated substantial agreement between the two stages, with kappa values significant at the \(p < .001\) level. This suggests that in both solutions, many of the participants who were
grouped together on the basis of hierarchical analyses were also grouped together on the basis of \( k \)-means analyses.

Mirroring these findings, all mean profiles in both solutions were found to have significant ICCs \((p < .001)\), suggesting that markedly similar profiles emerged during each stage of the analysis. Kappa values and ICCs for both the three- and four-cluster solution are presented in Table 8. On the basis of these results, both solutions were deemed adequately stable, and were sent on for multiple-method reliability assessment.

*Multiple-Method Reliability*

When the prospective solutions were subjected to three additional \( k \)-corrected hierarchical clustering algorithms (Average Linkage Between Groups, Average Linkage Within Groups, and Complete Linkage), the results suggested that the three-cluster solution best represented the data. Although kappa values and ICCs across the various \( k \)-corrected hierarchical algorithms were statistically significant for both solutions (see Table 9), a number of the cluster-mean profiles in the four-cluster solution were visually dissimilar, and some of the derived clusters were exceedingly small (e.g., \( n = 2 \)). Given that the three-cluster solution evidenced visually similar profiles and consistently adequate cluster sizes, it was selected as the final cluster solution in Study 2.

**3.2d Description of Clusters**

As in Study 1, the three clusters generated on the basis of the initial two-stage Ward’s analysis were assigned descriptive labels reflecting the most salient features of each mean WISC-IV profile. These representative profiles are illustrated in Figures 7 through 9. Figure 10 depicts the similarities and differences among all three cluster profiles.
Table 8

*Kappa values and ICCs comparing prospective solutions generated during Stages 1 and 2 of the Initial Ward’s Cluster Analysis in Study 2*

<table>
<thead>
<tr>
<th>Ward’s Cluster Solution</th>
<th>Kappa</th>
<th>Cluster Number</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Cluster</td>
<td>.674***</td>
<td>1</td>
<td>.959***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>.939***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.948***</td>
</tr>
<tr>
<td>4-Cluster</td>
<td>.743***</td>
<td>1</td>
<td>.937***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>.989***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.983***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>.967***</td>
</tr>
</tbody>
</table>

Note. ICC = Intraclass Correlation Coefficient.

* p < .05. ** p < .01. *** p < .001.
Table 9

*Kappa Values and ICCs comparing Ward’s solutions to solutions derived using three alternative hierarchical clustering techniques in Study 2*

<table>
<thead>
<tr>
<th>Ward’s Cluster Solution</th>
<th>Kappa</th>
<th>Cluster Number</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AbG</td>
<td>AwG</td>
<td>CoL</td>
</tr>
<tr>
<td>3-Cluster</td>
<td>.380***</td>
<td>.945***</td>
<td>.972***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Cluster</td>
<td>.428***</td>
<td>.698***</td>
<td>.403***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


* p < .05. ** p < .01. *** p < .001.
Figure 7. Mean WISC-IV profile for Cluster 1 in Study 2 (Globally Low2).

Note: BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; PCm (Picture Completion); VOC = Vocabulary; COM = Comprehension; SIM = Similarities; IND = Information; DS = Digit Span; LNS = Letter Number Sequencing; COD = Coding; SS = Symbol Search.
Figure 8. Mean WISC-IV profile for Cluster 2 in Study 2 (Low VCI²).

Note: BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; PCm (Picture Completion); VOC = Vocabulary; COM = Comprehension; SIM = Similarities; IND = Information; DS = Digit Span; LNS = Letter Number Sequencing; COD = Coding; SS = Symbol Search.
Figure 9. Mean WISC-IV profile for Cluster 3 in Study 2 (Low WMI and PSI²).

Note: BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; PCm (Picture Completion); VOC = Vocabulary; COM = Comprehension; SIM = Similarities; INF = Information; DS = Digit Span; LNS = Letter Number Sequencing; COD = Coding; SS = Symbol Search.
Figure 10. Mean WISC-IV profiles for all clusters in Study 2.
Note: GL² = Globally Low²; LVCI² = Low Verbal Comprehension Index²; LWMPS² = Low Working Memory Index and Processing Speed Index²; BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; PCm = Picture Completion; VOC = Vocabulary; COM = Comprehension; SIM = Similarities; INF = Information; DS = Digit Span; LNS = Letter Number Sequencing; AR = Arithmetic; COD = Coding; SS = Symbol Search.
Given the similarities between the cluster-mean profiles generated in Study 1 and Study 2, with the exception of the superscript 2 (designating that the cluster is from the second study), the profile labels used in this study were identical to those used in the first study. Thus, Cluster 1 \((n = 51)\) was labeled Globally Low\(^2\) (GL\(^2\)), reflecting below average scores on all subtests with the exception of Picture Concepts, which was average. Cluster 2 \((n = 22)\) was designated Low VCI\(^2\) (LVCI\(^2\)), reflecting below average scores on subtests associated with the VCI and average scores on all other subtests. Cluster 3 \((n = 42)\) was labeled Low WMI and PSI\(^2\) (LWMPS\(^2\)) reflecting below average scores on those subtests associated with the WMI and PSI and average scores on the remaining subtests. In general, the Information subtest varied with core subtests that contribute to the VCI, Arithmetic varied with subtests associated with the WMI, and there was no profile characterized by subtest variation supporting the Gv (i.e., Block Design and Picture Completion) - Gf (Picture Concepts, Matrix Reasoning, and Arithmetic) distinction. In fact, Picture Completion varied with Matrix reasoning in all clusters, and with Block Design in two of the three clusters. Similar to the results of Study 1, the Picture Concepts subtest was among the highest score in every profile.

WISC-IV subtest- and index-score means and standard deviations for each of the three clusters are presented in Table 10. Age and gender distributions for each cluster are presented in Table 11. There were no significant differences in mean age as a function of cluster membership \((F [2, 112] = .217, p = .805)\).

### 3.2e External Validation

To assess the external validity of the final cluster solution, a MANOVA was employed to compare the three WISC-IV clusters on the basis of mean performance on
Table 10

WISC-IV Means and standard deviations for the final 3-cluster solution in Study 2

<table>
<thead>
<tr>
<th>WISC-IV Subtest/Index</th>
<th>Cluster 1 GL² (n = 51)</th>
<th>Cluster 2 LVCI² (n = 22)</th>
<th>Cluster 3 LWMPS² (n = 42)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Block Design</td>
<td>7.11 (2.09)</td>
<td>11.09 (2.50)</td>
<td>8.11 (2.31)</td>
</tr>
<tr>
<td>Picture Concepts</td>
<td>8.56 (2.49)</td>
<td>11.00 (2.63)</td>
<td>9.97 (2.57)</td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>7.19 (2.56)</td>
<td>9.68 (2.39)</td>
<td>8.54 (2.01)</td>
</tr>
<tr>
<td>Picture Completion</td>
<td>6.96 (2.46)</td>
<td>9.13 (2.51)</td>
<td>8.71 (2.57)</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>5.98 (1.44)</td>
<td>6.63 (1.49)</td>
<td>9.50 (1.61)</td>
</tr>
<tr>
<td>Comprehension</td>
<td>6.41 (1.79)</td>
<td>7.63 (1.55)</td>
<td>9.26 (2.10)</td>
</tr>
<tr>
<td>Similarities</td>
<td>5.64 (1.97)</td>
<td>7.63 (1.91)</td>
<td>9.23 (2.23)</td>
</tr>
<tr>
<td>Information</td>
<td>6.29 (1.62)</td>
<td>7.68 (1.46)</td>
<td>8.90 (1.64)</td>
</tr>
<tr>
<td>Digit Span</td>
<td>7.50 (2.42)</td>
<td>9.09 (2.26)</td>
<td>6.85 (1.98)</td>
</tr>
<tr>
<td>LN Sequencing</td>
<td>7.54 (2.24)</td>
<td>9.77 (1.41)</td>
<td>6.71 (2.70)</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>7.50 (1.79)</td>
<td>8.68 (2.21)</td>
<td>7.45 (2.06)</td>
</tr>
<tr>
<td>Coding</td>
<td>7.37 (1.82)</td>
<td>9.09 (2.38)</td>
<td>7.30 (2.03)</td>
</tr>
<tr>
<td>Symbol Search</td>
<td>7.86 (1.64)</td>
<td>9.63 (2.47)</td>
<td>7.45 (2.52)</td>
</tr>
<tr>
<td>PRI</td>
<td>84.43 (9.00)</td>
<td>101.90 (12.50)</td>
<td>92.07 (9.28)</td>
</tr>
<tr>
<td>VCI</td>
<td>77.03 (7.27)</td>
<td>84.54 (5.44)</td>
<td>96.47 (8.12)</td>
</tr>
<tr>
<td>WMI</td>
<td>85.74 (9.23)</td>
<td>96.59 (7.18)</td>
<td>82.00 (10.04)</td>
</tr>
<tr>
<td>PSI</td>
<td>86.27 (7.70)</td>
<td>96.45 (11.52)</td>
<td>85.69 (10.05)</td>
</tr>
<tr>
<td>FSIQ</td>
<td>79.00 (4.53)</td>
<td>92.36 (4.35)</td>
<td>86.90 (5.93)</td>
</tr>
</tbody>
</table>

Note. GL² = Globally Low²; LVCI² = Low VCI²; LWMPS² = Low WMI and PSI².
### Table 11

*Age and gender distribution for the final 3-cluster solution in Study 2*

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Gender</th>
<th>Age</th>
<th>M</th>
<th>F</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GL²</td>
<td></td>
<td></td>
<td>30</td>
<td>21</td>
<td>13</td>
<td>10</td>
<td>7</td>
<td>9</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>10.27 (1.94)</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td></td>
<td>59</td>
<td>41</td>
<td>25</td>
<td>20</td>
<td>14</td>
<td>18</td>
<td>4</td>
<td>15</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>---</td>
</tr>
<tr>
<td>LVCI²</td>
<td></td>
<td></td>
<td>14</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>10.41 (2.00)</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td></td>
<td>64</td>
<td>36</td>
<td>23</td>
<td>23</td>
<td>14</td>
<td>14</td>
<td>1</td>
<td>18</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>---</td>
</tr>
<tr>
<td>LWMPS²</td>
<td></td>
<td></td>
<td>31</td>
<td>11</td>
<td>15</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>10.08 (2.01)</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td></td>
<td>74</td>
<td>26</td>
<td>36</td>
<td>14</td>
<td>10</td>
<td>14</td>
<td>12</td>
<td>10</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>---</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>75</td>
<td>40</td>
<td>33</td>
<td>21</td>
<td>13</td>
<td>18</td>
<td>9</td>
<td>16</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>10.22 (1.97)</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td></td>
<td>65</td>
<td>35</td>
<td>28</td>
<td>18</td>
<td>11</td>
<td>15</td>
<td>9</td>
<td>14</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>---</td>
</tr>
</tbody>
</table>

Note. GL² = Globally Low²; LVCI² = Low VCI²; LWMPS² = Lower WMI and PSI². All percentages are rounded to the nearest whole number.
WIAT-II Word Reading, Spelling, and Numerical Operations subtests. The MANOVA was not significant on the basis of Pillai’s Trace (F [6, 222] = 1.13, p = .341), Wilks’ Lambda (F [6, 220] = 1.13, p = .345), Hotelling’s Trace (F [6, 218] = 1.12, p = .350), or Roy’s Largest Root (F [3, 111] = 1.587, p = .197), suggesting that the clusters did not differ significantly on WIAT-II variables. For exploratory purposes, univariate analyses were conducted on each WIAT-II subtest. Consistent with the results of the MANOVA, no significant differences were obtained. A listing of univariate F-scores, along with means and standard deviations for the WIAT-II subtests, can be found in Table 12. Figure 11 illustrates mean WIAT-II subtest performance for each of the WISC-IV clusters.
Table 12

*Means, standard deviations, and univariate F-scores for differences on WIAT-II subtests (standard scores) based on cluster membership in Study 2*

<table>
<thead>
<tr>
<th>WIAT-II Subtests</th>
<th>GL²</th>
<th>LVCI²</th>
<th>LWMPS²</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>73.45</td>
<td>76.00</td>
<td>77.23</td>
<td>.900</td>
<td>.41</td>
</tr>
<tr>
<td>$(SD)$</td>
<td>(12.24)</td>
<td>(15.30)</td>
<td>(14.75)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Spelling</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>76.45</td>
<td>79.72</td>
<td>79.26</td>
<td>1.05</td>
<td>.35</td>
</tr>
<tr>
<td>$(SD)$</td>
<td>(9.09)</td>
<td>(13.05)</td>
<td>(11.86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Numerical Operations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>77.37</td>
<td>83.27</td>
<td>76.83</td>
<td>2.26</td>
<td>.10</td>
</tr>
<tr>
<td>$(SD)$</td>
<td>(11.94)</td>
<td>(11.06)</td>
<td>(13.12)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. GL = Globally Low; LVCI = Low VCI; LWMPS = Lower WMI and PSI.
Figure 11. Mean WIAT-II profiles for the WISC-IV clusters derived in Study 2
Note: GL² = Globally Low², LVCI² = Low Verbal Comprehension Index²; LWMP² = Lower Working Memory Index and Processing Speed Index²; Num Ops = Numerical Operations.
CHAPTER 4
DISCUSSION

The purpose of the current investigation was to determine if reliable and valid WISC-IV patterns, consistent with those identified in taxonomic research using the WISC-R and WISC-III, could be empirically derived in a sample of children referred for psychoeducational assessment. To this end, two cluster analytic studies were conducted. In an attempt to derive a typology that would be relevant to the majority of WISC-IV users, only the ten core subtests were used to generate the clusters in Study 1. Conversely, the clusters in Study 2 were derived on the basis of the ten core subtests plus three supplemental subtests (Information, Picture Completion, and Arithmetic). These particular supplemental subtests were selected for inclusion to (1) facilitate comparison to clusters derived on the basis of previous WISC versions, and (2) determine if clusters representing subtest variations associated with Gv (Block Design and Picture Completion) and Gf (Picture Concepts and Matrix Reasoning) would emerge. Below is a summary of the results of each study and the implications of these findings. The methodological and theoretical limitations of this investigation are then discussed, followed by suggestions for future research.

4.1 Study 1

4.1a Summary of Findings

On the basis of previous research, it was hypothesized that six distinct clusters would be derived in Study 1. Although all expected clusters did not emerge (i.e., Low VCI and WMI; Low PRI and PSI; Low PRI), three hypothesized clusters were identified. Reflecting the salient features of each representative WISC-IV profile, the three clusters
were labeled: (1) Globally Low (GL); (2) Low VCI (LVCI); and (3) Lower WMI and PSI (LWMPS).

The GL cluster was characterized by relatively low scores on all WISC-IV subtests with a slight elevation of the Picture Concepts subtest. This cluster comprised 38.1% of the sample, making it the largest of the three identified clusters. With respect to its relation to previous research, variations of the GL profile have consistently been identified in empirical clustering research using the WISC-R and WISC-III (e.g., Holcomb et al., 1987; Snow et al., 1985; Borsuk et al., 2006; Donders, 1996; Glutting et al., 1994; Konold et al., 1999; Saunders et al., 2001; Waxman et al., 2003; Ward et al., 1999; Waxman & Casey, 2006), and in clinical inferential research exploring WISC-IV performance in children diagnosed with combined Reading, Written Expression, and Mathematics Disorder (Wechsler, 2003b). In referred and clinical samples, GL profiles have been empirically linked to pervasive neuropsychological and academic weaknesses (Holcomb et al., 1987; Glutting et al., 1994; Waxman & Casey, 2006; Wechsler, 2003b), and have been reported in older students evidencing Specific Language Impairment (Leonard, 2000).

The LVCI cluster, which comprised 35.1% of the sample, was characterized by relatively low scores on the Vocabulary, Comprehension, and Similarities subtests of the WISC-IV. Variations of this profile have been identified in WISC-R and WISC-III taxonomic studies involving normative (e.g., Konold et al., 1999), referred (e.g., Saunders et al., 2001; Waxman & Casey, 2006), and clinical samples (e.g., Bender & Golden, 1990; Borsuk et al., 2006; Glutting et al., 1994; Snow et al., 1985), and in

The LWMPS cluster comprised 26.7% of the sample and was characterized by a broadly average profile with a pattern of lower scores on WMI and PSI subtests within the context of a relative strength on PRI subtests, particularly Picture Concepts. This general pattern of performance (i.e., lower WMI and PSI) has been widely reported in the WISC-R and WISC-III taxonomic literature (e.g., Hale & Saxe, 1983; Holcomb et al., 1987; Packman et al., 1994; Saunders et al., 2001; Snow et al., 1985; Vance et al., 1978; Waxman et al., 2003; Waxman & Casey, 2006), and in WISC-IV research involving typically-developing children (e.g., Cheramie, Stafford, & Mire, 2008), as well as those with high functioning autism (Mayes and Calhoun, 2007c), combined reading and writing disabilities (Wechsler, 2003b, Weiss & Gabel, 2008), and ADHD with and without comorbid learning challenges (Mayes & Calhoun, 2007b; Saklofske, Weiss, Raiford, & Prifitera, 2006; Weiss & Gabel, 2008). The WISC-IV mean profile identified in this cluster supports the rationale for the development of the Global Ability Index (GAI). As previously described, the GAI is a composite score that is calculated on the basis of the VCI and PRI without the influence of the WMI and PSI. According to the WISC-IV developers, the GAI provides the best estimate of general cognitive ability in cases where working memory and/or processing speed weaknesses unduly affect FSIQ performance (Raiford et al., 2005).

As hypothesized, the typology derived in Study 1 was considered to be internally valid in that it remained stable across first and second stages of the initial cluster analysis, and was well-replicated across hierarchical methods and split-half samples. Also
consistent with the stated hypotheses, the GL cluster was found to be externally valid in the sense that it differed from the other two groups on measures not used in the cluster analysis. Specifically, this cluster yielded significantly lower mean scores on WIAT-II Word Reading, Spelling, and Numerical Operations subtests, with effect size estimates suggesting a clinically meaningful finding on the latter subtest. The LVCI and LWMPS clusters did not differ significantly on the external variables employed in this study.

4.1b Implications

Despite the manifold changes made to the WISC during its latest revision, it has been recommended that interpretation of the WISC-IV proceed in a manner similar to that of previous WISC editions (e.g., Berninger et al., 2005; Wechsler, 2003c). In order to assess the appropriateness of this recommendation, research is needed to compare the new WISC to its predecessors, particularly in populations of children for which the WISC-IV is most likely to be used (American Educational Research Association et al., 1999). To this end, the results of the current cluster analytic investigation will be considered within the context of the taxonomic literature involving the WISC-R and WISC-III.

In keeping with previous empirical clustering research, the results of Study 1 suggest that stable and reliable patterns of WISC-IV core subtest scores can in fact be derived using cluster analysis in children referred for psychoeducational assessment. As previously indicated, the mean WISC-IV profiles identified in this investigation are markedly similar to those reported in taxonomic research using earlier WISC editions. This suggests that, to some degree, the new WISC performs in a manner similar to that of its predecessors, at least in terms of identifying groups of children with global-,
comprehension-, and cognitive proficiency (i.e., combined working memory and processing speed) weaknesses. That being said, reflecting the substantive changes made to the WISC during its latest revision, the results of this investigation also lend credence to claims that the WISC-IV deviates considerably from its predecessors with respect to the cognitive processes being measured (Allen et al., 2010; Baron, 2005; Yeates and Donders, 2005). Demonstrating this, a number of subtest profiles frequently reported in empirical clustering studies involving the WISC-R and WISC-III failed to emerge in the current investigation.

In contrast to the stated hypotheses, but in keeping with WISC-IV factor analytic research and the concomitant structural changes made to this instrument, the current investigation did not derive clusters characterized by relative weaknesses on subtests historically associated with the VIQ (i.e., VCI and WMI subtests) or PIQ (PRI and PSI subtests). That these subtests did not systematically covary in the current study suggests that the goal of creating indexes measuring more discrete cognitive domains (Wechsler, 2003c) may have been achieved with the latest WISC revision. It has yet to be empirically established how this change has influenced the identification of children with conditions traditionally diagnosed, at least in part, on the basis of VIQ-PIQ discrepancies (e.g., Specific Language Disorders, Nonverbal Learning Disabilities). In any case, the results of this study suggest that profiles characterized by subtest fluctuations consistent with the former VIQ and PIQ may not emerge in the new WISC at the same rate as in previous editions. This is an important finding to consider when interpreting the results of the WISC-IV, particularly for seasoned clinicians who may be inclined to base their
interpretation of this instrument on “internalized profiles derived from their years of experience with earlier WISC versions” (Baron, 2005 p. 474).

On the basis of the extant literature, it was also expected that a cluster characterized by relatively circumscribed weaknesses on the PRI subtests would be derived. Contrary to this prediction, unlike taxonomic studies involving the WISC-R (e.g., Snow et al., 1985; Vance et al., 1978) and WISC-III (e.g., Saunders et al., 2001; Waxman & Casey, 2006), a low PRI group did not emerge in the present investigation. This finding is not entirely surprising considering that, relative to the other indexes, the PRI underwent the greatest amount of change during the latest WISC revision, with the current index ostensibly measuring skills untapped by its predecessor, the POI (Weiss, Saklofske, & Prifitera, 2005). In addition to the reduction of speed and motor demands on the PRI, according to Weiss et al. (2005), “the construct measured by this index has changed from primarily perceptual organization with some fluid reasoning in WISC-III to primarily fluid reasoning with some perceptual organization in WISC-IV” (pg. 74). That a low PRI group did not emerge in the current investigation implies that, compared to the POI of earlier WISC versions, the WISC-IV PRI may not be as sensitive to the nonverbal weaknesses observed in some children with persistent academic difficulties (e.g., Rourke, 1989). A similar conclusion was drawn in studies investigating WISC-IV profiles in children with high-functioning autism (Mayes & Calhoun, 2008), TBI (Allen et al., 2010; Donders & Janke, 2008), and ADHD (Mayes & Calhoun, 2006), arguing against the recommendation made in the WISC-IV manual that the PRI should be substituted for the analogous WISC-III score “in clinical decision-making and other situations where [this score was] previously required” (Wechsler, 2003c, p. 6).
Curiously, the Picture Concepts subtest represented the highest score in every cluster, failing to consistently vary in a predictable manner with the other PRI subtests. Unexpectedly high mean Picture Concepts scores have also been found in studies exploring WISC-IV performance in TBI samples (Donders, 2008; 2010) and combined psychoeducational and clinic-referred samples (P. R. Ricciardi, personal communication, December 4, 2007), lending credence to questions regarding the skills measured by this subtest (Donders, 2010; Kain, 2006). Perhaps these findings indicate a relative strength in categorical reasoning abilities, as would be assumed based on the description of this subtest in the WISC-IV manual (Wechsler, 2003c). Another possibility is that these findings demonstrate the sensitivity of the Picture Concepts subtest to children’s tendencies to compensate for neuropsychological weaknesses through the use of intact cognitive skills (Wells, 2005) and behavioral strategies (Taylor, 2007). These things considered, until the skills tapped by the Picture Concepts subtest in clinical and referred samples are more clearly delineated, care should be taken not to misinterpret a high Picture Concepts score, recognizing that this may represent a common finding in these children.

Consistent with the vast literature demonstrating the relationship between global intellectual functioning and achievement across all academic domains (e.g., Glutting, Watkins, Konold, & McDermott, 2006; Mayes and Calhoun, 2007), the GL cluster in this study scored significantly lower than the other two clusters on all three WIAT-II subtests. Even so, that the only clinically meaningful difference between the groups occurred on the Numerical Operations subtest provides some support for the notion that WISC-IV profile shape in addition to level may play a role in academic achievement patterns (Hale
et al., 2007). That is, based on effect size estimates and the results of the discriminant function analysis, the three clusters derived in this study performed more similarly on Word Reading and Spelling subtests than on the Numerical Operations subtest. Thus, consistent with what would be expected (although not specifically stated), the pattern of WIAT-II performance differed between groups such that the LVCI and LWMPS groups exhibited lower Word Reading and Spelling scores within the context of better Numerical Operations performance, whereas the GL group performed particularly poorly on the latter subtest. Thus, in addition to level of WISC-IV performance, information regarding pattern of performance may be helpful for clinicians interested in formulating hypotheses about individual strengths and weaknesses (Hale et al., 2007).

That the LVCI and LWMPS groups did not differ on the WIAT-II was expected given that these groups were hypothesized to perform similarly on this instrument, differing only from a globally low group and groups characterized by poor performance on PRI subtests in isolation and PRI and PSI subtests combined. The latter hypotheses could not be tested because low PRI/PSI groups did not emerge in this investigation. As will be discussed below, different external validity variables will be needed to determine the extent to which the LVCI and LWMPS clusters represent distinct subgroups evidencing clinically meaningful differences.

4.2 Study 2

4.2a Summary of Findings

Employing virtually the same methodology as in Study 1, three distinct clusters were empirically derived on the basis of WISC-IV core and supplemental (Picture Completion, Information, and Arithmetic) subtest scores. Consistent with the stated
hypotheses, the representative WISC-IV profiles associated with these clusters were markedly similar to those derived in Study 1. Reflecting this finding, they were labeled: (1) Globally Low\(^2\) (GL\(^2\)); (2) Low VCI\(^2\) (LVCI\(^2\)); and (3) Low WMI and PSI\(^2\) (LWMPS\(^2\)).

The GL\(^2\) cluster was characterized by relatively low scores on all WISC-IV subtests with the exception of Picture Concepts, which was slightly elevated. This cluster comprised 44.3% of the sample, again making this the largest of the identified clusters. In general, Picture Completion, Information, and Arithmetic varied with the other subtests in their indexes, supporting in a referred sample the factor structure reported in the WISC-IV manual (Wechsler, 2003c). Similar to the results of Study 1, however, Picture Concepts did not vary in a predictable manner with the other subtests in the PRI, raising questions about the skills tapped by this subtest.

The LVCI\(^2\) cluster was characterized primarily by relative weaknesses on subtests contributing to the VCI and comprised 19.1% of the sample. In comparison to the analogous cluster identified in Study 1, this cluster evidenced higher scores on the Picture Concepts, Similarities, and Block Design subtests, with the latter difference being the most profound. Generally speaking, consistent with the factor structure reported by the WISC-IV developers (Wechsler, 2003c), Information and Arithmetic varied with the other subtests in the VCI and WMI, respectively. Conversely, the PRI was somewhat divided, with Block Design and Picture Concepts evidencing similar scores, and Matrix Reasoning and Picture Completion evidencing similar scores.

The LWMPS\(^2\) group, which comprised 36.5% of the sample, was characterized by below average scores on subtests contributing to the WMI and PSI. Although the
defining features of this cluster were similar to those associated with the analogous cluster derived in Study 1 (i.e., low WMI and PSI), the two mean profiles differed in a number of ways. Relative to the LWMPs group from Study 1, this cluster exhibited markedly lower scores on all PRI subtests, failing to represent a relative strength in this profile. Additionally, the Similarities subtest was shifted down approximately one scaled score in this cluster, as were the constituent subtests of the WMI and PSI. With respect to the supplemental subtests, Picture Completion varied with Block Design and Matrix Reasoning; however, Picture Concepts was notably higher than the other PRI subtests. Information varied almost uniformly with the other VCI subtests, while Arithmetic was slightly elevated relative to Digit Span and Letter Number Sequencing.

As predicted, the typology described in this study was found to be stable and reliable. Thus, similar clusters were derived during hierarchical and k-means analyses, and across four hierarchical algorithms. In contrast, the external validation hypotheses were not supported. That is, the three clusters did not differ significantly with respect to academic achievement as measured by WIAT-II Word Reading, Spelling, and Numerical Operations subtests.

4.2b Implications

The representative profiles derived in this study bear a close resemblance to those identified in Study 1. That similar clusters emerged in different samples and with the inclusion of additional variables supports the internal validity of the three cluster solution identified in this investigation (Morris, Blashfield, & Satz, 1981).

As predicted, the Information subtest varied with the other VCI subtests in all three derived clusters. This finding supports the factor structure reported by the
WISC-IV developers (Wechsler, 2003c), arguing against novel groupings of the VCI subtests, such as Information and Vocabulary forming a Long-Term Memory Cluster and Information and Comprehension forming a General Information Cluster (Flanagan & Kaufman, 2004). Further, these results imply that the routine administration of the Information subtest may not be warranted, as it appears this subtest may not regularly provide information above and beyond that provided by the core subtests alone.

Contrary to the stated hypotheses, clusters characterized by subtest score variations reflecting Gv (Block Design and Picture Completion) and Gf (Matrix Reasoning, Picture Concepts, and Arithmetic) did not emerge in this study. This finding fails to support the confirmatory factor analytic results reported by Keith et al. (2006), and suggests that subtest score variations along these two dimensions may be relatively rare in children with persistent academic difficulties.

In keeping with the conclusions of Study 1, the results of Study 2 raise questions regarding the skills tapped by the PRI subtests. Consistent with the first study, the Picture Concepts subtest in Study 2 represented one of the highest scores in every cluster, failing to consistently vary with the other PRI subtests. Additionally, one cluster (LVCI²) evidenced a novel division of the PRI subtests such that Block Design and Picture Concepts varied together, and Matrix Reasoning and Picture Completion varied together. The findings from this investigation, in combination with research indicating marginal stability coefficients for the PRI (Ryan, Glass, & Bartels, 2010), suggest the need to exercise caution when interpreting this index, especially when it comes to the Picture Concepts subtest.
Because the clusters derived in this investigation did not differ significantly on WIAT-II subtests, the external validity of this typology remains unclear. Thus, as was the case in Study 1, more extensive external validation measures will be needed to fully understand the extent to which these groups differ on measures not used in the cluster analysis.

4.3 Limitations

It is important to consider the current investigation within the context of its methodological limitations. One limitation involves the potential variability associated with data collection. Although the psychoeducational assessments giving rise to these data were conducted by Registered Psychologists, Psychological Associates, and trained personnel under the supervision of the Chief Psychologist, there is still likely to be some degree of variability with regard to test selection-, administration-, and scoring- practices. Thus, while it is reasonable to assume that the basic standards of practice were upheld during data collection, individual differences in assessment procedures may have affected the results of this investigation.

Another possible limitation relates to sample characteristics. Although the samples used in this investigation were heterogeneous in the sense that they included students presenting with a wide range of academic difficulties, in some ways the heterogeneity of the samples were limited. First, all children involved in this investigation resided in the same geographical region (i.e., Essex County). This may impact the extent to which these findings can be applied to students outside of this general area. Second, the age-range of the samples were somewhat truncated. Because referrals for psychoeducational assessment are less frequently made for children in later
grades, there was limited representation of older students in this investigation. Reflecting this, approximately 90% of the children in both studies were under the age of 14 years. Children at the lower end of the age continuum were also under-represented in this investigation. In an attempt to maintain consistency with previous taxonomic research, students under the age of 8 years were excluded from analyses. This restricted age variability may have precluded the identification of age effects, and may negatively impact the degree to which the current results can be applied to children at both ends of the age spectrum. Third, in keeping with previous empirical clustering research, children with FSIQ scores under 70 and over 130 were excluded from the present investigation. Again, this truncated range likely limits the generalizability of these results, and may have reduced the number of clusters derived in this investigation. Finally, prior to being referred for psychological assessment, the students in this investigation underwent a series of preliminary interventions. That being said, these interventions were not part of a systematic RTI program, which is the current trend in school psychology (Hale et al., 2010). Such a program has been implemented at the GECDSB since this study was completed.

A number of limitations related to sample characteristics are unique to Study 2. Specifically, the sample employed in this study was relatively small. As samples become larger, less frequently occurring patterns have the opportunity to emerge as distinct clusters rather than being subsumed into other clusters. Thus, with a larger sample, representative profiles reflecting subtest variations consistent with Gv and Gf may have emerged. Also, the children in this sample were selected because they had been administered the supplemental subtests of interest. It is possible that supplemental
subtests were administered only to children for whom additional clinical information was required. Thus, this sample may have comprised children with more complicated academic or cognitive presentations compared to the sample employed in Study 1. Finally, although replication of cluster analytic findings using a subset of the original sample is considered an acceptable research design (Morris et al., 1981; Speece, 2003), ideally, Study 2 would have been conducted in an independent sample.

Although using retrospectively gathered data enables researchers to carry out studies that otherwise may not be possible, such investigations are constrained by the available data. In the case of the present work, the range of variables on hand to explore the external validity of the derived clusters was limited. As previously noted, the WIAT-II subtests employed in this investigation were inadequate for truly assessing whether clinically meaningful differences exist between groups. Similarly, follow up research including the remaining WISC-IV supplemental subtests (i.e., Cancellation and Word Reasoning) was not possible due to the limited number of participants to whom these tests had been administered.

Other limitations of the present investigation relate to the use of cluster analytic methodology. Despite painstaking attempts to ensure the stability and reliability of the derived typology, the fact remains that cluster analysis represents a relatively subjective research tool (Lange et al., 2002). Thus, although efforts were made to ensure that selections regarding the similarity coefficient, grouping algorithm, and association indexes followed conventional standards and were empirically driven, in the end, a somewhat subjective decision was made by the present author. Additionally, it is important to recognize that with the use of cluster analysis, all participants in a sample
are forced into clusters on the basis of relative similarity to other participants without consideration of similarity in an absolute sense (DeLuca et al., 1991; Everitt, 1984; Hair & Black, 2000). Thus, the clusters generated in this investigation likely include some individuals who bear only a minimal similarity to the mean profile derived for that cluster. Finally, in an attempt to simultaneously evaluate profile elevation and pattern, Squared Euclidean Distance was used as the measure of similarity in the current investigation. Although Squared Euclidean Distance is clearly the most commonly used similarity index in taxonomic research on the Wechsler scales, it has recently been argued that methodology that maximizes the influence of profile shape and minimizes the influence of profile magnitude may derive clusters that provide more meaningful information (Lange, 2007).

4.4 Future directions

Considering that this investigation represents the first attempt to empirically delineate patterns of performance using the WISC-IV, it will be necessary to evaluate the reliability and validity of these findings through replication and cross-validation. To this end, cluster analysis of WISC-IV data should be conducted on similar samples of children to determine the extent to which the same mean profiles emerge. Additional research will then be needed to address the limitations associated with the current investigation.

To enable less frequently occurring WISC-IV patterns to emerge, it is recommended that future studies consistently employ large samples. Additionally, to maximize the generalizability of findings, care should be taken to employ heterogeneous samples of children spanning the full age range and representing the full IQ spectrum. In
keeping with current trends in the field of school neuropsychology (Hale et al., 2010), the current investigation should be replicated in children for whom increasingly intensive and individualized evidence-based academic interventions have not been successful. This would provide a clearer picture of the WISC-IV patterns that emerge exclusively in children whose academic difficulties cannot be directly attributed to insufficient educational experiences.

With respect to the variables employed, future research is needed to explore the effects of including all supplemental subtests in the cluster analysis. This would provide information regarding the extent to which the two subtests not included in the current investigation contribute to the identification of novel clusters. Additionally, given that the external validation variables used in the current investigation were limited to a few highly correlated achievement subtests, more comprehensive measures are needed to adequately explore the differences among the clusters derived in this investigation. Because the WISC-IV is often used in the process of generating hypotheses regarding psychological strengths and weaknesses, it would be wise to seek external validation through measures tapping a wide range of neuropsychological abilities (e.g., language processing, visuoperceptual/visuomotor skills, learning and memory, attention, and executive functioning) and academic-skill sets (e.g., phonemic awareness, decoding skills, reading fluency, reading comprehension, math fluency, math reasoning, written expression), as was done by Waxman & Casey (2006). Considering the moderating and mediating variables that are likely to influence the manifestation of learning difficulties (Taylor, 2010), it would also be interesting to explore the manner in which the clusters differ with respect to environmental- (e.g., sociodemographic status, parent psychological
distress, home/school support), socioemotional- (e.g., anxiety, depression, personal adjustment), and behavioural- (e.g., impulsivity, hyperactivity) factors. Indeed, some of the most important information will ultimately come from external validation studies examining the relationship between cluster membership and response to intervention efforts (Lange et al., 2002). Considering the multiply-determined nature of learning difficulties (Taylor, 2010), perhaps single-case research designs would be most appropriate here (Braden & Kratochwill, 1997) as group models are limited by their tendency to obscure individual differences (Hale & Fiorello, 2004).

To offset the inherent limitations associated with cluster analysis, attempts should be made to replicate the current findings using alternative taxonomic methodology. One possibility would be to employ Profile Analysis via Multidimensional Scaling (PAMS). This relatively new empirical clustering technique represents a variation of the more traditional factor- analytic model, but possesses considerable advantages over Q-factor analysis (Kim, Davison, & Frisby, 2007). One advantage of the PAMS technique is that, unlike cluster analysis, this method does not force participants into clusters, but instead provides information regarding the extent to which each participant's profile resembles one of the prototypical profiles identified in the data (Kim, Frisby, & Davison, 2004). Thus, through the use of this technique, 'purer' clusters would be derived given that individuals with profiles bearing only minimal similarity to the representative profile could be identified and excluded from analyses. Further, in contrast to cluster analysis, PAMS permits both exploratory and confirmatory data analyses (Kim et al., 2007).

To address the recommendations made by Lange (2007), future cluster analytic research exploring WISC-IV patterns in children with academic difficulties should
employ correlation rather than distance coefficients as the index of similarity. According to Lange (2007) clusters derived exclusively on the basis of profile shape rather than elevation provide the most clinically meaningful information. This assumption has yet to be empirically tested, thereby inviting future investigation.

In light of the current findings, studies are needed to explore the skills tapped by the Picture Concepts subtest in children with persistent academic difficulties. One method for accomplishing this goal would be to conduct an exploratory factor analysis including Picture Concepts and a wide range of neuropsychological measures. Identifying the factor(s) on which Picture Concepts loads may provide valuable information regarding the skills measured by this subtest in academically-struggling students.

4.5 Conclusions

The results of this investigation generally support the factor structure reported in the WISC-IV manual and suggest that in some ways the WISC-IV performs similarly to its predecessors, and in other ways it performs differently. In students with persistent academic difficulties, the WISC-IV seems to identify patterns reflecting global-, verbal comprehension-, and cognitive proficiency- weaknesses in a manner similar to that of its predecessors. This suggests that the WISC-IV VC-, PS-, and WM- indexes may be tapping constructs similar to those tapped by analogous indexes on earlier WISC editions, supporting the claim that these scores can be interpreted in like fashion (e.g., Berninger et al., 2005). Unlike in previous taxonomic research, verbal (i.e., VIQ) and performance (i.e., PIQ) weaknesses were not identified as commonly-occurring patterns using the WISC-IV. Thus, clinicians should not expect to see these profiles at the same rate using
the WISC-IV as they did using the WISC-R or WISC-III. Similarly, the PRI appears to measure skills untapped by the POI, thereby arguing against the recommendation to simply substitute the former score for the latter score in clinical practice (e.g., Berninger et al., 2005). Finally, a relative strength on the Picture Concepts subtest appears to be a common finding in children with persistent academic difficulties. Thus, care should be taken when interpreting this subtest, as it is unclear what skills are being measured.

On a final note, it is important to recognize that the results of group research do not necessarily apply to all children. As such, the mean WISC-IV profiles described in this investigation are not likely to be found in every child referred for psychoeducational assessment. Thus, while findings such as these may inform clinical practice, individual case formulations must be made on the basis of informed clinical judgment that takes into consideration a wide range of variables and recognizes the complexity of the determinants of learning (Hale et al., 2010; Flanagan & Kaufman, 2009; Taylor, 2009).
REFERENCES


Cohen, J. C. (1959). The factorial structure of the WISC at ages 7-6, 10-6, and 13-6.


Individuals with Disabilities Education Improvement Act (IDEA) of 2004, PL 108-446, 20, USC §§ 1400 et seq.


Taylor, H. G. (2009, February). Neuropsychological assessment has a specific though limited role in identification and treatment of LD. Invited debate: Cognitive/neuropsychological assessment is critical for learning disabilities – or is it? Symposium conducted at the 37th Annual International Neuropsychological Society Meeting, Atlanta, GA.


Table A1. *Summary of empirical clustering studies using the WISC-R in Clinical and Referred Samples*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering Method</td>
<td>Cluster analysis</td>
<td>Q-type factor analysis</td>
<td>Cluster analysis</td>
<td>Cluster analysis</td>
<td>Q-type factor analysis</td>
<td>Cluster analysis</td>
<td>Cluster analysis</td>
</tr>
<tr>
<td>Sample</td>
<td>N = 106 Ss = LD</td>
<td>N = 104 Ss = RD</td>
<td>N = 111 Ss = LD, ADD, ADHD</td>
<td>N = 119 Ss = LD</td>
<td>N = 269 Ss = PE referrals</td>
<td>N = 57 Ss = LD</td>
<td>N = 59 Ss = LD</td>
</tr>
<tr>
<td>WISC-R Variables</td>
<td>Factor Indexes 12 subtests</td>
<td>Not Specified</td>
<td>10 core subtests</td>
<td>10 core subtests</td>
<td>Sim, Voe, Con, BD, OA, Cod</td>
<td>Not specified</td>
<td></td>
</tr>
<tr>
<td>Clusters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal Weaknesses</td>
<td>Cluster 2 - Low VCI/High POI</td>
<td>Factor 3 - Language Disability (Automatic)</td>
<td>Cluster 4 - Impaired Verbal</td>
<td>Type A - Low Verbal &amp; Coding subtests</td>
<td></td>
<td>Cluster 3 - Language Deficits</td>
<td></td>
</tr>
<tr>
<td>Visual-Spatial Weaknesses</td>
<td>Cluster 6 - Low POI</td>
<td>Factor 2 - Perceptual Organization</td>
<td>Cluster 3 - Impaired Spatial &amp; Attention</td>
<td>Type B - Low Performance subtests</td>
<td>Profile 3 - Low Performance subtests</td>
<td>Cluster 2 - Visual Deficits</td>
<td>Subtype I - Spatial, sequential, &amp; behavioral deficits</td>
</tr>
<tr>
<td>Attention/Sequencing Weaknesses</td>
<td>Cluster 3 - Low FDI</td>
<td>Factor 1 - Distractibility</td>
<td>Cluster 2 - Impaired Attention/Sequencing</td>
<td>Type C - ACID Pattern (Partial)</td>
<td>Profile 1 - Low Sequencing</td>
<td></td>
<td>Subtype IV - Spatial &amp; sequential deficits</td>
</tr>
<tr>
<td>Global Weaknesses</td>
<td>Cluster 1 - Low VCI, POI, &amp; FDI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Weaknesses</td>
<td>Cluster 5 - Average VCI, POI, &amp; FDI</td>
<td></td>
<td>Cluster 1 - Flat profile</td>
<td>Type F - At least average</td>
<td></td>
<td></td>
<td>Subtype III - Unremarkable profile</td>
</tr>
<tr>
<td>Other Profiles</td>
<td></td>
<td>Factor 5 - Behavioral Comprehension &amp; Coding</td>
<td></td>
<td>Profiles 2 &amp; 4 - Not clinically interpretable</td>
<td></td>
<td></td>
<td>Subtype II - Low Info, Arith, PA; High OA</td>
</tr>
</tbody>
</table>

* Analysis included additional clustering variable
Appendix B

Table B1. Summary of empirical clustering studies using the WISC-III in Clinical and Referred Samples

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering Method</td>
<td>Similarity (D⁹) to Konold et al. (1999) core subtest taxonomy</td>
<td>Cluster analysis</td>
<td>Q-type factor analysis</td>
<td>Cluster analysis</td>
<td>Cluster analysis</td>
</tr>
<tr>
<td>Sample</td>
<td>N = 585</td>
<td>N = 343</td>
<td>Saunders et al. (2001) sample</td>
<td>N = 201</td>
<td>N = 182</td>
</tr>
<tr>
<td></td>
<td>Ss = LD</td>
<td>Ss = Heterogeneous referred</td>
<td></td>
<td>Ss = LD</td>
<td>Ss = Heterogeneous referred</td>
</tr>
<tr>
<td>WISC-III Variables</td>
<td>10 core subtests</td>
<td>12 subtests (Mazes omitted)</td>
<td>12 subtests (Mazes omitted)</td>
<td>Factor Indexes</td>
<td>12 subtests (Mazes omitted)</td>
</tr>
<tr>
<td>Verbal Weaknesses</td>
<td>Profile 6 – Below average ability and PIQ&gt;VIQ</td>
<td>Cluster 2 – Deficient language abilities</td>
<td>Group 1 – Verbal processing deficits</td>
<td>Cluster 2 – Low VCI &amp; FDI with underachievement in reading &amp; writing</td>
<td>Group 5 – Verbal processing deficits</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cluster 3 – Deficient nonverbal abilities</td>
<td>Cluster 4 – Nonverbal processing deficits</td>
<td></td>
<td>Group 3 – Low visual spatial/processing skills</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cluster 4 – ACID pattern</td>
<td>Group 2 – SCAD pattern</td>
<td></td>
<td>Group 4 – ACID</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cluster 5 – Deficient working memory (low Digit Span and Coding)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention/Sequencing Weaknesses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Processing Speed Weaknesses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global Weaknesses</td>
<td>Profile 8 – Low ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Weaknesses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Profiles</td>
<td>Profiles 1, 2, 3, 4, 5, 7, 9 – Failed to meet reliability criteria</td>
<td>Cluster 6 – Deficits in tasks involving visual sequencing and language abilities</td>
<td>Group 3 – Visual sequencing and language deficits</td>
<td>Cluster 3 – Average with underachievement in writing</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group 5 – General processing deficits (with intact processing speed)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cluster 4 – Low average with PSI strength</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cluster 5 Low average</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Analysis included additional clustering variables.
Appendix C

VCI
- Similarities
- Vocabulary
- Comprehension
  - Information
  - Word Reasoning

PRI
- Block Design
- Picture Concepts
- Matrix Reasoning
  - Picture Completion

FSIQ

WMI
- Digit Span
- LN Sequencing
  - Arithmetic

PSI
- Coding
- Symbol Search
  - Cancellation

Figure C1. Relationship of the WISC-IV subtests to the Indexes and FSIQ
Note: Supplemental subtests are shown in italics. VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed
Adapted from Wechsler (2003c)
Appendix D

Description of selected WISC-IV subtests

Verbal Comprehension Subtests

Similarities
Requires the child to identify essential similarities between objects or concepts. Questions and responses are both oral. Total score reflects the number of correct responses (item scores range from 0 – 2).

Vocabulary
Requires the child to name pictures presented in the Stimulus Book (items 1 – 4) or define words presented orally and visually (items 5 – 36). Total score reflects the number of correct responses (item scores range from 0 – 2).

Comprehension
Requires the child to explain situations, actions, or activities on the basis of his or her understanding of general principles and social norms. Questions and responses are both oral. Total score reflects the number of correct responses (item scores range from 0 – 2).

Information*
Requires the child to answer questions related to a wide range of general knowledge topics including body parts, calendar information, historical figures, science, and geography. Questions and responses are both oral. Total score reflects the number of correct responses (item scores range from 0 – 1).

Perceptual Reasoning Subtests

Block Design
Requires the child to arrange dichromatic blocks to recreate geometric designs presented as constructed models or pictures as quickly as possible. Total score reflects speed and accuracy of block arrangements. Time bonus scores are calculated for the last six items (item scores range from 0 – 7).

Matrix Reasoning
Requires the child to select the missing portion of incomplete matrices from an array of five choices. Responses may be oral or gestural. Total score reflects the number of correct responses (item scores range from 0 – 1).

Picture Concepts
Requires the child to view two or three rows of common items and select one picture from each row to form a group with a common characteristic. Responses may be oral or gestural. Total score reflects number of correct responses (items scores range from 0 – 1).

Picture Completion*
Requires the child to identify important missing details of pictures within a specified time limit. Responses may be oral or gestural. Total score reflects number of correct responses (item scores range from 0 – 1).
Working Memory Subtests

Digit Span
Requires the child to repeat a series of orally presented digits verbatim or in reverse sequence. Total score reflects number of correct responses (item scores range from 0 – 1).

Letter Number Sequencing
The child is required to listen to a series of numbers and letters presented in a specified random order and orally recall the numbers in ascending order and the letters in alphabetical order. Total score reflects number of correct responses (item scores range from 0-1).

Arithmetic*
The child is required to mentally solve a series of orally presented arithmetic problems within a specified time limit. Responses are oral. Total score reflects number of correct responses (item scores range from 0 – 1).

Processing Speed Subtests

Symbol Search
The child is required to identify target symbol(s) within a visual search group as quickly as possible. Responses involve the manual checking of yes or no boxes. Total score reflects the number of correct responses within a specified time limit minus number of incorrect response (item scores range from 0 – 1).

Coding
The child is required to copy symbols that are paired with shapes or numbers using a key as quickly as possible. Total score reflects number of correct responses within a specified time limit (item scores range from 0 -1).

* Supplemental subtest
Descriptions adapted from Wechsler (2003c) and Sattler & Dumont (2004).
Appendix E

Description of selected WIAT-II subtests

Word Reading
The examinee is required to name letters of the alphabet, work with sounds within words, and read words aloud from a list. Accuracy of pronunciation (not comprehension) is scored. Total score reflects the number of correct responses (item scores range from 0 - 1).

Spelling
The examinee is required to write orally presented letters and sounds and spell dictated words presented within sentences. Homonyms are included in the word list to assess the examinee's ability to use contextual cues to facilitate accurate spelling. Total score reflects the number of correct responses (item scores range from 0 - 1).

Numerical Operations
The examinee is required to identify and write numbers, count, and solve increasingly complex written calculation problems and equations involving addition, subtraction, multiplication, and division. Use of pencil and paper is permitted. Total score reflects the number of correct responses (item scores range from 0 - 1).

Descriptions adapted from Wechsler (2002a).
Appendix F

<table>
<thead>
<tr>
<th>Stage</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>7</td>
<td>5.0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4</td>
<td>6.2</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>9</td>
<td>9.4</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>6</td>
<td>10.3</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>3</td>
<td>16.4</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>8</td>
<td>17.3</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>5</td>
<td>18.1</td>
</tr>
</tbody>
</table>

Disproportionate increase in sum of squares coefficient

Figure F1. Abbreviated example of an agglomeration schedule
Note: This hypothetical agglomeration schedule suggests the presence of a four cluster solution. Adapted from Norusis (2005).

Figure F2. Abbreviated example of a dendrogram
Note: This hypothetical dendrogram suggests the presence of a four cluster solution. Adapted from http://www.mathworks.com
Appendix G

Table G1

*Kappa value interpretation system used in present investigation*

<table>
<thead>
<tr>
<th>Kappa value</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0</td>
<td>Less than chance agreement</td>
</tr>
<tr>
<td>0.00 – 0.20</td>
<td>Slight agreement</td>
</tr>
<tr>
<td>0.21 – 0.40</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>0.41 – 0.60</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>0.61 – 0.80</td>
<td>Substantial agreement</td>
</tr>
<tr>
<td>0.81 – 1.00</td>
<td>Almost perfect agreement</td>
</tr>
</tbody>
</table>

Note. Adapted from Landis & Koch (1977).
Appendix H

Figure H1. WISC-IV profile of Multivariate Outlier 1 removed from analyses

Note: BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; VOC = Vocabulary; COM = Comprehension; SIM = Similarities; DS = Digit Span; LNS = Letter Number Sequencing; COD = Coding; SS = Symbol Search.
Figure H2. WISC-IV profile of multivariate outlier 2 removed from analyses

Note: BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; VOC = Vocabulary; COM = Comprehension; SIM = Similarities; DS = Digit Span; LNS = Letter Number Sequencing; COD = Coding; SS = Symbol Search.
Appendix H

Figure H3. WISC-IV profile of Multivariate Outlier 3 removed from analyses

Note: BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; VOC = Vocabulary; COM = Comprehension; SIM = Similarities; DS = Digit Span; LNS = Letter Number Sequencing; COD = Coding; SS = Symbol Search.
Figure H4. WISC-IV profile of Multivariate Outlier 4 removed from analyses
Note: BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; VOC = Vocabulary; COM = Comprehension; SIM = Similarities; DS = Digit Span; LNS = Letter Number Sequencing; COD = Coding; SS = Symbol Search.
Figure H5. WISC-IV Profile of Multivariate Outlier 5 removed from analyses

Note: BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; VOC = Vocabulary; COM = Comprehension; SIM = Similarities; DS = Digit Span; LNS = Letter Number Sequencing; COD = Coding; SS = Symbol Search.
VITA AUCTORIS

NAME: Corinne Rene Hale
PLACE OF BIRTH: Detroit, Michigan
YEAR OF BIRTH: 1971
EDUCATION: University of Michigan, Ann Arbor, Michigan
B.Sc. with High Distinction, Biopsychology 1994
University of Windsor, Windsor, Ontario
M.A. Clinical Neuropsychology 2000
University of Windsor, Windsor, Ontario
Ph.D. Clinical Neuropsychology 2010