A Cluster Analytic Study of the Wechsler Intelligence Test for Children-IV in Children Referred for Psychoeducational Assessment Due to Persistent Academic Difficulties

Corinne R. Hale¹, Joseph E. Casey¹,*, Philip W. R. Ricciardi²

¹Department of Psychology, University of Windsor, Windsor, Ontario, Canada
²Greater Essex County District School Board, Windsor, Ontario, Canada

*Corresponding author at: Department of Psychology, University of Windsor, Windsor, ON, Canada N9B 3P4. Tel.: +1-519-253-3000; Fax: +1-519-973-7021.
E-mail address: jecasey@uwindsor.ca (J.E. Casey).
Accepted 28 October 2013

Abstract

Wechsler Intelligence Test for Children-IV core subtest scores of 472 children were cluster analyzed to determine if reliable and valid subgroups would emerge. Three subgroups were identified. Clusters were reliable across different stages of the analysis as well as across algorithms and samples. With respect to external validity, the Globally Low cluster differed from the other two clusters on Wechsler Individual Achievement Test-II Word Reading, Numerical Operations, and Spelling subtests, whereas the latter two clusters did not differ from one another. The clusters derived have been identified in studies using previous WISC editions. Clusters characterized by poor performance on subtests historically associated with the VIQ (i.e., VCI + WMI) and PIQ (i.e., POI + PSI) did not emerge, nor did a cluster characterized by low scores on PRI subtests. Picture Concepts represented the highest subtest score in every cluster, failing to vary in a predictable manner with the other PRI subtests.

Keywords: Classification; Developmental disorders of scholastic skills; Intelligence tests; Learning disorders; Profile analysis; Subtypes

Introduction

Within the context of psychoeducational assessment, intelligence tests have historically been employed to rule out the presence of a broad-based intellectual disability or to identify a learning disability (LD) 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 (Meyer, 2000; Sternberg & Grigorenko, 2002; Stuebing et al., 2002). Reflecting this, at least in the USA, the latest re-authorization of the Individuals with Disabilities Education Act (IDEA, 2004) permits schools to abandon the use of the IQ-Achievement discrepancy model of LD identification in favor of a more empirically defensible approach that involves delineating the psychological profiles of children who fail to respond to evidence-based academic interventions (Smith, 2005). Nonetheless, 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, which can play a role in educational programming (e.g., Dombrowski et al., 2004; Hale, Fiorello, Kavanagh, Holdnack, & Aloe, 2007; Kavale, Holdnack, & Mostert, 2006).

The Wechsler Intelligence Test for Children (WISC) is the most widely used individually administered intelligence test worldwide (Kaufman, Flanagan, Alfonso, & Mascolo, 2006). The latest edition of this measure, the WISC-IV (Wechsler, 2003a) is structurally distinct from its predecessors and represents the most dramatic change ever made to this battery (Flanagan & Kaufman, 2009). 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, American Psychological Association, & National Council on...
Measurement in Education, 1999). Contrary to these Standards, however, the WISC-IV has not been adequately studied in children exhibiting persistent academic difficulties (Hale & Fiorello, 2004; Mayes & Calhoun, 2007). The present study contributes to 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. Over the years, numerous studies employing empirical clustering methodology (e.g., cluster analysis) with the WISC-R and WISC-III have reported distinct performance patterns in normative, clinical, and referred samples (e.g., Bodin, Pardini, Burns, & Stevens, 2009; Borsuk, Watkins, & Canivez, 2006; Donders, 1996; Holcomb, Hardesty, Adams, & Ponder, 1987; Konold, Glutting, McDermott, Kush, & Watkins, 1999; Snow, Cohen, & Holliman, 1985; Vance, Wallbrown, & Blaha, 1978; Waxman & Casey, 2006). Despite considerable methodological and interpretive variability across investigations, some consistencies in this research have emerged. Commonly reported in cluster-analytic studies involving clinical and referred children are subgroups characterized by subtest patterns reflecting weaknesses in (a) broad-based cognitive processing (e.g., Borsuk et al., 2006; Holcomb et al., 1987; McDermott, Kush, & Watkins, 1999; Snow, Cohen, & Holliman, 1985; Vance, Wallbrown, & Blaha, 1978; Waxman & Casey, 2006). Despite considerable methodological and interpretive variability across investigations, some consistencies in this research have emerged. Commonly reported in cluster-analytic studies involving clinical and referred children are subgroups characterized by subtest patterns reflecting weaknesses in (a) broad-based cognitive processing (e.g., Borsuk et al., 2006; Holcomb et al., 1987; McDermott, Kush, & Watkins, 1999; Snow, Cohen, & Holliman, 1985; Vance, Wallbrown, & Blaha, 1978; Waxman & Casey, 2006). Despite considerable methodological and interpretive variability across investigations, some consistencies in this research have emerged. Commonly reported in cluster-analytic studies involving clinical and referred children are subgroups characterized by subtest patterns reflecting weaknesses in (a) broad-based cognitive processing (e.g., Borsuk et al., 2006; Holcomb et al., 1987; McDermott, Kush, & Watkins, 1999; Snow, Cohen, & Holliman, 1985; Vance, Wallbrown, & Blaha, 1978; Waxman & Casey, 2006).

In contrast to the numerous cluster analytic studies using the WISC-R and WISC-III, no study has been published using empirical clustering methodology to delineate patterns of performance in children referred for assessment using the WISC-IV. The current investigation was designed to fill this gap in the scientific literature. To this end, a cluster analytic study was conducted using WISC-IV subtest scores obtained from a large heterogeneous sample of children referred for psychoeducational assessment due to persistent academic difficulties. (The term “persistent academic difficulties” is used in this manuscript to describe conditions in which significant academic weaknesses persist despite adequate educational opportunities including the implementation of individualized programming accommodations and in the absence of a broad-based intellectual disability.) Toward the goal of developing a typology that would be relevant to the majority of WISC-IV users, this study was conducted using only the 10 core WISC-IV subtests.

It was anticipated that reliable and valid clusters would emerge in this investigation and that the mean WISC-IV profile patterns associated with these clusters would be consistent with the factor structure of the WISC-IV and with the results of empirical clustering studies based on previous WISC editions. Thus, the following mean WISC-IV subtest patterns were expected to emerge based on studies utilizing previous versions of the WISC: (a) relative weaknesses on subtests associated with the VCI; (b) relative weaknesses on subtests associated with the VCI and WMI (i.e., similar to the VIQ in previous WISC editions); (c) relative weaknesses on subtests associated with the PRI and PSI (i.e., similar to the PIQ in previous WISC editions). It was further predicted that some of the derived subgroups would differ on Wechsler Individual Achievement Test (WIAT)-II Word Reading, Spelling, and Numerical Operations subtests based on the findings from previous studies demonstrating a relationship between WISC patterns and academic achievement profiles (e.g., Rourke, Young, & Flewelling, 1971; Ward et al., 1999; Waxman & Casey, 2006; the “Special Group Studies” section of the Technical and Interpretive Manual [Wechsler, 2003b]).

Specifically, a cluster characterized by low scores across all subtests (i.e., globally low [GL]) would be expected to perform poorly relative to the other clusters on all three WIAT-II subtests. Clusters characterized by low scores on subtests contributing to the VCI and WMI (i.e., low VIC; low VCI [LVCI] and WMI; low WMI and PSI [LWMPS]) 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 and PSI clusters). Finally, clusters characterized by low scores on subtests contributing to the PRI and PSI would be expected to exhibit lower mean scores on the Numerical Operations subtest compared with those clusters characterized by low scores on VCI and WMI subtests.

**Method**

**Participants**

Participants included in this study were part of a larger data set originally collected by the psychology staff of a district school board located in southern Ontario in response to referrals for psychoeducational assessment. 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 behavioral concerns \((N = 56; 11.9\%)\). No child was included if he or she had only a behavioral or socioemotional problem. Referrals were initiated by the school-based learning support team, which usually included the classroom teacher, learning support teacher, and the school’s principal or vice-principal. 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 subtests \(\text{[Wechsler, 2002]}\). All participants were free of uncorrected sensory impairments, significant medical histories, psychiatric conditions, and pervasive developmental disorders. Children previously diagnosed with Attention-Deficit/Hyperactivity Disorder \(\text{(ADHD)}\) were included in the study \((n = 17)\). Five of the original 477 participants were identified as multivariate outliers \(\text{(Mahalanobis Distance; } p < .001)\) and were excluded from the analysis. The final sample comprised 472 participants ranging in age from 8 years, 0 months to 16 years, 11 months. To assess the replicability of the cluster solution, following the initial cluster analysis, the sample was randomly split in half. Descriptive data for the full sample and both split samples are presented in Table 1.

### Measures

The measures employed in this investigation were administered and scored by trained examiners employed by the school board under the direction of the Supervisor for Psychological Services between the years of 2004 and 2008. Scaled scores from the 10 core \(WISC-IV\) subtests were used to derive representative clusters. To determine the extent to which the derived clusters differed on measures of academic achievement \(i.e., \text{external validation}\), the subgroups were compared on the basis of \(WIAT-II\) subtest standard score performance. Canadian age-based norms were used for both \(WISC\) tests. To provide a means for identifying achievement patterns in the three fundamental academic domains, Word Reading, Spelling, and Numerical Operations subtests were selected for inclusion in this study as these were the subtests most consistently administered.

### Procedures

**Initial cluster analysis.** Considering that data input order can affect the results of hierarchical cluster analyses due to potential ties in similarity coefficients \(\text{(Podani, 1997)}\), 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 \(\text{(van der Kloot, Spaans, & Heiser, 2005)}\). 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 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. First, a hierarchical cluster analysis using Ward’s method with squared Euclidean distance was conducted to estimate the number of clusters present in the data. Second, using the centroids from the first stage as initial seeds, \(K\)-means iterative partitioning was conducted to correct fusion errors. For all suggested solutions, mean subtest scores for each derived cluster were plotted and visually inspected for clinical interpretability, an important consideration in empirical clustering research \(\text{(Kamphaus, DiStefano, & Lease, 2003)}\). Those cluster solutions characterized by interpretable profiles were forwarded for reliability analyses.

**Reliability.** The reliability procedures are depicted in Fig. 1.

**Hierarchical and \(K\)-means reliability.** Morris, Blashfield, and Satz \(\text{(1981)}\) contended that the number of cases changing cluster membership from the first stage \(\text{(hierarchical)}\) solution to the second stage \(\text{(K-means)}\) solution reflects an index of cluster “stability” \(i.e., \text{reliability}\). 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 random effects intraclass correlation coefficients \(\text{(ICCs)}\) were used to compare the solutions in terms of membership agreement and profile similarity, respectively. All reliable solutions were forwarded for additional analyses.
Multiple-method reliability. Multiple-method stability 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 forwarded for split-half reliability analyses.

Split-half reliability. To determine the extent to which the derived cluster solutions 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 standard score performances 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. Effect sizes were calculated using Cohen’s $d$ and were interpreted on the basis of suggestions made by Cohen (1988): $d = 0.20$ (small effect); $d = 0.50$ (medium effect); $d = 0.80$ (large effect).

Results

Initial Cluster Analyses

A review of the agglomeration schedule coefficients, pseudo-$F$ statistics, and dendrogram from the initial Ward’s analysis supported two-, three-, five-, and seven-cluster solutions. Following $K$-means analysis, it was determined that two-, three-, and
five-cluster solutions were most interpretable. The two-cluster solution was excluded from further analyses due to limited subtest scatter (i.e., clusters represented low and high performance exclusively). In light of these findings, the three- and five-cluster solutions were subjected to reliability analyses.

**Reliability Analyses**

**Hierarchical and K-means reliability.** The reliability of the three- and five-cluster solutions were evaluated by comparing the profiles derived from the initial Ward’s analysis (stage 1) to those derived from the K-means analysis (stage 2). Significant ICCs were found for all clusters in both solutions ($p < .05$). Kappa values were also significant ($p < .01$). Both solutions were deemed reasonably stable and were sent on for multiple-method reliability assessment.

**Multiple-method reliability.** When the prospective solutions were subjected to three additional hierarchical clustering algorithms, followed by a K-means pass through the data, good agreement was obtained for both solutions. Kappa values for both solutions were significant ($p < .001$) and suggested Fair to Almost Perfect agreement in the three-cluster solution, and Moderate agreement in the five-cluster solution (based on a Kappa interpretive system suggested by Landis and Koch, 1977). All ICCs in the three-cluster solution were significant ($p < .01$); conversely, a non-significant ICC was found for one profile within the five-cluster solution. Both solutions were deemed adequately replicable and were subjected to split-half reliability analyses.

**Split-half reliability.** The split-half profiles associated with the three-cluster solution had good visual agreement, and all ICCs were significant ($p < .05$). Conversely, the split-half profiles from the five-cluster solution were difficult to match, and two of the ICCs were non-significant. Based on these findings, the three-cluster solution was considered representative of the data and was selected as the final cluster solution.

**Assumption of non-multicollinearity.** Multicollinearity among variables can affect the results of a cluster analysis, potentially undermining the ability to draw meaningful conclusions from the final cluster solution (Hair & Black, 2000). To explore the possibility of multicollinearity within the data, a multiple regression analysis was conducted using the 10 core WISC-IV subtests as independent variables and participant age as the dependent variable. A review of the correlation matrix of the predictor variables did not suggest the presence of multicollinearity, as none of the correlations exceeded .80 (Field, 2009). Correlations ranged from .104 to .663. To further explore the possibility of multicollinearity among the subtests, a review of the Variance Inflation Factors (VIFs) was conducted. Non-multicollinearity is suggested when the maximum VIF is less than 10.0 (e.g., Field, 2009; Marquardt, 1970). VIFs in the current investigation ranged from 1.32 to 2.30, all substantially below 10 (mean VIF = 1.57). Together, the diagnostics indicated a lack of multicollinearity among the clustering variables.

**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 profiles are illustrated in Figs. 2–4.

![Mean WISC-IV profile for Cluster 1 (GL). 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.](image-url)
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 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 the VCI subtests. As such, this cluster was designated LVCI. Cluster 3 ($n=126$) was labeled LWMPS to reflect a generally average profile characterized by relatively low scores on subtests contributing to the WMI and PSI. Unexpectedly, the Picture Concepts subtest represented the highest score in all three clusters. There were no significant differences in mean age as a function of cluster membership—$F(2, 469)=0.111, p=.895$.

**External Validation**

The external validity of the final solution was assessed by determining the extent to which the derived clusters differed 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—$F(6, 936)=14.74, p<.001$.

Three univariate ANOVAs were computed to assess the significance of each WIAT-II variable. All ANOVAs were significant ($p<.001$), indicating that the empirically derived WISC-IV clusters differed on all three WIAT-II subtests. Fig. 5 illustrates the 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 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 estimates are presented in Table 2.
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

---

**Table 2.** Post hoc analyses (Games–Howell Test) and effect size estimates (Cohen’s $d$) for mean differences between WISC-IV clusters on WIAT-II subtests

<table>
<thead>
<tr>
<th>Cluster</th>
<th>GL</th>
<th>LVCI</th>
<th>LWMPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>WIAT-II Word Reading</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GL</td>
<td>—</td>
<td>3.97*</td>
<td>7.07***</td>
</tr>
<tr>
<td></td>
<td>$d$ 0.28</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>LVCI</td>
<td>—</td>
<td>—</td>
<td>3.09</td>
</tr>
<tr>
<td></td>
<td>$d$ 0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WIAT-II spelling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GL</td>
<td>—</td>
<td>3.82*</td>
<td>4.93**</td>
</tr>
<tr>
<td></td>
<td>$d$ 0.31</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>LVCI</td>
<td>—</td>
<td>—</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>$d$ 0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WIAT-II numerical operations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GL</td>
<td>—</td>
<td>9.43***</td>
<td>11.84***</td>
</tr>
<tr>
<td></td>
<td>$d$ 0.80</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>LVCI</td>
<td>—</td>
<td>—</td>
<td>–2.40</td>
</tr>
<tr>
<td></td>
<td>$d$ 0.09</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** GL = Globally Low; LVCI = Low VCI; LWMPS = Lower WMI and PSI. Cohen’s $d$ values in bold denote large effect sizes.

* $p < .05.$
** $p < .01.$
*** $p < .001.$
Clusters 1 (GL) and 2 (LVCI; $d = 0.80$) and between Clusters 1 and 3 (LWMPS; $d = 0.97$). 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’ Lambda = 0.828, $p < .001$). 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.

Discussion

The results suggest that reliable patterns of WISC-IV core subtest scores can be derived using cluster analysis in children referred for psychoeducational assessment. The mean WISC-IV profiles identified in this investigation are markedly similar to those reported in taxonomic research using earlier WISC editions, although not all previously identified profiles emerged. Replicated were profiles characterized by global, verbal comprehension, and combined working memory and processing speed weaknesses. Reflecting the substantive changes made to the WISC during its latest revision, the results of this investigation lend credence to claims that the WISC-IV deviates considerably from its predecessors with respect to the cognitive processes being measured (Allen, Thaler, Donohue, & Mayfield, 2010; Yeates and Donders, 2005). 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 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, 2003b) were achieved with the latest WISC revision in a referred sample. 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).

Unlike taxonomic studies involving the WISC-R (e.g., Snow et al., 1985) and WISC-III (e.g., Waxman & Casey, 2006), a low PRI or “non-verbal” 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 ostensibly measuring skills untapped by its predecessor, the Perceptual Organization Index (POI; Weiss, Saklofske, & Prifitera, 2005). In addition to the reduction in speed and motor demands on the PRI, “the construct measured by this index has changed from primarily perceptual organization with some fluid reasoning . . .to primarily fluid reasoning with some perceptual organization in WISC-IV” (Weiss et al., 2005, p. 74). That a low PRI group did not emerge in the current investigation implies that, compared with 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., those with Nonverbal Learning Disorder; Casey, 2012). A similar conclusion was drawn in studies investigating WISC-IV profiles in children with high-functioning autism (Mayes & Calhoun, 2008), Traumatic Brain Injury (TBI) (Allen et al., 2010; Donders & Janke, 2008), and in children with ADHD (Mayes & Calhoun, 2006).

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. A high mean Picture Concepts score was also reported by Donders and Janke (2008) exploring WISC-IV performance in a TBI sample. This lends credence to questions regarding the skills measured by this subtest (Kain, 2006). Until the skills tapped by the Picture Concepts subtest in clinical and referred samples are more clearly delineated, care should be taken when interpreting the Picture Concepts score, recognizing that a relatively higher score on this subtest may represent a common finding in referred children.

Consistent with the vast literature demonstrating a relationship between global intellectual functioning and academic achievement generally (e.g., Glutting, Watkins, Konold, & McDermott, 2006; Mayes & Calhoun, 2007), the GL cluster in this study was characterized by significantly lower scores than the other two clusters on all three WIAT-II subtests. Even so, from the standpoint of effect size the only meaningful difference between the groups occurred on the Numerical Operations subtest, supporting the notion that WISC-IV profile shape in addition to level plays a role in academic achievement patterns. That is, the three clusters derived in this study performed more similarly on Word Reading and Spelling subtests than on the Numerical Operations subtest. Thus, 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).

It is important to consider the current investigation within the context of its methodological limitations. A possible limitation relates to sample characteristics. Although the sample used in this investigation was heterogeneous in that it included students presenting with a wide range of academic difficulties, in some ways the heterogeneity of the sample was limited. First, the age
range of the sample was 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 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. Second, in keeping with previous empirical clustering research, children with full scale IQ 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.

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. Because the full range of WIAT-II subtests were not routinely administered, the three WIAT-II subtests used in this investigation may have been inadequate for truly assessing whether clinically meaningful differences exist between groups. Similarly, follow-up analysis 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 efforts to ensure the reliability and validity of the derived typology, the fact remains that cluster analysis represents a relatively subjective research tool (Lange, Iverson, Senior, & Chelune, 2002). 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 is required by the researchers. 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 (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).

Considering that this investigation represents the first attempt to empirically delineate patterns of performance using the WISC-IV, it is 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. Studies including larger samples and the combination of WISC-IV core and supplemental subtests might result in less frequently occurring WISC-IV patterns to emerge.

Given that the external validation variables used in the current investigation were limited to achievement subtests that are highly correlated with the WISC-IV, possibly leading to conflation, non-academic measures are needed to adequately explore the external validation of the clusters derived in this investigation. Considerations could include environmental factors (e.g., sociodemographic status, parent psychological distress, home/school support), socioemotional factors (e.g., anxiety, depression, personal adjustment), and behavioral factors (e.g., impulsivity, hyperactivity). 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).

To offset the inherent limitations associated with cluster analysis, attempts should be made to replicate the current findings using alternative taxonomic methodologies. 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, one of which 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. 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.

On a final note, it is important to recognize that the results of group research do not apply to all children. As such, any one of the mean WISC-IV profiles described in this investigation will not be found in every child referred for psychoeducational assessment. 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 (Flanagan & Kaufman, 2009; Hale et al., 2010).

Acknowledgements

This manuscript is based on a dissertation submitted by the first author to the Department of Psychology at the University of Windsor in partial fulfillment of the requirements for the PhD degree. Parts of this project were presented at the 38th Annual Meeting of the International Neuropsychological Society, Acapulco, Mexico, February 2010, and the 39th Annual Meeting of the International Neuropsychological Society, Boston, Massachusetts, February 2011. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

References


