

Spearman's Law of Diminishing Returns (SLODR): Examining Effects at the Level of Prediction

Ryan J. McGill¹

Abstract

The current study presents the results of an incremental validity investigation using a hierarchical multiple regression design, examining standard score data from the standardization sample ($N = 4,722$) of the Woodcock-Johnson III Tests of Cognitive Abilities (WJ-III COG) and Achievement (WJ-III ACH). Specifically, this study examined whether or not the differentiation hypothesis known as Spearman's law of diminishing returns (SLODR) impacts the incremental predictive validity of the Cattell-Horn-Carroll (CHC) broad cluster scores on the WJ-III COG. Consistent with the SLODR hypothesis, differential patterns of prediction were observed across ability classes (e.g., Below Average, Average, Above Average). Specifically, the general factor score (GIA) accounted for significantly less portions of achievement variance in the Above Average group ($R^2 = .11$ to $.20$) when compared to estimates for the Below Average group ($R^2 = .24$ to $.32$). Secondary chi-square goodness of fit tests provided additional evidence of predictive variance across ability classes. The results of this study indicate that the effects of SLODR must be accounted for when interpreting the results of the contemporary cognitive ability tests in clinical settings.

Keywords: SLODR, incremental validity, CHC, general intelligence

1. Introduction

Buoyed by the statistical developments of his cousin Karl Pearson, British statistician Charles Spearman proposed the first unified psychometric theory of cognitive abilities with the publication of the paper "General Intelligence, Objectively Determined and Measured" in 1904. Spearman (1904) determined that different cognitive measures tended to correlate, a phenomenon he referred to as positive manifold (Thurstone, 1947). In his analysis he found that 62.9% of the total variance between all of the tests was accounted for by a single factor which he identified as the general factor or g . The remaining 37.1% of variance was attributed to specific factors unique to the individual tests. According to Schneider and Flanagan (2015), the resulting two-factor theory "assumes that, aside from error, [cognitive] test scores are determined by two kinds of ability: general and specific" (p. 321). Whereas g exerts its influence on all types of cognitive tasks (i.e., indifference of the indicator), specific abilities only influence measures that sample more precise cognitive skills. Critics of Spearman's theory (e.g., Thompson, 1916, Thorndike, Lay, & Dean, 1909) soon emerged, suggesting that g was a statistical artifact and thus did not represent a unitary intelligence factor. Although a more nuanced discussion regarding the nature of the g -factor is beyond the scope of the present study, it should be noted that despite these and subsequent critiques of the construct within the technical literature (e.g., Horn, 1991), strong empirical support for a general ability dimension has consistently been found in factor analytic research over the course of the last century (e.g., Carroll, 1993; 2003; Gignac, 2014; Golay et al., 2013; Johnson & Bouchard, 2005). As a result, the most recent iteration of the widely referenced Cattell-Horn-Carroll model of intelligence (CHC; Schneider & McGrew, 2012) includes g at the apex of a multilayered matrix of empirically supported broad and specific cognitive abilities. According to McGrew (2009), most comprehensive intelligence batteries implicitly or explicitly acknowledge the influence of the CHC model within their test design framework.

¹ Department of Psychology and Philosophy, Texas Woman's University, P.O. Box 425470, Denton, TX 76204.
E-Mail: rmcgill@twu.edu

Despite their widespread use, a long-standing controversy remains about how to best interpret cognitive measures. Interpretation of intelligence tests involves making inferences about an individual based upon the scores that are obtained from a standardized administration of a particular test (Canivez, 2013). Due to the fact that contemporary cognitive measures appraise performance at multiple levels (e.g., subtest, composite, full-scale); practitioners are provided with a multitude of options for making inferences regarding individual performance. Because intelligence tests are utilized primarily for predicting important outcomes such as achievement and occupational attainment, the cognitive variables that provide the most optimal weighting for prediction are of greatest importance (Gottfredson, 1997; Wiggins, 1973). Simultaneous interpretation of scores at all multiple levels ignores the fact that the variance in lower-order factor and composite measures is multidimensional (Reise, Bonifay, & Haviland, 2013). That is, some variance is apportioned to the higher-order general ability dimension (as represented in contemporary test batteries as the full-scale IQ score), some to the first-order factor (specific ability), and some is attributable to measurement error (Carroll, 1995). As a consequence, interpreting all obtained scores may result in redundancy because clinicians are not able to disaggregate disparate sources of score variance at the level of the individual. Thus, consideration of the incremental validity of lower-order scores beyond that of higher-order scores is important when interpreting tests at multiple levels as is advocated for most contemporary cognitive measures (Canivez et al., 2014).

1.1 Incremental Validity

Incremental validity refers to the “extent to which a measure adds to the prediction of a criterion beyond what can be predicted with other data” (Hunsley, 2003, p. 443). Incremental validity is rooted in the scientific law of parsimony which states “what can be explained by fewer principles is needlessly explained by more” (Jones, 1952, p. 620). When applied to intelligence tests, interpretation of the full-scale IQ score is more parsimonious than interpretation at the lower-order factor or composite level. Thus, to interpret primarily at lower-order levels, practitioners should have a compelling reason for doing so (e.g., lower-order measures should account for non-trivial portions of criterion variance beyond that accounted for already by the full-scale score). Hierarchical multiple regression analysis is a well-established statistical procedure for assessing incremental validity in the social sciences and has been successfully applied in the technical literature in studies utilizing cognitive-achievement assessment data (Canivez, 2013). In this procedure, the full-scale IQ score is entered first into a regression equation followed by the lower-order factor or composite scores to predict a criterion achievement variable. This entry technique allows for the unique predictive effects of the lower-order scores to be assessed while controlling for the effects of the full-scale IQ score and operates conceptually in very much the same way as the Schmid and Leiman technique (1957) for residualizing variance in exploratory factor analysis. Incremental validity studies (e.g., Canivez et al., 2014; Glutting, Watkins, Konold, & McDermott, 2006; McGill & Busse, 2014; McGill, 2015) have consistently demonstrated that the omnibus full-scale score on intelligence tests accounts for most of the reliable achievement variance in prediction models and that little additional incremental variance is accounted for by lower-order cognitive scores after controlling for the effects of the general factor. As a result, Canivez (2013) encouraged users of IQ tests to limit most, if not all, of their interpretation to the omnibus full-scale IQ score. However, this suggestion assumes that the contribution of g is relatively consistent across individuals (Wolfe, 1940). Unfortunately, previous incremental validity investigations have failed to account for psychometric mechanisms, such as Spearman’s law of diminishing returns, that can potentially moderate the influence of g across individuals.

1.2 Spearman’s Law of Diminishing Returns (SLODR)

First described by Spearman, the law of diminishing returns (SLODR), also known as the differential hypothesis, suggests that the contribution of g decreases as the level of ability increases. Initial evidence for SLODR was provided by Spearman (1927) who compared correlation matrices between 78 “normal” children and 22 “defective” children and found that the mean correlation coefficient between tasks was .47 for the normal sample and .78 for the sample of children with below average intellectual abilities. From these data he concluded that “the more ‘energy’ a person has available already, the less advantage accrues to his ability from further increments of it” (p. 219). SLODR is less understood than other factors influencing g due to the fact that it has only begun to be studied relatively recently within the scientific literature. The first large scale empirical study ($N = 4,080$) of the effects of SLODR was conducted by Detterman and Daniel (1989).

They divided a sample of individuals who had been administered the Wechsler Scale into five ability groups and found that inter-correlations between subtests decreased across the low to high ability groups. In a more sophisticated large-scale study ($N = 10,535$), Deary and colleagues (1996) demonstrated differential g loadings across ability groups along with evidence of an interaction effect across ages. In order to further examine the tenability of the differentiation hypothesis, Coyle and Rindermann (2013) examined national performance across multiple international standardized tests and found that g loadings on the measures were lower for higher ability nations. As a result, the authors suggested that the predictive validity of cognitive tests may be lower for higher ability nations. Additionally, several recent investigations (e.g., Reynolds, Hajovsky, Nilleksela, & Keith, 2011; Reynolds, Keith, & Beretvas, 2010) using factor mixture modeling have provided latent variable evidence for differential loadings across ability groups on contemporary intelligence tests. On the basis of these investigations, Reynolds (2013) encouraged practitioners to interpret lower-order ability profiles in place of general factor scores for individuals with higher estimated levels of general intelligence. Despite these results, the evidence-base for the tenability of the SLODR hypothesis is far from conclusive. Several empirical investigations have challenged the law of diminishing returns (e.g., Arden & Plomin, 2007; Facon, 2004; Fogarty & Stankov, 1995; Hartmann & Reuter, 2006; Jensen, 2003), suggesting that the effects of SLODR are not consistent across the lifespan as well as not expressed equally across all types of cognitive measures. Using confirmatory factor analysis (CFA) to test the invariance of latent factor structures for lower and higher ability groups on the Kaufman Assessment Battery for Children-Second Edition (KABC-II; Kaufman & Kaufman, 2004), Reynolds and Keith (2007) found that SLODR was not produced by subtests with lower loadings on the general factor. Most recently, an investigation (Murray, Dixon, & Johnson, 2013) using a combination of simulated and observed cognitive test data ($N = 436$) questioned the empirical support for SLODR, concluding that it was an artifact of method variance resulting from skewness in the distributions of the manifest variables produced by the *a priori* sorting of individuals into distinct ability groups. Murray and colleagues concluded that conventional methods for assessing SLODR (e.g., CFA) cannot distinguish between effects at different levels of g that have nothing to do with SLODR and true SLODR effects.

1.3 Purpose of the Current Study

Despite the implications of previous SLODR research (e.g., Reynolds, 2013), little is known about the effect that SLODR may have on the predictive validity of IQ tests in clinical practice. This is due to the fact that observed scores (e.g., standard scores) employed by psychologists are not compatible with the latent dimensions examined in the structural equation modeling studies employed by most SLODR researchers (Glutting et al., 2006). Therefore additional research is needed to determine the extent to which SLODR manifests itself at the observed level of measurement. To wit, the purpose of the current study is to examine the tenability of the SLODR hypothesis at the level of prediction using an incremental validity design featuring cognitive-achievement variables from the Woodcock-Johnson III Psycho educational Battery (WJ-III; Woodcock, McGrew, & Mather, 2001a). The WJ-III is an ideal reference battery for such analyses as it explicitly references CHC in its test design framework and has served as the principle psychometric instrument for making refinements to CHC theory over the course of the last decade (Schneider & McGrew, 2012). An incremental validity design is a potentially useful vehicle for testing the effects of SLODR across different ability classes as the results from such analyses are directly applicable to practitioners due to use of observed level variables. If the SLODR hypothesis is correct than differential prediction should be observed across different ability classes. More specifically, it is hypothesized that g will exert diminished predictive effects as general ability increases which will result in concomitant increases in predictive effects at the lower-order factor level. The current investigation is the first incremental validity design to account for the effects of SLODR and to this authors knowledge is the first empirical examination of the potential effects of SLODR on composite or factor level variables. The current study is an extension of previous research and will potentially provide practitioners with additional information regarding correct interpretation of contemporary cognitive measures in clinical practice.

2. Method

2.1 Participants

The participants were children and adolescents ages 6-0 to 18-11 ($N = 4,722$) drawn from the standardization sample for the WJ-III. Table 1 presents the relative proportions across demographics for sex, race, ethnicity, region, community type, and parent education level for the sample along with comparable 2005 U.S. census estimates. The participants ranged in grade from kindergarten to grade 12 with a mean age of 11.48 ($SD = 3.51$).

2.2 Measurement Instruments

The WJ-III COG (Woodcock, McGrew, & Mather, 2001c) is a multidimensional test of general intelligence for ages 2 to 90 years. The measure is comprised of 20 subtests, 14 of which contribute to the measurement of seven CHC-based broad cluster scores: Comprehension-Knowledge (Gc), Fluid Reasoning (Gf), Auditory Processing (Ga), Visual-Spatial Thinking (Gv), Short-Term Memory (Gsm), Long-Term Retrieval (Glr), and Processing Speed (Gs). All of the CHC clusters are differentially weighted according to their relative *g* loadings and then combined to form a full-scale General Intellectual Ability (GIA) composite. All variables on the WJ-III COG are expressed as standard scores with a mean of 100 and a standard deviation of 15. Extensive normative and psychometric data can be found in the WJ-III technical manuals (McGrew, Schrank, & Woodcock, 2007; McGrew & Woodcock, 2001). The WJ-III ACH (Woodcock, McGrew, & Mather, 2001b) is a comprehensive academic assessment battery designed to measure five academic domains: Reading, Written Language, Mathematics, Oral Language, and Academic Knowledge. The WJ-III ACH is comprised of 22 subtests that combine to provide 17 broad clusters and a total achievement composite score. Broad clusters are expressed as standard scores with a mean of 100 and a standard deviation of 15. Additional technical information for the WJ-III ACH can be found in the WJ-III technical manuals (McGrew et al., 2007; McGrew & Woodcock, 2001).

2.3 Data Analysis

To assess the potential effects of SLODR on the predictive validity of the WJ-III COG available cases were sorted into ability groups that were differentiated according to overall level of cognitive ability as estimated by the GIA score. Specifically, a "Below Average" group composed of individuals with GIA scores of 80 and below, an "Average Ability" group (81 to 119), and an "Above Average" group (120 and above) were formulated. Hierarchical multiple regression analyses (HMR) were conducted to assess the proportions of WJ-III ACH broad score variance accounted for by the observed WJ-III COG GIA and CHC cluster scores. The WJ-III COG GIA was entered into the first block, and the CHC cluster scores were entered jointly into the second block of the SPSS version 21 linear regression analysis. WJ-III ACH analyses included the Broad Reading, Broad Mathematics, and Broad Written Language scores as criterion variables, in order to examine the potential effects of SLODR over a range of general academic abilities. The change in the WJ-III ACH achievement variance predicted by the CHC cluster scores in the second block of the regression model provided an estimate of the incremental prediction beyond the GIA in the first block of the model. According to Pedhazur (1997), these variance partitioning procedures are appropriate given the predictive nature of the current study. The results were interpreted using the resulting R^2 statistic as an effect size. Guidelines for interpreting R^2 as an effect size are found in Cohen (1988); they are "small," .01; "medium," .09; and "large," .25. The critical coefficient in hierarchical multiple regression analysis is the incremental squared multiple correlation coefficient (ΔR^2). The ΔR^2 represents the amount of variance that is explained by an independent variable (IV) after controlling for the effects of IVs previously entered in the regression equation. At present there are no conventional guidelines for interpreting the ΔR^2 coefficient, thus Cohen's interpretive framework for R^2 was applied in the current investigation. To further examine the SLODR hypothesis, the obtained R^2 estimates for the three groups were converted to proportions and analyzed using a chi-square goodness of fit test (Glass & Hopkins, 1996) utilizing the R^2 estimates obtained from the Average sample as reference parameters. Total variance for each prediction model was summed to 100 and divided into three categories (e.g., Carroll, 1995): a) achievement variance attributable to the GIA; b) variance attributable to the CHC clusters; and c) unpredictable score variance (which encompasses an unspecified combination of error variance and dimensions not specified within the prediction model). Non-parametric statistics were utilized to assess for group-level effects due to the fact that the proportions utilized for the analyses do not meet the assumptions for conventional parametric tests.

3. Results

The means, standard deviations, skewness, and kurtosis statistics for all of the WJ-III cognitive and achievement variables are listed in Table 2. The mean (99.99 to 101.38) and standard deviation ranges (14.62 to 15.70) for the cognitive and achievement variables generally reflect values that would be expected for normally distributed standard score variables. Skewness values provided evidence of normally distributed symmetry for all the variables, ranging from -0.41 to 0.11. Additionally, inspection of the residual plots of the data indicated that the regression models utilized in this study met the assumptions for homoscedasticity of the residuals.

To assess whether Spearman's law of diminishing returns (SLODR) impacted the incremental validity of the CHC cluster scores on the WJ-III COG, the school age sample was divided into three subgroups based upon estimated level of general ability, via the observed GIA score for each individual. The results of the HRA analyses that were conducted on the Below Average, Average, and Above Average subgroups can be found in Table 4. Additionally, secondary chi-square goodness of fit tests were conducted that compared differences between the Below Average and Above Average group variance coefficients and those obtained from the Average group (serving as a control for parameter estimates). Goodness of fit test results can be found in Table 5.

3.1 Average Group

Inspection of the zero-order correlations for the independent WJ-III COG variables in Table 3 indicate that all of the CHC cluster scores demonstrated statistically significant ($p < .01$, two-tailed) correlations with the full-scale GIA score. In the Average group ($M_{GIA} = 100.47$), the GIA accounted statistically significant ($p < .05$) portions of each of the WJ-III ACH broad cluster scores. Across the 3 regression models utilized to predict Broad Reading, Broad Mathematics, and Broad Written Language on the WJ-III ACH, the GIA accounted for 27% (Broad Mathematics) to 33% (Broad Reading) of the criterion variable variance. The R^2 coefficients that corresponded with those values are all considered to be large effect sizes. Conversely, the CHC broad cluster scores entered jointly into the second block of the regression equations accounted for statistically significant increments of achievement variance ranging from 4% (Broad Mathematics) to 8% (Broad Reading) of additional achievement variance in the prediction models. The ΔR^2 values that corresponded with those variance increments all represented small effect sizes.

3.2 Below Average Group

Inspection of the zero-order correlations for the independent WJ-III COG variables in Table 3 reveals a pattern of descriptive relationships that is somewhat inconsistent with the SLODR hypothesis. Whereas all of the CHC cluster scores demonstrated statistically significant ($p < .01$, two-tailed) correlations with the full-scale GIA score, which is concordant with the SLODR hypothesis, the strength of the correlations was slightly weaker than the relationships that were observed in the Average group matrix (discordant with the SLODR hypothesis). In the Below Average group ($M_{GIA} = 72.11$), the GIA accounted statistically significant ($p < .05$) portions of each of the WJ-III ACH broad cluster scores. Across the 3 regression models utilized to predict Broad Reading, Broad Mathematics, and Broad Written Language on the WJ-III ACH, the GIA accounted for 24% (Broad Mathematics) to 32% (Broad Written Language) of the criterion variable variance. The R^2 coefficients that corresponded with those values are all considered to be large effect sizes. Conversely, the CHC broad cluster scores entered jointly into the second block of the regression equations accounted for statistically significant increments of achievement variance ranging from 10% (Broad Mathematics and Broad Written Language) to 11% (Broad Reading) of additional achievement variance in the prediction models. The ΔR^2 values that corresponded with those variance increments all represented moderate effect sizes. Goodness of fit tests results in Table 5, examining the significance of differences observed between the variance coefficients obtained from the HMR analyses of the Below Average group and control estimates from the Average subsample, indicated statistically significant group differences for predicting Broad Mathematics, $\chi^2(2) = 9.46$, $p = .001$.

3.3 Above Average Group

Inspection of the zero-order correlations for the independent WJ-III COG variables in Table 3 reveals a pattern of descriptive relationships that is consistent with the SLODR hypothesis. Although all of the CHC cluster scores, with the exception of Processing Speed, demonstrated statistically significant ($p < .01$, two-tailed) correlations with the full-scale GIA score, the strength of these relationships was noticeably weaker than those observed in the Below Average and Average group matrices. Additionally, a consistent pattern of negative inter-cluster correlations suggests enhanced differentiation at the cluster level. In the Above Average group ($M_{GIA} = 126.85$), the GIA accounted statistically significant ($p < .05$) portions of each of the WJ-III ACH broad cluster scores. Across the 3 regression models utilized to predict Broad Reading, Broad Mathematics, and Broad Written Language on the WJ-III ACH, the GIA accounted for 11% (Broad Written Language) to 20% (Broad Reading) of the criterion variable variance. The R^2 coefficients that corresponded with those values reflect moderate to large effect sizes. Conversely, the CHC broad cluster scores entered jointly into the second block of the regression equations accounted for statistically significant increments of achievement variance ranging from 4% (Broad Mathematics) to 15% (Broad Reading) of additional achievement variance in the prediction models. The ΔR^2 values that corresponded with those variance increments represented small to moderate effect sizes.

Goodness of fit tests results in Table 5, examining the significance of differences observed between the variance coefficients obtained from the HMR analyses of the Above Average group and control estimates from the Average subsample, indicated statistically significant group differences for predicting Broad Reading, $\chi^2(2) = 11.86$, $p = .002$, Broad Mathematics, $\chi^2(2) = 10.10$, $p = .001$, and Broad Written Language, $\chi^2(2) = 21.10$, $p < .001$.

3.4 Summary

The pattern of GIA predictive effects across levels of ability for each of the criterion achievement variables is presented in Figure 1. Consistent with the SLODR hypothesis, inspection of the graphic array reveals a general decrease in predictive effects associated with the general factor for all of the WJ-III ACH variables that were assessed in the current study. As indicated by SLODR, the GIA accounted for more achievement variance in the Below Average ability group than in the Above Average ability group, although the proportions of achievement variance predicted by CHC cognitive abilities were relatively equal across groups. However, in the case of the Above Average group, CHC abilities provided for increases in prediction that were equal to or exceeded proportions already accounted for by the general factor.

4. Discussion

SLODR refers to the phenomenon that g loses some of its explanatory value relative to other lower-order cognitive abilities at higher levels of g . In this study, the presence of SLODR in the WJ-III standardization sample data was investigated with hierarchical multiple regression analysis (HMR). At a descriptive level, general factor variance decreased in predicting outcomes from the Below Average group to the Above Average group across all three WJ-III ACH criterion domains that were examined. These findings are consistent with the differentiation hypothesis. A second characteristic of SLODR is that lower-order ability effects are more prevalent in higher ability groups as a consequence of decreased g loadings (Reynolds, 2013). In the Broad Reading and Broad Written Language models, CHC cluster effects were greater for the Above Average group than for the Below Average group, consistent with the SLODR hypothesis. However, in the Broad Mathematics model, a discordant pattern of prediction was observed wherein the variance attributable to the CHC clusters unexpectedly decreased from the Below Average group to the Above Average group. The overall findings from the present study are consistent with previous research examining the effects of SLODR at the latent level (e.g., Reynolds, 2013; Reynolds, Hajovsky, Niileksela, & Keith, 2011; Reynolds, Keith, & Beretvas, 2010). To this author's knowledge, this is the first empirical examination of such effects at the level of prediction. The results of the current study indicate that SLODR had a consistent moderating effect in predicting Broad Reading and Broad Written Language outcomes, whereas less consistent effects were observed in the Broad Mathematics prediction models.

In sum, the current results add to the growing literature suggesting that g does not exert its effects equally across levels of ability. Although it has been suggested that clinical interpretation of contemporary IQ measures should be limited to the full scale IQ score (e.g., Canivez, 2013; Glutting et al., 2006), these findings suggest that more interpretative weight for broad and specific cognitive abilities may be warranted for individuals with higher levels of g (Reynolds et al., 2010). In the Above Average group, the general factor accounted for a significantly lower portion of achievement variance compared to other estimates obtained within this study. Across the prediction models, the variance coefficients associated with the general factor were consistent, with values that corresponded to large effect size estimates. Low general factor values such as these have only been reported in incremental validity studies with clinical samples (e.g., Nelson & Canivez, 2012; Nelson, Canivez, & Watkins, 2013). In the Below Average and Average groups, general factor estimates were slightly larger but not within previously estimated ranges (e.g., 40% to 60%). These results indicate that relationships between achievement and the general factor, as estimated by the WJ-III COG may be less stable across subgroups and disaggregated samples than previously thought (e.g., Brody, 1992; Jensen, 1998; Mackintosh, 2011). Ultimately, more research is needed to determine if these results are an artifact of low base rates or a legitimate challenge to validity estimates obtained with larger sample sizes (Matarrazo & Herman, 1985). Although small to moderate incremental variance was accounted for by the CHC factors across the regression models, the proportion of additional variance accounted for at the cluster level was substantially higher than previous incremental validity studies with this same measurement instrument (e.g., McGill, 2015; McGill & Busse, 2014).

As an example, in the Broad Mathematics model for the Above Average group, the ΔR^2 coefficient for the CHC factors corresponded to a small effect size estimate, however, due to the fact that the general factor accounted for such a low proportion of math achievement in the first block of the regression equation, the CHC factors provided a 31% increase in predictive effects in the model. This example illustrates the danger of interpreting effect size estimates in isolation because the practical validity of lower-order scores is inextricably tied to the magnitude of effect attributed to the full-scale score (Keith, 2015). Consistent with previous incremental validity researches using cognitive measures, multicollinearity between the GIA and the first-order cluster scores was observed across all of the multiple regression analyses in the current study. Multicollinearity refers to a potential threat to validity in multiple regression research that is introduced when a prediction model utilizes independent variables that are significantly correlated (Pedhazur, 1997). It should be noted that multicollinearity is not a threat to validity in predictive studies that are limited to interpreting the R^2 statistic (Cohen et al., 2003), nor does it invalidate the use of HMR analysis to detect improvements in R^2 such as those provided by the CHC cluster scores beyond the GIA composite (Dana & Dawes, 2007; Schneider, 2008).

4.1 Implications for Future Research

Additional research is needed to determine the mechanism responsible for the discordant cluster-level effects in Broad Mathematics in addition to examining the generalizability of these findings to other measurement instruments. Furthermore, previous research (e.g., Arden & Plomin, 2007; Facon, 2004) suggests that SLODR does not manifest itself equally across the lifespan. Although the present study provides evidence for the effects of SLODR at the level of prediction during childhood and adolescence, additional research examining the generality of these effects with adult samples would be beneficial.

4.2 Limitations

This study is not without limitations that should be considered when interpreting the results. The most important limitation of the present study is the use of an archived standardization sample. Although the sample was relatively large and nationally representative, additional research is needed to determine if these results generalize to other measurement instruments. Additionally, the creation of arbitrary ability groups resulted in within-group range restriction for the cognitive variables utilized in the current study (Murray, Dixon & Johnson, 2013). Examination of the ability group correlation matrices in Table 3 indicates attenuation of coefficients in the Below Average and Above Average groups. Although the potential threat of range restriction should be considered when interpreting these results, Hunt (2011) argued that restriction of range is not an issue in predictive research. Finally, power was low for some of the subgroup analyses due to small sample sizes and the categorical treatment of continuous data (Cohen, 1988; MacCallum, Zhang, Preacher, & Rucker, 2002). Based on a power level of .80 and an alpha of .05, the smallest effect sizes that could be detected when CHC cluster scores were entered jointly in the regression equation for sample sizes of 183 and 191 were .09 and .08 respectively. Nevertheless, cluster-level effects exceeded these thresholds in all of the regression models for the implicated subgroups with smaller sample sizes.

4.3 Conclusion

Overall, the results obtained within this study provide evidence for the effects of SLODR at the level of prediction in the domains of Broad Reading and Broad Written Language. As a consequence, in some of the prediction models, CHC clusters jointly accounted for more incremental variance than that already accounted for by the general factor, a finding that has not previously been reported within the published incremental validity literature. Less consistent cluster-level effects were evidenced in the Broad Mathematics models. The results from the current study add to the growing SLODR literature base as this is the first study to assess the differentiation hypothesis at the level of prediction. Although replication of these results with other measurement instruments is needed, these findings suggest that professionals in psychology and education should be mindful of the effects of SLODR when appraising the intellectual functioning of individuals in clinical practice.

References

- Arden, R., & Plomin, R. (2007). Scant evidence for Spearman's law of diminishing returns in middle childhood. *Personality and Individual Differences, 42*, 743-753. doi: 10.1016/j.paid.2006.08.010
- Brody, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.
- Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwab (Eds.), *The Oxford handbook of child psychological assessment* (pp. 84-112). New York: Oxford University Press.
- Canivez, G. L., Watkins, M. W., James, T., James, K., & Good, R. (2014). Incremental validity of WISC-IV UK factor index scores with a referred Irish sample: Predicting performance on the WIAT-II UK. *British Journal of Educational Psychology, 84*, 667-684. doi: 10.1111/bjep.12056
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.
- Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research, 30*, 429-452. doi: 10.1207/s15327906mbr3003_6
- Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5-21). New York: Pergamon Press.
- Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). New York: Psychology Press.
- Coyle, T. R., & Rindermann, H. (2013). Spearman's law of diminishing returns and national ability. *Personality and Individual Differences, 55*, 406-410. doi: 10.1016/j.paid.2013.03.023
- Dana, J., & Dawes, R. (2007). Comment on Fiorello et al. interpreting intelligence test results for children with disabilities: Is global intelligence relevant? *Applied Neuropsychology, 14*, 21-25. doi: 10.1080/09084280701280379
- Deary, I. J., Egan, V., Gibson, G. J., Austin, E., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence, 23*, 105-132. doi: 10.1016/S0160-2896(96)90008-2
- Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are the highest for low IQ groups. *Intelligence, 13*, 349-310. doi: 10.1016/S0160-2896(89)80007-8
- Facon, B. (2004). Are correlations between cognitive abilities highest in low-IQ groups during childhood? *Intelligence, 32*, 391-401. doi: 10.1016/j.intell.2004.06.002
- Fogarty, G. J., & Stankov, L. (1995). Challenging the "law of diminishing returns." *Intelligence, 21*, 157-174.
- Gignac, G. E. (2014). Dynamic mutualism versus g factor theory: An empirical test. *Intelligence, 42*, 89-97. doi: 10.1016/j.intell.2013.11.004
- Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Needham Heights, MA: Allyn & Bacon.
- Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *Journal of Special Education, 40*, 103-114. doi: 10.1177/00224669060400020101
- Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2013). Further insights on the French WISC-IV factor structure through Bayesian structural equation modeling. *Psychological Assessment, 25*, 496-508. doi: 10.1037/a0030676
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*, 79-132. doi: 10.1016/S0160-2896(97)90014-3
- Hartmann, P., & Reuter, M. (2006). Spearman's "Law of Diminishing Returns" tested with two methods. *Intelligence, 34*, 47-62. doi: 10.1016/j.intell.2005.06.002
- Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. McGrew, J. K. Werder, & R. W. Woodcock, *WJ-R technical manual* (pp. 197-232). Itasca, IL: Riverside Publishing.
- Hunt, E. B. (2011). *Human intelligence*. New York: Cambridge University Press.
- Hunsley, J. (2003). Introduction to the special section on incremental validity and utility in clinical assessment. *Psychological Assessment, 15*, 443-445. doi: 10.1037/1040-3590.15.4.443
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

- Jensen, A. R. (2003). Regularities in Spearman's law of diminishing returns. *Intelligence*, 31, 95-105.
- Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, 33, 393-416. doi: 10.1016/j.intell.2004.12.002
- Jones, W. T. (1952). *A history of western philosophy*. New York: Harcourt-Brace.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children-Second Edition*. Circle Pines, MN: American Guidance Service.
- Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling (2nd ed.)*. New York: Routledge.
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19-40. Doi: 10.1037//1082-989X.7.1.19
- Mackintosh, N. J. (2011). *IQ and human intelligence (2nd ed.)*. New York: Oxford University Press.
- Matarazzo, J. D., & Herman, D. O. (1985). Clinical uses of the WAIS-R: Base rates of differences between VIQ and PIQ in the WAIS-R standardization sample. In B. B. Wolman (Ed.), *Handbook of Intelligence* (pp. 899-932). New York: John Wiley.
- McGill, R. J. (2015). Incremental criterion validity of the WJ-III COG clinical clusters: Marginal predictive effects beyond the general factor. *Canadian Journal of School Psychology*, 30, 51-63. doi: 10.1177/0829573514560529
- McGill, R. J., & Busse, R. T. (2014). Incremental validity of the WJ III COG: Limited predictive effects beyond the GIA-E. *School Psychology Quarterly*. Advance online publication. doi: 10.1037/spq0000094
- McGrew, K. S. (2009). Editorial: CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1-10. doi: 10.1016/j.intell.2008.08.004
- McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Woodcock-Johnson III Normative Update technical manual*. Rolling Meadows, IL: Riverside.
- McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III technical manual*. Itasca, IL: Riverside.
- Murray, A. L., Dixon, H., & Johnson, W. (2013). Spearman's law of diminishing returns: A statistical artifact? *Intelligence*, 41, 439-451. doi: 10.1016/j.intell.2013.06.007
- Nelson, J. M., & Canivez, G. L. (2012). Examination of the structural, convergent, and incremental validity of the Reynolds Intellectual Scales (RIAS) with a clinical sample. *Psychological Assessment*, 24, 129-140. doi: 10.1037/a0024878
- Nelson, J. M., Canivez, G. L., & Watkins, M. W. (2013). Structural and incremental validity of the Wechsler Adult Intelligence Scale-Fourth Edition with a clinical sample. *Psychological Assessment*, 25, 618-630. doi: 10.1037/a0032086
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction (3rd ed.)*. New York: Holt, Rinehart, & Winston.
- Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95, 129-140. doi: 10.1080/00223891.2012.725437
- Reynolds, M. R. (2013). Interpreting intelligence test composite scores in light of Spearman's law of diminishing returns. *School Psychology Quarterly*, 28, 63-76. doi: 10.1037/spq0000013
- Reynolds, M. R., Hajovsky, D. B., Niileksela, C. R., & Keith, T. Z. (2011). Spearman's Law of Diminishing Returns and the DAS-II: Do g effects on subtest scores depend on g? *School Psychology Quarterly*, 26, 275-289. doi: 10.1037/a0026190
- Reynolds, M. R., & Keith, T. Z. (2007). Spearman's law of diminishing returns in hierarchical models of intelligence for children and adolescents. *Intelligence*, 35, 267-281. doi: 10.1016/j.intell.2006.08.002
- Reynolds, M. R., Keith, T. Z., & Beretvas, S. N. (2010). Use of factor mixture modeling to capture Spearman's law of diminishing returns. *Intelligence*, 38, 231-241. doi: 10.1016/j.intell.2010.01.002
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22, 53-61. doi: 10.1007/BF02289209
- Schneider, W. J. (2008). Playing statistical Ouija board with communality analysis: Good questions, wrong assumptions. *Applied Neuropsychology*, 15, 44-53. doi: 10.1080/09084280801917566
- Schneider, W. J., & Flanagan, D. P. (2015). The relationship between theories of intelligence and intelligence tests. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 317-340). New York: Springer.
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D.

- P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99-144). New York: Guilford Press.
- Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 201-293. Retrieved from <http://psychclassics.yorku.ca/Spearman>
- Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.
- Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, 8, 271-281. doi: 10.1111/j.2044-8295.1916.tb00133.x
- Thorndike, R. L., Lay, W., & Dean, P. R. (1909). The relation of accuracy in sensory discrimination to general intelligence. *American Journal of Psychology*, 20, 364-369. Retrieved from <http://www.jstor.org/stable/1413366>
- Thurstone, L. L. (1947). *Multiple factor analysis: A development and expansion of the vectors of the mind*. Chicago: University of Chicago Press.
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.
- Wolfe, D. (1940). *Factor analysis to 1940*. Chicago: University of Chicago Press.
- Woodcock, R. W., McGrew, K. S., Mather, N. (2001a). *Woodcock-Johnson III*. Itasca, IL: Riverside.
- Woodcock, R. W., McGrew, K. S., Mather, N. (2001b). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside.
- Woodcock, R. W., McGrew, K. S., Mather, N. (2001c). *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside.

Table 1: Demographic Information for the Study Sample (N = 4,722)

Variable	n	Percent of Sample	Percent of U.S. Population
Sex			
Male	2382	50.4	51.2
Female	2340	49.6	48.8
Race			
European American	3702	78.4	78.5
African American	684	14.5	16.1
Asian American	241	5.1	4.1
Native American	95	2.0	1.3
Hispanic/Latino/Chicano			
Yes	567	12.0	18.7
No	4155	88.0	81.3
Census Region			
Northeast	1133	24.0	17.8
Midwest	978	20.7	22.3
South	1487	31.5	35.9
West	1124	23.8	24.0
Community Size			
Large City	2800	59.3	68.3
Suburban	1025	21.7	10.7
Rural	897	19.0	21.0
Type of School			
Public	4099	86.8	86.5
Private	571	12.1	11.3
Home	52	1.1	2.2
Foreign Born			
No	4486	95.0	94.3
Yes	236	5.0	5.7
Father's Education			
Less than High School	527	11.7	13.3
High School	1507	33.5	31.8
More than High School	2465	54.8	54.9
Not Available	223		
Mother's Education			
Less than High School	432	9.6	10.9
High School	1485	33.0	29.5
More than High School	2583	57.4	59.6
Not Available	222		

Table 2: Univariate Descriptive Statistics for WJ-III (Woodcock, McGrew & Mather, 2001a) Cognitive-Achievement Variables

Variables	<i>NMSD</i>	Skewness	Kurtosis		
GIA	2130	100.36	14.94	-0.12	0.39
Comprehension-Knowledge	2903	100.85	14.86	-0.40	0.53
Fluid Reasoning	3253	100.30	15.45	-0.34	0.32
Auditory Processing	3435	100.25	15.64	0.11	0.40
Visual-Spatial Thinking	2775	100.36	14.62	-0.15	0.37
Long-Term Retrieval	3078	100.16	14.88	0.03	0.34
Short-Term Memory	3746	100.74	15.49	-0.07	0.48
Processing Speed	3326	99.99	15.11	-0.07	0.47
Broad Reading	3845	101.38	15.26	-0.36	0.99
Broad Mathematics	3954	100.76	15.70	-0.20	0.68
Broad Written Language	3877	100.51	15.44	-0.41	1.07

Note. GIA = General Intellectual Ability Composite. Obtained values rounded to the nearest hundredth.

Table 3: Zero-Order Correlation Coefficients between Independent Variables on the WJ-III COG Across Levels of Ability

<u>Below Average (n = 183)^a</u>									
Variable	GIA	Gc	Gf	Ga	Gv	Gsm	Glr	Gs	
GIA	-								
Comprehension-Knowledge (Gc)	.59*	-							
Fluid Reasoning (Gf)	.51*	-.06	-						
Auditory Processing (Ga)	.46*	.29*	-.01	-					
Visual-Spatial Thinking (Gv)	.31*	.16	.05	.16*	-				
Short-Term Memory (Gsm)	.52*	.00	.10	.16*	.02	-			
Long-Term Retrieval (Glr)	.57*	.38*	.11	.14	.27	.22*	-		
Processing Speed (Gs)	.38*	.01	.14	.15*	.15*	.07	.30*	-	
<u>Average (n = 1,735)^a</u>									
Variable	GIA	Gc	Gf	Ga	Gv	Gsm	Glr	Gs	
GIA	-								
Comprehension-Knowledge (Gc)	.73*	-							
Fluid Reasoning (Gf)	.72*	.38*	-						
Auditory Processing (Ga)	.46*	.30*	.18*	-					
Visual-Spatial Thinking (Gv)	.37*	.18*	.26*	.12*	-				
Short-Term Memory (Gsm)	.56*	.16*	.20*	.18*	.10*	-			
Long-Term Retrieval (Glr)	.56*	.36*	.37*	.19*	.23*	.20*	-		
Processing Speed (Gs)	.36*	.09*	.17*	.19*	.16*	.13*	.23*	-	
<u>Above Average (n = 191)^a</u>									
Variable	GIA	Gc	Gf	Ga	Gv	Gsm	Glr	Gs	
GIA	-								
Comprehension-Knowledge (Gc)	.42*	-							
Fluid Reasoning (Gf)	.43*	-.02	-						
Auditory Processing (Ga)	.20*	.04	-.12	-					
Visual-Spatial Thinking (Gv)	.29*	-.01	.03	-.05	-				
Short-Term Memory (Gsm)	.51*	-.02	-.08	.06	.12	-			
Long-Term Retrieval (Glr)	.28*	.09	.10	.13	.16	-.11	-		
Processing Speed (Gs)	.17	-.11	-.07	.05	.06	.01	.06	-	

Note. Values rounded to nearest hundredth. *Coefficients were statistically significant ($p < .01$, two-tailed). GIA = General Intellectual Ability Composite.

^a Denotes valid cases listwise.

Table 4: Hierarchical Regression Analyses Assessing the Impact of SLODR on the Predictive Validity of WJ-III COG (Woodcock, McGrew, & Mather, 2001c) Variables across Differential Levels of General Ability

Broad Reading		Below Average				Average Above Average				Below Average		Average Above Average				
Predictor	R ²	95% CI	Δ R ²	Inc. ^b	R ²	95% CI	Δ R ²	Inc. ^b	R ²	95% CI	Δ R ²	Inc. ^b	R ²	95% CI	Δ R ²	Inc. ^b
GIA	.30*	[.19, .42]	-	30%	.33*	[.30, .37]	-	33%	.20*	[.07, .33]	-	20%				
CHC Factors (df = 7) ^a	.41	[.30, .52]	.11*	11%	.41	[.37, .45]	.08*	8%	.35	[.25, .45]	.15*	15%				
Gc, Gf, Ga, Gv, Glr																
Gsm, Gs																

Broad Mathematics		Below Average				Average Above Average				Below Average		Average Above Average				
Predictor	R ²	95% CI	Δ R ²	Inc. ^b	R ²	95% CI	Δ R ²	Inc. ^b	R ²	95% CI	Δ R ²	Inc. ^b	R ²	95% CI	Δ R ²	Inc. ^b
GIA	.24*	[.13, .37]	-	24%	.27*	[.24, .31]	-	27%	.13*	[.04, .21]	-	13%				
CHC Factors (df = 7) ^a	.34	[.24, .45]	.10*	10%	.31	[.27, .35]	.04*	4%	.17	[.07, .26]	.04*	4%				
Crystallized Ability																
Gc, Gf, Ga, Gv, Glr																
Gsm, Gs																

Broad Written Language		Below Average				Average Above Average				Below Average		Average Above Average				
Predictor	R ²	95% CI	Δ R ²	Inc. ^b	R ²	95% CI	Δ R ²	Inc. ^b	R ²	95% CI	Δ R ²	Inc. ^b	R ²	95% CI	Δ R ²	Inc. ^b
GIA	.32*	[.20, .43]	-	32%	.27*	[.23, .30]	-	27%	.11*	[.00, .21]	-	11%				
CHC Factors (df = 7) ^a	.42	[.31, .53]	.10*	10%	.32	[.28, .36]	.06*	6%	.24	[.14, .34]	.14*	14%				
Gc, Gf, Ga, Gv, Glr																
Gsm, Gs																

Note. GIA = General Intellectual Ability Composite, CHC = Cattell-Horn-Carroll, Gc = Comprehension-Knowledge, Gf = Fluid Reasoning, Ga = Auditory Processing, Gv = Visual-Spatial Thinking, Glr = Long-Term Retrieval, Gsm = Short-Term Memory, Gs = Processing Speed. Squared multiple correlation coefficients for the CHC clusters reflect the total variance accounted for by that variable when it is included in a prediction equation in addition to the GIA.

^a Degrees of freedom reflects controlling for GIA.

^b Represents proportion of variance accounted for by clusters entered jointly into the regression equation after controlling for the effects of the GIA. All values rounded to the nearest hundredth, may not sum to 100 due to rounding errors.

* $p < .05$.

Table 5: Chi-Square Goodness of Fit Test Results Assessing the Significance of the Impact of SLODR on Obtained Variance Percentages across Levels of General Ability

GIA Group Level	Broad Reading		χ ²	p	Broad Mathematics		χ ²	p	Broad Written Language		χ ²	p
	P _E	P _O			P _E	P _O			P _E	P _O		
Below Average			1.40	.50			9.46	.00			4.80	.09
General Factor	33	30			27	24			27	32		
CHC Factors	8	11			4	10			6	10		
Unpredictable ^a	59	59			69	66			67	58		
Above Average			11.86	.00			10.10	.00			21.10	.00
General Factor	33	20			27	13			27	11		
CHC Factors	8	15			4	4			6	14		
Unpredictable ^a	59	65			69	83			67	75		

Note. P_E = expected percentages, P_O = observed percentages, CHC = Cattell-Horn-Carroll. df = 2. Expected values obtained from HMR analysis of average ability subsample for each dependent variable. Variance percentages are R² and ΔR² multiplied by 100. CHC cluster variance obtained using joint block entry procedures.

^a Combination of error and unaccounted for variance in the prediction models.

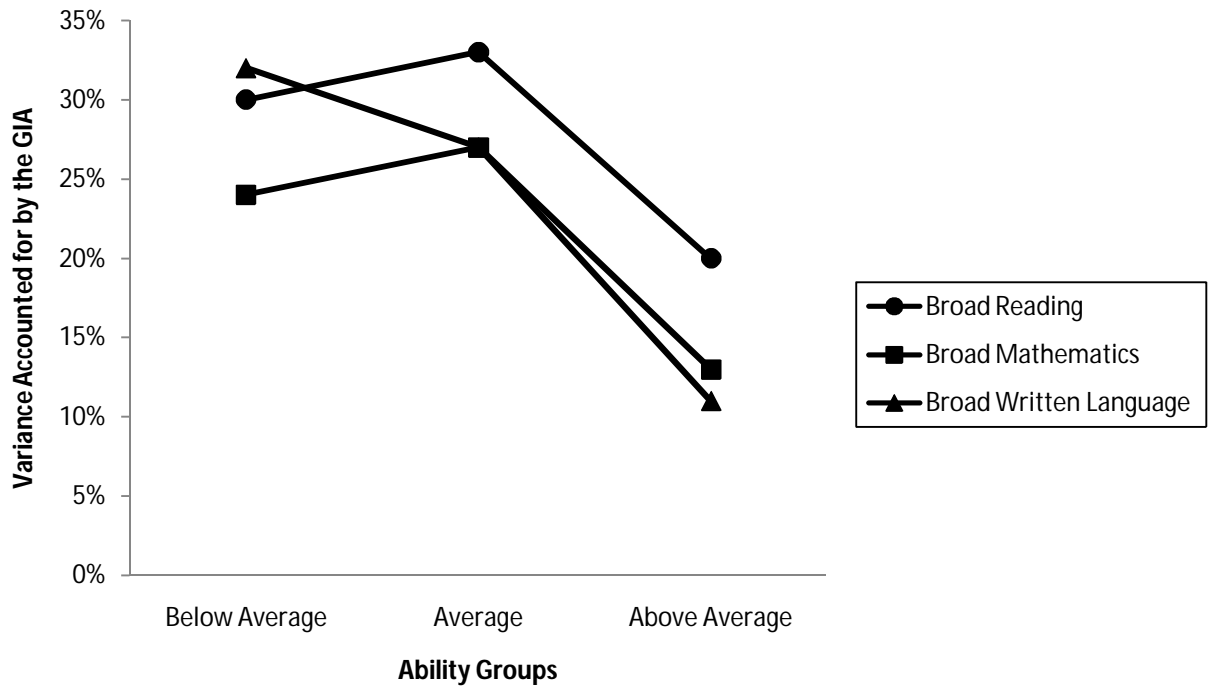


Figure 1: Pattern of General Factor Predictive Effects for Reading, Mathematics, and Written Language Outcomes across Levels of Ability on the WJ-III (Woodcock, McGrew, & Mather, 2001a)