The Report Committee for Jamison Elder Carrigan
Certifies that this is the approved version of the following Report:

Does Spearman's Law of Diminishing Returns Differ by Race/Ethnicity or by Sex?

APPROVED BY
SUPERVISING COMMITTEE:

Timothy Z. Keith, Supervisor

Erin M. Rodriguez

Does Spearman's Law of Diminishing Returns Differ by Race/Ethnicity or by Sex?

by

Jamison Elder Carrigan

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

December 2018

Abstract

Does Spearman's Law of Diminishing Returns Differ by Race/Ethnicity or by Sex?

Jamison Elder Carrigan, M.A.

The University of Texas at Austin, 2018

Supervisor: Timothy Z. Keith

Abstract: The purpose of this study is to determine whether Spearman's Law of Diminishing Returns (SLODR) operates differentially by race and by sex. In its simplest terms, SLODR is the assertion that a general intelligence factor (*g*) is a less meaningful determinant of cognitive ability at higher levels of ability. Using the standardization data from the Differential Ability Scales, Second Edition (DAS-II), separate multigroup bifactor models (by race and by sex) will be created in which loadings and squared loadings from *g* to subtests will be estimated.

Table of Contents

List of Tables ... vi

List of Figures ... vii

INTRODUCTION ... 1

INTEGRATIVE ANALYSIS AND INTERPRETATION ... 4

A general intelligence factor ... 4

Explaining Spearman's Law of Diminishing Returns ... 6

Methodologial considerations when studying SLODR ... 8

Race differences and *g* ... 15

Sex differences and *g* ... 16

PROPOSED RESEARCH STUDY ... 19

Statement of Problem/Purpose ... 19

Research Questions and Hypotheses ... 20

METHOD ... 21

Instrumention ... 21

Participants ... 23

ANALYSES AND EXPECTED RESULTS ... 25

DISCUSSION ... 32

Summary ... 32

Limitiations ... 32

Future Directions and Implications ... 33

References ... 35

List of Tables

Table 1: CHC Broad Abilities and Descriptions 6

Table 2: Description of the DAS-II Subtests Included in the School-Age Battery 22

Table 3: Demographic Characteristics of the DAS-II Standardization Sample..........24

List of Figures

Figure 1: Two hierarchical conceptions of intelligence 11

Figure 2: Within-class confirmatory factor analysis model for the DAS-II................25

Figure 3: Model-implied relations between the score on the g factor and each broad ability 25

Introduction

The purpose of this study is to determine whether Spearman's Law of Diminishing Returns (SLODR) operates differentially by race/ethnicity or by sex. In its simplest terms, SLODR is the notion that at higher levels of ability, a general intelligence factor (*g*) is a less meaningful determinant of cognitive ability than at lower levels of ability (Detterman & Daniel, 1989; Spearman, 1927).

Test scores on measures of cognitive ability are positively correlated; a general intelligence factor (*g*) is commonly cited as the primary reason for these correlations. This *g* factor is therefore useful in explaining individual differences in cognitive ability test performance (Bartholomew, 2004; Jensen, 1998). A key assumption of *g* is that it operates in a constant manner across individuals (Wolfle, 1940). Spearman (1927), however, found that the influence of *g* is differentiated by cognitive ability: children of lower ability had higher intercorrelations between mental ability tasks than did children of average ability. This phenomenon has been found in other samples and has become known as Spearman's Law of Diminishing Returns. If SLODR is indeed accurate, it would call into question the usefulness of a single factor analytic model to represent those of low and high ability (Detterman & Daniel, 1989).

The study of intelligence as a construct has been controversial from the beginning. Concerns of cultural bias in test measures have existed since the introduction of the first measure of intelligence, the Binet-Simon Scales (Binet & Simon, 1917; Burt, 1921; Stern, 1914). Because these measures were not often created with an eye towards cultural fairness, they often were subject to several criticisms, including a lack of inclusion of minority members in standardization samples (Jensen, 1980; C. R. Reynolds & Lowe, 2009; Valencia & Suzuki, 2001).

Evidence suggests that modern intelligence tests lack cultural bias on the basis of race (Valencia & Suzuki, 2001). Furthermore, there is evidence to support equivalent factor structures across racial groups on well-known measures like the Woodcock Johnson Cognitive battery, the Wechsler Intelligence Scale for Children—Fifth Edition, and the Differential Ability Scales, Second Edition (Edwards & Oakland, 2006; Scheiber, 2016; Trundt, Keith, Caemmerer, & Smith, 2017).

Although no sex differences likely exist for full scale IQ, there are notable differences in certain tasks (Halpern & LaMay, 2000). For example, males demonstrate a consistent advantage on tasks of visual-spatial ability (Voyer, Voyer, & Bryden, 1995). Meanwhile, females tend to perform higher on tasks related to long-term memory and processing speed (e.g., Geffen, Moar, O'Hanlon, Clark, & Geffen, 1990; Stumpf & Jackson, 1994). Another curious phenomenon is that the tails of the IQ distribution tend to be more male-heavy; that is, males are overrepresented in the highest percentiles of overall intelligence and at the lowest (DeFries & Gillis, 1993; Halpern, 2012; Johnson, Carothers, & Deary, 2008). Males are more likely to be diagnosed with learning disabilities, intellectual disabilities, and language disorders (Halpern, 2012; Henning-Stout & Close-Conoley, 1992).

The issues of test bias and fairness remain relevant, especially as new intelligence measures and revisions are developed. These issues are more pressing considering the use of intelligence testing to make determinations of special education eligibility and gifted program placement (Holdnack & Weiss, 2006; Kaufman & Harrison, 1986). The presence of SLODR indicates that the current theoretical model of intelligence may not be the most practically useful.

For example, SLODR may help to explain why overall *g* is an inherently limited way of assessing people of above-average ability and making determinations about giftedness curricula (Deary et al., 1996).

Because intelligence testing is so crucially important, and because the implications of SLODR are so compelling, it therefore follows that SLODR be understood as fully and completely as possible. Despite the amount of research into SLODR, racial differences, and sex differences, no study has attempted to understand the interrelationship of these topics.

The aim of this study is to answer the following question: *Does Spearman's Law of Diminishing Returns (SLODR) operate differentially by race/ethnicity or by sex?* This aim will be accomplished using standardization data from the Differential Ability Scales, Second Edition (DAS-II). Separate multigroup higher-order models (by race/ethnicity and by sex) will be created in which a nonlinear interaction term (*g* x *g*) will be included. These models will be compared using qualitative methods, as traditional fit indices (such as χ^2 or CFI) are not calculated for nonlinear models.

Integrative Analysis and Interpretation

This integrative analysis will begin by explaining Spearman's original hypotheses about the nature of intelligence, including the presence of a single, general intelligence factor. This presentation will be followed by a description of current intelligence theory with special emphasis on Cattell-Horn-Carroll (CHC) theory. Then, Spearman's Law of Diminishing Returns (SLODR) and recent research about this phenomenon will be explained. Specific methodological considerations will be discussed as they relate to the current investigation. Last, a discussion of race/ethnicity and sex differences in intelligence will place the current investigation of SLODR in context of these previous research inquiries.

A general intelligence factor

Spearman's (1927) theory of general intelligence posited that there is a single general factor of intelligence underlying all aspects of cognitive ability. He described this general factor, or *g*, as "something analogous to an 'energy'," a force that could be expended to complete various cognitive tasks. He noted that tasks designed to measure different aspects of intelligence (e.g., reasoning, memory, abstract thought) were highly correlated with one another, evidence of the universality of *g*. Spearman regarded *g* as the most important aspect of his research, as it was such a strong and consistent indicator of mental performance. This consistency is illuminated in his theorem of indifference of the indicator, which states that "for the purpose of indicating the amount of *g* possessed by a person, any test will do just as well as any other, provided only that its correlation with *g* is equally high." This theorem is still used to explain high correlations among intelligence tests (Jensen, 1992).

Spearman (1927) also proposed specific factors, or *s*, that were subsumed by *g*. In his words, *g* is the "amount of a general mental energy," while *s* is "the efficiency of specific mental engines." Other researchers became interested in these specific factors after Spearman published his 1927 work (Holzinger & Harman, 1938; Kelley, 1928; Thurstone, 1938), though some researchers regarded *g* as a lower-order factor in their models (Burt, 1949) or did not consider overall *g* at all (Guilford, 1959).

First conceptualized in the 1940s and fully developed in the 1960s, Cattell and Horn's theory of fluid and crystallized intelligence (Gf-Gc theory) outlined multiple broad factors of intelligence. In its first rendition, five broad factors were hypothesized, including a personality factor (Horn & Cattell, 1966). Horn & Cattell has revised this theory a number of times, eventually dropping the personality factor, and settling on eight broad factors of intelligence: fluid reasoning (Gf); comprehensive knowledge (Gc); visual processing (Gv); auditory processing (Ga); processing speed (Gs); short-term memory (Gsm); long-term retrieval (Glr); and quantitative ability (Gq). The two main features of the theory, Gf and Gc, represent abilities gained primarily from genetic factors and from educational or cultural opportunities, respectively (Horn, 1991).

As factor analytic studies continued to show that the broad abilities described in Gf-Gc theory loaded heavily onto a second-order *g* factor, a theory that combined the *g* model and the Gf-Gc model was introduced (Carroll, 1992, 1993). This theory is represented by a three-stratum factor structure with a general intelligence factor, *g*, at the top, several broad abilities at the second stratum, and many specific abilities at the first stratum. The broad abilities are largely the same as those listed in Gf-Gc theory, though Gq is now generally regarded as an achievement factor as opposed to a cognitive factor. Carroll, after having reanalyzed over 460 factor analytic studies on the nature of intelligence, stated that this model, now called the Cattell-Horn-Carroll theory of intelligence, was the best at describing human cognitive ability. A list of the seven most researched broad abilities, along with their descriptions, can be found in Table 1; this list is adapted from Schneider and McGrew (2012).

Table 1

CHC Broad Abilities and Descriptions

Broad Ability	Description
Fluid Reasoning (*Gf*)	The deliberate but flexible control of attention to solve novel, on-the-spot problems that cannot be performed by relying exclusively on previously learned habits, schemas, and scripts
Crystallized Intelligence (*Gc*)	The depth and breadth of knowledge and skills that are valued by one's culture
Visual Processing (*Gv*)	The ability to make use of simulated mental imagery (often in conjunction with currently perceived images) to solve problems
Auditory Processing (*Ga*)	The ability to detect and process meaningful nonverbal information in sound
Short-Term Memory (*Gsm*)	The ability to encode, maintain, and manipulate information in one's immediate awareness
Long-Term Storage and Retrieval (*Glr*)	The ability to store, consolidate, and retrieve information over periods of time measured in minutes, hours, days, and years
Processing Speed (*Gs*)	The ability to perform simple, repetitive cognitive tasks quickly and fluently

Note: Adapted from "The Cattell-Horn-Carroll model of intelligence", by W.J. Schneider & K.S. McGrew in *Contemporary intellectual assessment: Theories, tests, and issues*, 2012, New York: Guilford Press (pp. 99-112).

Explaining Spearman's Law of Diminishing Returns

Spearman (1927), as noted above, regarded *g* as a sort of mental "energy." Following this same analogy, he proposed a law of diminishing returns for *g* that operates in the same way as in engine mechanics. He posited that the more "energy" a person has available (that is, the more overall intelligence), the less advantage is added with constant, increased amounts. This concept was illustrated by a correlation matrix of 12 cognitive ability tasks for what he called "normal" and "mentally defective" children, with the "normal" group showing a smaller average intercorrelation among the tests. The mean correlation among the tasks for the "normal" children (n = 78) was .466, while for the "defective" children (n = 22) it was .782.

Detterman and Daniel (1989) seemingly rediscovered this phenomenon when analyzing subtests on the Wechsler Adult Intelligence Scale—Revised (WAIS-R) and the Wechsler Intelligence Scale for Children—Revised (WISC-R). They found that subtest intercorrelations were similarly related to ability group; the intercorrelations were highest in the low ability group, and "declined systematically with increasing IQ." For both tests, the intercorrelations were twice as high in the low IQ groups than in the high IQ groups. This publication, however, was apparently unaware of this rediscovery, claiming that Spearman's (1904) emphasis on "positive manifold" implied that the positive manifold was "uniformly distributed over the full range of ability" (Detterman & Daniel, 1989). Here, positive manifold refers to how tests of cognitive ability are all positively correlated with one another. It was Deary and Pagliani (1991) who pointed out that indeed Spearman had hypothesized the very thing that Detterman and Daniel (1989) had discovered.

Nevertheless, Detterman and Daniel (1989) understood the importance of such a finding, stating that "if the finding that correlations between mental tests vary systematically by level of ability is found to be a general one, not specific to certain tests, then the implications of this finding are substantial." Many researchers have replicated these findings using this methodology (or one very similar) on different test batteries since this study (e.g., Deary et al., 1996; Deary & Pagliari, 1991; Evans, 1999; Legree, Pifer, & Grafton, 1996; Lynn, 1992; Lynn & Cooper, 1993; Maxwell, 1972).

Anderson's (1992) theory of minimal cognitive architecture has become a useful and common way for researchers to conceptualize the mechanism by which such a phenomenon is possible. This theory states that a single general processing mechanism constrains the effectiveness of various, more specific domains of cognitive performance. This helps explain why different cognitive domains are correlated (i.e., positive manifold). For those for whom the general processing mechanism is faster or more efficient, the specific mechanisms are less constrained, meaning that these specific abilities are less correlated with one another.

This theory is not necessarily antithetical to the hypotheses of Spearman (1927), who likened *g* to fuel for engines to complete domain-specific functions, but that provides diminishing returns in increased engine efficiency as incremental increases are made. Detterman and Daniel (1989) also claimed that if "central processes are deficient, they limit the efficiency of all other processes in the system. So all processes in subjects with deficits tend to operate at the same uniform level. However, subjects without deficits show much more variability across processes because they do not have deficits in important central processes" (p. 358).

Thurstone's (1938) sampling theory, another common way for researchers to frame this phenomenon, posits that low ability individuals have fewer cognitive resources on which to base their behaviors, and that high ability individuals have more cognitive resources, such that complex behaviors can only be expressed by those with the capacity to do so (i.e., higher ability individuals). This theory, too, is compatible with the ideas of Spearman (1927) and Detterman and Daniel (1989), as it suggests that, at lower ability levels, fewer cognitive resources may constrain a wide range of different behaviors.

Methodological considerations when studying SLODR

Traditionally, SLODR has been measured by dividing samples into high and low ability groups and either comparing the intercorrelations of different tasks or the proportion of variance accounted for by a common (*g*) factor (Tucker-Drob, 2009). In the classic Detterman and Daniel (1989) study, intercorrelations were compared between intellectually disabled individuals and their same-age college-going peers, as well as low and high IQ high school students, on the

WAIS-R. The intellectually disabled and low IQ groups had substantially higher intercorrelations than the college and high IQ groups. Detterman and Daniel (1989) also divided the standardization sample from the WAIS-R and WISC-R into five ability groups; these ability groups could not be created by divided on Full Scale IQ, since that would introduce spurious negative correlations among subtests (since those subtests contribute to the calculation of FSIQ). Instead, the groups were created based on the scores from a single subtest (Vocabulary), as well as a replication subtest (Information). In the lowest ability groups (IQ equivalent of less than 78), the average subtest intercorrelations were about .7, while in the highest ability groups (IQ equivalent of greater than 122), the intercorrelations were about .4.

Since this investigation, different researchers have taken various approaches to studying this phenomenon. Deary et al. (1996), for example, created similarly distributed subsamples of low and high IQ 14- to 17-year-olds from a sample of 10,500 school-age children, and found that the first principal component accounted for about 2% more variance in the high IQ groups. Abad, Colom, Juan-Espinosa, and García (2003) used this same methodology on a sample of 3,430 university applicants, as well as the Spanish standardization sample of the WAIS-III (n = 823). In the university sample, a single common factor accounted for about 2% more variance in the low ability groups than in the high ability groups; in the WAIS-III sample, this difference was about 12%. The same methodology was also applied by Kane, Oakland, and Brand (2006), who created high and low IQ groups from the Woodcock-Johnson Psycho-Educational Battery—Revised (WJ-R) standardization sample (n = 6,359, ages 2 to 95). A single common factor accounted for 52% of the variance in the test scores in the low IQ group, while the variance explained in the high IQ group was only 29%.

In M. R. Reynolds and Keith (2007), the standardization sample of the Kaufman Assessment Battery for Children—Second Edition (KABC-II) was used (ages 6 to 18); high (n = 594) and low (n = 593) ability groups were selected based on Anderson-Rubin factor scores obtained from principal factor analysis in order to minimize selection effects. Here, the researchers used confirmatory factor analysis to fit two hierarchical ability models; one model was a higher-order model, in which *g* is at the apex and imposes a structure on the covariances of the first-order factors. This model is consistent with CHC three-stratum theory, and allows for the higher-order *g* factor to influence subtest performance indirectly via the broad abilities. They also fit a bifactor model, in which there are fewer mathematical constraints and the *g* factor influences subtest performance directly. An illustration of these models (which comes from M. R. Reynolds & Keith, 2007) is found in Figure 1. It is worth noting that the Gf factor in the bifactor model is underidentified as it only has two indicators; M. R. Reynolds and Keith (2007) did not include a Gf factor in their bifactor model for the actual analyses.

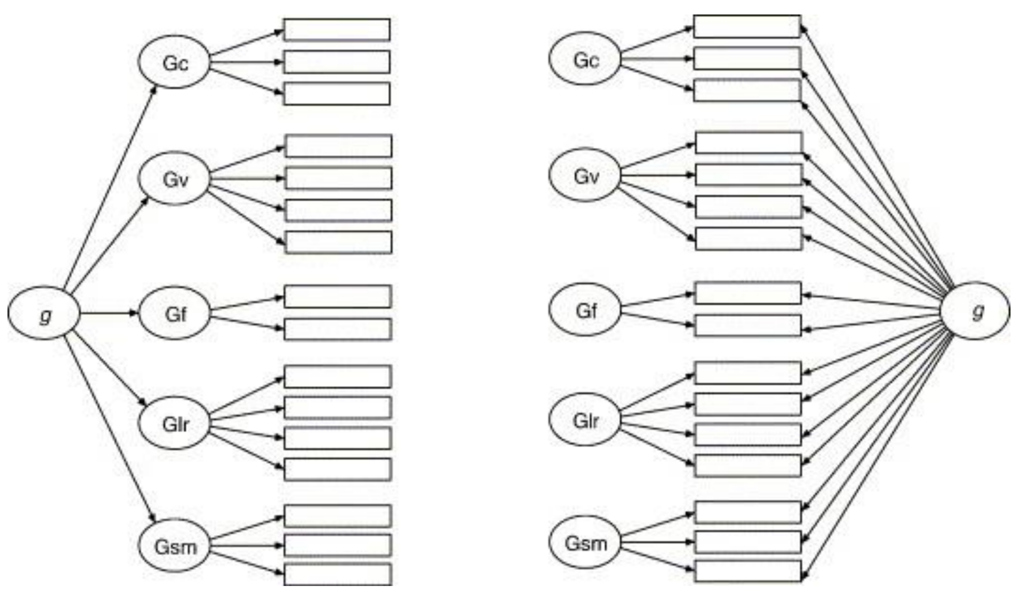

Figure 1. Two hierarchical conceptions of intelligence: the higher-order model is on the left and the bifactor model is on the right. From M. R. Reynolds & Keith (2007).

In the higher order model investigation, *g* explained more variance, on average, in the subtests for the low ability group (.16) compared to the higher ability group (.07), suggesting that "*g* is less important in explaining individual variation in performance on cognitive tests for higher ability children," just as SLODR would predict. Interestingly, the average proportion of variance explained in the subtests by the broad abilities did not change across the low and high ability groups (.22 and .23, respectively). The average proportion of residual, or unexplained, variance was .62 for the low ability group and .70 for the high ability group. Taken together, this finding suggests that the smaller amount of variance explained by *g* in higher ability individuals is not explained adequately by the broad abilities, but by a greater proportion of unique variance than lower ability individuals. In other words, SLODR acts on *g* and not on the broad abilities (even though, as is pointed out, the contribution of *g* to various broad abilities is lower in the high ability group). These findings were nearly identical in the bifactor model, suggesting that SLODR operates similarly for either hierarchical model.

M. R. Reynolds and Keith (2007) also tested SLODR with the higher-order model by using the model with all first-order intercepts invariant (except for Gc, as it was determined that the low and high ability groups differ in Gc beyond their difference in *g*), testing whether the variances in *g* differed across ability groups. This model fit significantly worse than the specific factor means model, and the high ability groups showed significantly lower variability in *g* than the low ability group. This difference in variances suggests that *g* is "more homogenous in the high ability group," as SLODR would predict. When M. R. Reynolds and Keith (2007) analyzed the bifactor model, similar findings occurred. Broad ability factor variances were set to be invariant across groups, but the model did not fit significantly worse than the previous scalar invariant model (in which equality constraints were added to the subtest intercepts). When the *g* variances were constrained to be equal, however, the model fit was significantly worse. This finding, too, suggests that SLODR operates at the level of *g*, but not at the level of the broad abilities. Furthermore, SLODR was not produced primarily by subtests with low *g* loadings in either the higher-order or the bifactor model, meaning that SLODR is not a function of the magnitude of *g* loadings.

M. R. Reynolds and Keith's (2007) finding that SLODR operates on *g* and not on the broad abilities seems to be inconsistent with the findings in Carlstedt (2001), who found that, at higher ability levels, the contribution of Gc and Gv increased. Keith and Reynolds (2007) posited that these inconsistencies may be due to differing age ranges used in each study, the different number of groups used (Carlstedt, 2001, used more than two groups), or because Keith and Reynolds used multiple broad abilities as first-order factors.

The finding that *g* explains less variance in the broad abilities across both the high and low ability groups was also noted in, for example, Kane and Brand (2006), who found lower broad ability loadings on *g* in the high ability group. Keith and Reynolds (2007) noted that this sort of finding fits nicely into Anderson's (1992) theory of minimal cognitive architecture, in which a general processing mechanism is not allowing for differentiation in low ability groups because it is working harder on all cognitive tasks. Interestingly, M. R. Reynolds and Keith (2007) found a rather large decrease in Gf variance explained by *g* in the high ability group, a finding which seems to be in opposition to many researchers (e.g., Gustafsson, 1984) who have noted that Gf and *g* are often indistinguishable from one another. M. R. Reynolds and Keith (2007) explained this finding by stating that the KABC-II may in fact not represent Gf as strongly as other measures might have, a finding echoed by M. R. Reynolds, Keith, Fine, Fisher, and Low (2007).

Der and Deary (2003) were the first study to use nonlinear models to examine SLODR; polynomial regressions were used to predict scores on a test of verbal and numeric reasoning (Part I of the Alice Heim 4 [AH4]) from simple and complex reaction time tasks. The researchers posited that SLODR was perhaps better understood nonlinearly; that is, the relationship between ability level and *g* saturation is not a linear function and that the slope would be steeper at lower ability levels. Indeed, their findings confirmed that a quadratic model better described the relation between the reasoning measure and the simple reaction time task in the direction predicted by SLODR.

Further complicating the study of SLODR is the finding that the phenomenon may not operate invariantly across the age span. In Facon (2006), the French standardization sample of the WISC-III was split into three age groups (total sample included ages 7 to 15), and then each age group was split into high and low ability groups. By comparing the median intercorrelation between subtests for each subgroup, he found that the greatest difference between high and low ability groups was in the oldest group pairing. Facon (2004) compared high and low ability groups in children aged 4 to 9 and found that the intercorrelations were not significantly different from one another. He therefore concluded that SLODR must not manifest until later in the age span, and that age may be an important variable in the study of the phenomenon. Another study concluded the same thing using longitudinal sampling of British twins at ages 7, 9, and 10; the first principal component explains more variance in *g* in the low ability group at every age and in both sexes separately, but, except for 7-year-old females, these differences between the low and high ability groups were not significant (Arden & Plomin, 2007).

As is noted in Deary et al. (1996) and, more recently, in Tucker-Drob (2009), the use of arbitrary criteria (i.e., test score cutoffs) is potentially problematic. SLODR may not be a linear hypothesis at all, as was noted by Der and Deary (2003). Additionally, arbitrarily dividing a continuous variable may incur possible artifacts of that division (e.g., unequal ranges of scores). Tucker-Drob (2009) attempted to simultaneously address the issues of nonlinearity and age differentiation in a series of models; while he consistently found support for ability differentiation (i.e., SLODR), he was less able to find meaningful evidence of age modification of ability differentiation. Most notably, his model of ability differentiation was nonlinear and covered nearly the entirety of the lifespan (n = 6,273, ages 4 to 101). A similar approach to the problem will be used in the present study. Given the lack of support for age-related variables in Tucker-Drob's (2009) study, the present study will not attempt to include age variables or modifiers.

Race differences and g

Evidence suggests that most well-developed, modern intelligence tests lack cultural bias on the basis of race (Valencia & Suzuki, 2001). On well-known measures such as the Woodcock Johnson Cognitive battery and the Wechsler Intelligence Scale for Children—Fifth Edition, there is evidence to support equivalent factor structures across racial groups (Edwards & Oakland, 2006; Scheiber, 2016).

For the Differential Ability Scales, Second Edition (DAS-II), Trundt (2013) conducted a study in which a multi-group confirmatory factor analysis was used to test for measurement invariance across Asian, Black, Hispanic and White children and adolescents using the standardization sample. This methodology tested whether criteria for increasingly strict levels of invariance were met across groups. For the Asian and Hispanic subsamples (as compared to the White subsample), all three levels of invariance were supported, indicating that the underlying constructs measured by the DAS-II are similar between these groups. However, there were inconsistent findings between the Black and White subsamples. Comparisons were made between the groups twice, once with an initial comparison group (White 1) and again with a replication sample (White 2). When the Black subsample was compared to the White 1 subsample, configural and metric invariance was supported, meaning that the configuration of subtests and factors is consistent across groups. Full intercept invariance was not supported, though, and partial intercept invariance models were explored. By allowing the intercept of the Digits Forward subtest to vary across groups, model fit was significantly improved, indicating that the Digits Forward subtest on the DAS-II may be differentially more difficult for one group (White 1) than the other group (Black). This subtest does not contribute to the calculation of the overall general cognitive ability score; it contributes only to the diagnostic cluster for working memory. Furthermore, this finding of partial intercept invariance was not supported in the replication sample; in fact, the change in model fit at the intercept invariance level for the initial comparison group (White 1) was just barely significant (.011 value at a .01 cutoff).

The Trundt (2013) study concluded by stating that the subtests from the DAS-II do not appear to show any evidence of construct bias across any of the groups included in the study. The present study will use the same sample as this analysis.

***Sex differences and* g**

Intelligence tests are carefully constructed so that there will be no average overall difference between sexes (Brody, 1992). Most researchers have found that on these tests, there are no sex differences (Camarata & Woodcock, 2006; Colom, Juan-Espinosa, Abad, & García, 2000; Deary, Strand, Smith, & Fernandes, 2007; Deary, Thorpe, Wilson, Starr, & Whalley, 2003; Jensen, 1998; Mackintosh, 1996; Spinath, Spinath, & Plomin, 2008; van der Sluis et al., 2006), though some researchers have found either a small advantage for males (Deary, Irwing, Der, & Bates, 2007; Jackson & Rushton, 2006; Lynn, Raine, Venables, Mednick, & Irwing, 2005; Nyborg, 2005) or small advantage for females (Harnqvist, 1997; Rosén, 1995). One review found that these differences can be explained by non-representative samples, as well as greater male variance in general cognitive ability (Dykiert, Gale, & Deary, 2009), a finding that was confirmed by Hunt and Madhyastha (2008).

Although there is great debate about the nature and extent of sex differences in various cognitive abilities, the reality is that these differences may just depend on the test or measurement used (Halpern, Beninger, & Straight, 2011; Halpern & LaMay, 2000). One commonly replicated finding is that females tend to perform better on tests of verbal abilities, especially those that include long-term memory (Hedges & Nowell, 1995; Johnson & Bouchard, 2007) and processing speed components (Camarata & Woodcock, 2006; Hedges & Nowell, 1995; Keith, Reynolds, Patel, & Ridley, 2008; van der Sluis et al., 2006). Males, however, tend to perform better on tasks related to visuospatial working memory and fluid reasoning, especially as they relate to scientific and mathematical domains (Hedges & Nowell, 1995; Torres et al., 2006; Voyer et al., 1995). In Jensen (1998), tests that "load heavily on *g*," but that were not normed to eliminate sex differences, were analyzed. Jensen found no mean *g* differences, noting that males tended to do better in some areas, while females did better in other areas.

Keith, Reynolds, Roberts, Winter, and Austin (2011) analyzed the DAS-II standardization sample (ages 5 to 17) in order to test for sex differences in overall *g* and in the broad ability factors; this study used multi-group mean and covariance structural equation modeling in order to test these questions across age. Although no significant differences existed on the *g* factor or on the Gc factor, a few small advantages were discovered within the broad abilities. Girls performed better on tasks related to processing speed (Gs) and on free-recall memory, a narrow ability of long-term retrieval (Glr). Younger girls performed better on short-term memory (Gsm) tasks. Boys, as predicted, showed an advantage in visual-spatial ability tasks.

Another notable sex difference is that males tend to be overrepresented in the tails of the *g* distribution. That is, there are more males than females in the highest percentiles of *g* and in the lowest (DeFries & Gillis, 1993; Halpern, 2012). Males are more likely to be diagnosed with learning disabilities, intellectual disabilities, and language disorders (Halpern, 2012; Henning-Stout & Close-Conoley, 1992). One literature review placed the ratio of males to females at 3.6 to 1 across several categories of intellectual disability (Volkman, Szatmari, & Sparrow, 1993). This phenomenon may contribute to findings that support higher male intelligence; because test samples collected in school settings are not typically given to those with intellectual disabilities, the greater number of males at the lower end of the spectrum are often omitted (Halpern et al., 2011). This phenomenon may also help explain why, in Keith et al. (2011), boys showed larger variances for several broad abilities on the DAS-II; these differences, however, were not statistically significant.

While Spearman's Law of Diminishing Returns has been studied extensively, and invariance of measurement across racial/ethnic and sex groups has been studied extensively, no one has yet studied invariance of SLODR across racial/ethnic and sex groups.

Proposed Research Study

Statement of Problem/Purpose

Despite strong research support for Spearman's Law of Diminishing Returns (SLODR), no research study has attempted to determine whether SLODR operates differentially by race/ethnicity or by sex. This topic was illuminated in M. R. Reynolds (2013), in which the author noted that a study limitation was that "it is unknown whether or how SLODR operates differently in different subgroups, for example, related to sex or ethnicity." Even though most previous research inquiries have used representative samples, the question of subgroup differentiation is empirically unique.

As with any topic related to the field of intelligence, the stakes are high: intelligence testing is perhaps the most important aspect in determinations of special education eligibility and gifted program placement (Holdnack & Weiss, 2006; Kaufman & Harrison, 1986). It is therefore crucial that the nature of intelligence is well-understood, as theoretical models influence the construction and interpretation of those measures. The more well-understood SLODR becomes, the more accurately practitioners can interpret test findings; for example, the importance of broad ability scores may be more important when making decisions about giftedness curricula than they would be during decisions about placement into special education programs.

Furthermore, the study of SLODR has often relied on linear methodologies to establish (or disprove) its existence; by using a nonlinear modeling approach to this question, a more descriptive and better-fitting model can be used to understand the nature of SLODR in each of the subgroups studied.

Research Questions and Hypotheses

The research questions are:

1. Does Spearman's Law of Diminishing Returns (SLODR) operate differentially by race/ethnicity?
2. Does Spearman's Law of Diminishing Returns (SLODR) operate differentially by sex?

Considering the general lack of research on this topic, it is difficult to anticipate hypotheses based on prior studies. For Research Question 1 (RQ1), my hypothesis is the null; that is, that SLODR will not operate differentially by race/ethnicity. The DAS-II has demonstrated invariant factor structures across race/ethnicity subgroups (Trundt et al., 2017); in fact, most modern intelligence tests show a lack of cultural bias on the basis of race (Valencia & Suzuki, 2001). There is no reason, therefore, to anticipate a drastic departure from this finding in the present study.

For Research Question 2 (RQ2), my hypothesis is also the null; that is, that SLODR will not operate differentially by sex. This hypothesis is more difficult to predict, as males tend to be overrepresented in the tails of the *g* distribution (DeFries & Gillis, 1993; Halpern, 2012). This phenomenon could perhaps make it so that the "steeper slopes" found in the lower end of ability in the Der and Deary (2003) study are steeper for males. In the Facon (2004) study, male and female children demonstrated no significant differences in the amount of variance explained by *g* in the first principal component; however, in that same study, SLODR was not supported, as the ages of the participants were (presumably) too low. Because there have not been any clear research inquiries about sex differences and SLODR, the null hypothesis of no differences in SLODR across sexes seems plausible.

Method

Instrumentation

The Differential Ability Scales, Second Edition (DAS-II, Elliott, 2007) is a measure of cognitive abilities. The test is administered individually and is standardized for children and adolescents between the ages 2:6 and 17:11. The current study will focus on children between the ages of 7:1 and 17:11. The DAS-II is a popular measure for assessing children and adolescents because of its ease of use, as well as its non-verbal composite score availability (Dumont, Willis, & Elliott, 2008).

The DAS-II includes two separate batteries: the Early Years Battery for ages 2:6 to 6:11, and the School-Age Battery for ages 7:0 to 17:11. The Early Years Battery is further divided into Lower Level (2:6 to 3:5) and Upper Level (3:6 to 6:11) batteries. The present study will focus on children and adolescents that participated in the standardization of the School-Age Battery (ages 7:0 to 17:11).

The DAS-II yields an overall composite score called General Conceptual Ability, or GCA. This overall composite score summed from three cluster scores: Verbal Ability, Nonverbal Reasoning Ability, and Spatial Ability. Two additional cluster scores are provided (Working Memory and Processing Speed); however, these are not included in the calculation of the GCA. The DAS-II also includes five diagnostic subtests that do not contribute to the calculation of any cluster score. One of these diagnostic subtests (Phonological Processing) is only given to children up to age 12:11, and is thus not included in the present study. For the purposes of this study, the cluster scores will not be used; rather, the individual subtests will be used to estimate factors corresponding to CHC broad abilities. Overall, the DAS-II School-Age Battery includes fourteen unique subtests that are given to all children between the ages of 7:0 and 17:11. A description of each of these subtests is provided in Table 2.

Table 2

Description of the DAS-II Subtests Included in the School-Age Battery (Ages 7:0 to 17:11)

Subtest	Broad CHC Ability	Description
Word Definitions	*Gc*	Child tells the meaning of words
Verbal Similarities	*Gc*	Childs indicates how three named things are similar
Matrices	*Gf*	Child selects a figure that completes a 2x2 or 3x3 matrix
Sequential and Quantitative Reasoning	*Gf*	Child determines which object completes a series of objects or numbers
Pattern Construction	*Gv*	Child replicates designs created by the examiner or pictured using wooden blocks, plastic blocks, or flat squares
Recall of Designs	*Gv*	Child re-creates a drawing after viewing it for 5 seconds
Recognition of Pictures	*Gv*	Child is shown a picture of one or more objects for 5 seconds and then selects the previously viewed object(s) from a second pictorial array that includes distracters
Recall of Objects– Immediate	*Glr*	Examiner teaches the names of 20 objects; the child recalls as many objects as possible in 40–45 seconds immediately after card is removed
Recall of Objects– Delayed	*Glr*	Child recalls as many objects as possible 10–30 min after initial exposure, in 45 seconds, without viewing them again
Recall–Digits Forward	*Gsm*	Child repeats spoken digits
Recall–Digits Backwards	*Gsm*	Child repeats spoken digits in reverse order
Recall of Sequential Order	*Gsm*	Child recites body parts and other objects in sequential order stated by the examiner
Speed of Information Processing	*Gs*	Child quickly determines the circle that contains the greatest number of squares or the highest number in each row
Rapid Naming	*Gs*	Child indicates the color, object, or the color and object of pictures presented by the examiner

Information provided in the DAS-II manual indicates adequate to strong evidence of reliability (Elliott, 2007). In the overall standardization sample, test-retest reliability coefficients for all scores (subtest, cluster, and GCA) ranged from .63 to .91. Average internal consistencies for each cluster ranged from .87 to .96, indicating excellent internal reliability. Average internal consistencies for individual subtests range from .77 to .95, also indicating adequate to excellent internal reliability. External research further indicates that the DAS-II standardization data has internal structure validity, meaning that it appears to measure the constructs it intended (Canivez & McGill, 2016; Keith, Low, Reynolds, Patel, & Ridley, 2010; Keith et al., 2011). Trundt, Keith, Caemmerer, and Smith (2017) found a lack of construct bias among African-American, Asian, Hispanic, and Caucasian children within the standardization sample.

Participants

Participants will be selected from the DAS-II standardization sample, which is stratified according to age, sex, race/ethnicity, parent education level, and geographic region based on the 2005 U.S. Census. Table 3 provides demographic characteristics of the overall standardization sample (age 4:0 to 17:11). The sample for this study, however, will only include all children and adolescents in the overall standardization sample between the ages of 7:0 and 17:11. The sample will be age-limited as these children were administered a common battery of tests.

Table 3

Demographic Characteristics of the DAS-II Standardization Sample, Age 4 to 17

Variable	N
Total Sample	2,952
Sex	
Male	1,476
Female	1,476
Race/Ethnicity	
White	1,852
Hispanic	496
Black	458
Asian	115
Other	29
Parent Education (Average Years of Schooling)	
≤8	126
9 to 11	272
12 (High school degree or equivalent)	796
13 to 15 (Some college or 2-year degree)	959
≥16 (college or graduate degree)	799

Because the selected sample will include a much larger number of Caucasian participants, a random sample of Caucasian participants will be selected that is equivalent in size to the next largest subsample. For the sex differences research question, all males and females will be compared.

Analyses and Expected Results

Using the standardization data from the DAS-II, separate multigroup higher-order models (by race/ethnicity and by sex) will be created. A higher-order model approach was chosen because it will allow for the use of only two indicators on multiple broad abilities (J. C. Anderson & Gerbing, 1988).

In M. R. Reynolds, Hajovsky, Niileksela, and Keith (2011), a higher-order model was created to test for SLODR in the DAS-II standardization sample. Two models were created in this study, one for ages 5 to 8 and another for ages 9 to 17. The latter model includes the same subtests that will be included in the current study. Figure 3 shows the within-class confirmatory factor analysis model from this study. Model fit for this model was very good: $\chi^2 = 423.97$, RMSEA = .053, CFI = .967, SRMR = .037.

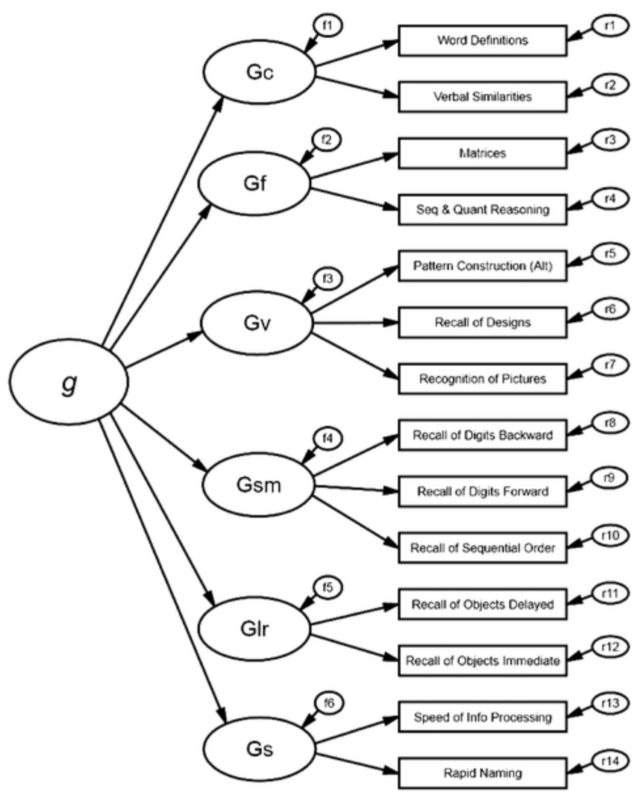

Figure 2. Within-class confirmatory factor analysis model for ages 9-17 on the DAS-II from M. R. Reynolds, Hajovsky, Niileksela, and Keith (2011).

With the current DAS-II School-Age sample, a Ga factor would be not be indicated by any subtest and will therefore not be included in this analysis. The Gv factor and the Gsm factor are indicated by three subtests, while the rest of the broad abilities are indicated by two. Keith et al. (2010) also noted in their study of the DAS-II factor structure that the Glr factor's indexing subtests (Recall of Objects, Immediate and Delayed) may reflect "a less broad ability than long-term retrieval," but for the purposes of this study, the Glr factor will be retained as is.

The higher-order model will be separately applied to groups divided by race and by gender for each of the six broad abilities included in the DAS-II validation model. As mentioned previously, I will select a Caucasian subsample at random to compare each of the race/ethnicity subsamples (African-American, Asian, and Hispanic). Likewise, the model will be applied separately to each of the two sex-divided subsamples.

The analyses will be conducted using Mplus (Muthén & Muthén, 2017) and will follow a procedure based in part on the method used by Tucker-Drob (2009). The first step will be to establish configural invariance across both the male and female groups and each of the groups divided on race/ethnicity. This test involves setting the factors and pattern of loadings to be the same for all groups; in other words, the model structure will be the same across groups, but the parameters to be estimated (means, intercepts, variances, etc.) will be free to vary. Means of latent variables will be fixed to zero, and the intercepts for the measured variables will be freely estimated for all groups. Configural invariance should be able to be established for the groups divided by race/ethnicity as it was also established by Trundt (2013) on the DAS-II overall standardization sample. Configural invariance should also be able to be established in the male and female groups, as Keith et al. (2011) demonstrated in the DAS-II standardization sample.

Support for configural invariance (and other forms of invariance) is assessed by comparing various indicators of model fit, such as the Likelihood Ratio Test, or the difference in chi-square between two models ($\Delta\chi^2$; Cheung & Rensvold, 2002); the χ^2 statistic is highly sensitive to sample size (Brannick, 1995), and therefore may not be practically useful for comparing competing models. Cheung and Rensvold (2002) suggested using comparative fit index difference (ΔCFI) across competing models. CFI is not affected by sample size or model complexity. The critical value for ΔCFI is -0.01, so a decrease in CFI greater than 0.01 across models suggests a lack of invariance across groups at that level.

The next step will be to establish metric invariance across both the male and female groups and each of the groups divided on race/ethnicity (assuming configural invariance is supported). This is accomplished by running a model where the factor loadings are constrained to be equal across groups, but the intercepts are still allowed to differ between groups. This procedure tests whether the relation of the subtests (measured variables) to the broad ability scores (latent variables), or the scale of the latent variables, is the same across groups. As with configural invariance, metric invariance should be able to be established for the groups divided by race/ethnicity in the current study as it was also established by Trundt (2013) on the DAS-II overall standardization sample. Metric invariance should also be able to be established in the male and female groups, as Keith et al. (2011) demonstrated in the DAS-II standardization sample. This too will be assessed using ΔCFI.

Intercept invariance will then be tested by constraining first-order factor intercepts to zero for one group only (e.g., female and White: one for each comparison). The second-order means will be constrained to zero for both groups. Any differences in means on the measured variables will therefore be due to differences in latent means. If intercept invariance is supported, then differences in latent means should account for any differences across the groups in subtest scores.

Residual invariance will then be assessed by keeping the model from the intercept invariance model and then setting subtest residual variances (and covariances, if any) equal across groups. If residual invariance is accomplished, that would mean that any differences in subtest scores are the result of the latent variables.

As the model in question is higher-order, structural invariance should be established between groups. In order to do this, a further series of steps must be taken. Firstly, the residual invariance model will be used (all future invariance steps are additive), and then second-order loadings will be set equal across groups. The second-order means will be set to zero for all groups. Invariance at this step would mean that the second-order factor (in other words, *g*) has the same meaning across groups.

Next, first-order intercepts will be set to zero for both groups. The second-order mean would be set to zero for only one group. Invariance at this step would mean that there is no mean differences across groups on the first-order latent variable (*g*). First-order unique variances will then be set equal across groups. Invariance at this step would mean that the unique aspects of the first-order factors (not explained by second-order factors) are the same across groups. Second-order variances will then be set equal groups; invariance here would mean that the second-order factors are equally variable across groups. Lastly, second-order factor means will be set equal across groups. Invariance at this step would mean that the means are equal across groups on the second-order factor. If, through all of these steps, invariance is continually supported, structural invariance of the second-order loading would be established. It is reasonable to assume that metric and structural invariance will be supported for all between-group comparisons (sex and race/ethnicity), as Trundt (2013) and Keith et al. (2011) found no evidence of construct bias by race or sex, respectively.

Assuming invariance has been supported, an interaction component (`gxg`) will then be created in Mplus using a `g XWITH g` command (Maslowsky, Jager, & Hemken, 2015). More specifically, the "`XWITH`" procedure in Mplus tests the interaction between two latent variables; in this case, the variables are just the higher-order *g* factor with itself (in other words, g^2). Then, the `gxg` term will be regressed on each broad ability (i.e., "`Gc ON gxg`").

When the "`XWITH`" command is used in Mplus, standard model fit indices such as CFI, RMSEA, or χ^2, are not computed. Instead, a two-step method must be used to assess overall fit for each interaction model (Klein & Moosbrugger, 2000). First, fit statistics such as CFI, RMSEA, or χ^2 are obtained from the model without the interaction (or Model 0). Then, the relative fit of Model 0, which here is the null model, is compared to Model 1 (the one with the interaction) using a log-likelihood ratio test. This test is used to determine whether Model 0, which is more parsimonious, represents a significant loss in fit compared to a more complex Model 1 (Satorra & Bentler, 2010). If Model 0 represents a significant loss in fit compared to Model 1, then it can be concluded that Model 1 is also a well-fitted model. The test-statistic for a log-likelihood ratio test (or *D*) is calculated as follows (Maslowsky et al., 2015):

$$D = -2[(\text{log-likelihood for Model 0}) - (\text{log-likelihood for Model 1})]$$

Values of *D* are approximately distributed as χ^2. Maslowsky et al. (2015) gives a fuller explanation of the procedure of comparing Models 0 and 1 using such a test.

Assuming Model 1, or the interaction model, is a well-fitted model for each of the groups, then it will be used to make comparisons across males and females and across each of the groups divided on race/ethnicity. This is a valid assumption, as Tucker-Drob's (2009) g^2

model showed a significant improvement in fit from a linear model, thus supporting the ability differentiation hypothesis (i.e., SLODR).

The same model fit comparison issues will be present during the group comparison stage, though the two-step model proposed by Klein and Moosbrugger (2000) will not be possible because both models will include interaction terms. In order to compare models using a log-likelihood ratio test, Model 0 must lack an `XWITH` term. Therefore, a more qualitative approach to assessing invariance must be used for between-group comparisons by sex and by race. This can be accomplished by plotting the model results for each group. The model-implied relations between the score on the *g* factor and each broad ability can be plotted, similar to the way in which Tucker-Drob (2009) plotted his model output (see Figure 3). In this illustration, multiple broad abilities are plotted at once; however, plotting multiple broad abilities for multiple groups would be overly complicated to qualitatively assess. This issue can be alleviated by plotting one broad ability at a time with each group's *g*-broad ability relation being compared across males and females and across each race/ethnicity group.

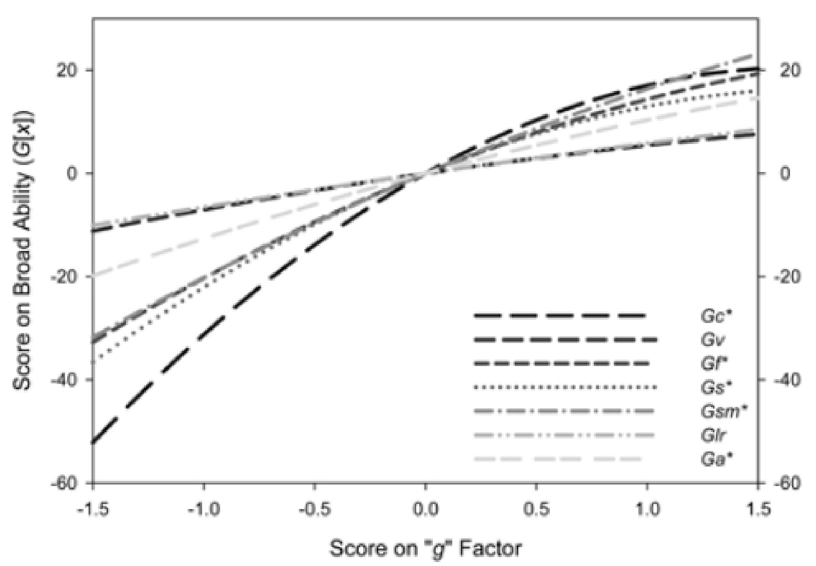

Figure 3. Model-implied relations between the score on the *g* factor and each broad ability for the adult sample. From Tucker-Drob (2009).

A formal way to measure invariance between two models with a latent interaction term does not yet exist other than to qualitatively measure differences using an illustration. Nevertheless, these illustrations can be illuminating if, as hypothesized, the plots of each *g*-broad ability relation do not appear to differ across groups. If there is indeed clear visual evidence of a difference across groups, a further examination of this invariance can be conducted.

Discussion

Summary

The purpose of this study is to determine whether Spearman's Law of Diminishing Returns operates differentially by race/ethnicity or by sex. As was discussed in the Integrative Analysis section, the extant research on SLODR supports the existence of the phenomenon. Various methods have been used over time to measure SLODR, including a comparison of intercorrelations, principal component analysis, and, more recently, confirmatory factor analysis and nonlinear modeling. Likewise, invariance of measurement between racial/ethnic groups and sex groups has been demonstrated repeatedly. Never before, though, has anyone measured the invariance of SLODR across racial/ethnic and sex groups.

In the present analysis, separate multigroup higher-order models (by race/ethnicity and by sex) will be created in which linear and nonlinear (squared) loadings will be estimated.

Limitations

The present study is limited firstly by its use of a single test measure, thereby constraining the analysis to the subtests used in the DAS-II School Age Battery. Though the psychometric properties of the DAS-II are sound, and the test is aligned with CHC theory, many of the broad abilities are indicated only by two subtests.

Another limitation is that no Ga factor was included; however, it is reasonable to expect that Ga would operate in a similar way as the other broad abilities (that is, *g* would explain less variance in Ga in high-ability groups).

Though the racial/ethnic subsample sizes are comparatively quite large (when looking at other intelligence-based research), the fact that the White subsample is much large than the other groups is an issue. This is partially alleviated in the study by randomly selecting out a White subsample to match the size of the next-largest subsample, but the disparate subsample sizes may make comparisons based on chi-square model fit indices problematic; chi-square is affected by sample size.

As was elaborated on in some depth in the Analysis section, the lack of model fit indices with an `XWITH` command make answering the research questions difficult using traditional methods. Though qualitative methods of assessing comparative model fit are not ideal, they can still be enlightening as a guide for future research. Nevertheless, a more traditional approach to this research question (which has not yet been studied using any methodology) may be more appealing to those who seek to find specific criteria for invariance.

Future Directions and Implications

Determining whether SLODR operates similarly across race/ethnicity and sex is a necessary step before continuing to do research about the phenomenon. If the null hypothesis turns out to be disproved, replications of the study using different test measures, samples, and methodologies must be completed. Indeed, this is a good next step regardless of the outcome of the present study. Given the importance of measurement invariance across these groups, replications are encouraged. If the null hypothesis is repeatedly supported, then research on SLODR can be conducted with the knowledge that the phenomenon is stable across race/ethnicity and sex groups.

As far as SLODR is concerned, there remain a number of questions about the nature of the phenomenon. As was suggested in the M.R. Reynolds and Keith (2007) article, the variance not explained by *g* in higher-ability individuals is not explained by the broad abilities, either. In that study, subtest-level unique and error variances were different across ability groups, but the precise nature of these differences is unknown. Assessing whether there is more unique variance or more error variance for higher-ability individuals could shed light on the nature of SLODR.

An understanding of SLODR can be important in schools, where the use of intelligence testing is often the central determinant for both special education services and placement into a gifted program. If overall intelligence is indeed a less adequate reflection of a higher-ability student's cognitive capabilities, then perhaps those making determinations about gifted programs should look more closely at broad ability (or cluster) scores to make more informed decisions about a student's particular strengths.

Considering intelligence is a highly predictive latent variable, capable of even predicting future job performance (Schmidt & Hunter, 2004), SLODR might predict that, for higher-ability individuals, *g* would be less predictive than domain-specific broad ability scores.

While it may be simpler to use previous methods of assessing SLODR for this sort of question (i.e., linear factor analytic modeling, or even principal component analysis), the future direction of SLODR research seems to be in nonlinear factor analytic modeling. Nevertheless, an examination of this question from a more traditional methodology may be warranted as a supplement to the present study.

References

Abad, F. J., Colom, R., Juan-Espinosa, M., & García, L. F. (2003). Intelligence differentiation in adult samples. *Intelligence*, *31*(2), 157–166. https://doi.org/10.1016/S0160-2896(02)00141-1

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423. https://doi.org/10.1037/0033-2909.103.3.411

Anderson, M. (1992). *Intelligence and development: A cognitive theory*. Malden: Blackwell Publishing.

Arden, R., & Plomin, R. (2007). Scant evidence for Spearman's law of diminishing returns in middle childhood. *Personality and Individual Differences*, *42*(4), 743–753. https://doi.org/10.1016/J.PAID.2006.08.010

Bartholomew, D. J. (2004). *Measuring intelligence: Facts and fallacies*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511490019

Binet, A., & Simon, T. (1917). *The development of intelligence in children* (Vol. 46). New York: Arno. https://doi.org/10.1097/00005053-191710000-00024

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, *16*(3), 201–213. https://doi.org/10.1002/job.4030160303

Brody, N. (1992). *Intelligence* (2nd ed.). Gulf Professional Publishing.

Burt, C. (1921). *Mental and scholastic tests*. London: P.S. King and Son.

Burt, C. (1949). The structure of the mind: A review of the results of factor analysis. *British Journal of Educational Psychology*, *19*(3), 176–199. https://doi.org/10.1111/j.2044-8279.1949.tb01612.x

Camarata, S., & Woodcock, R. (2006). Sex differences in processing speed: Developmental effects in males and females. *Intelligence*, *34*(3), 231–252. https://doi.org/10.1016/j.intell.2005.12.001

Canivez, G. L., & McGill, R. J. (2016). Factor structure of the Differential Ability Scales–Second Edition: Exploratory and hierarchical factor analyses with the core subtests. *Psychological Assessment*, *28*(11), 1475–1488. https://doi.org/10.1037/pas0000279

Carlstedt, B. (2001). Differentiation of cognitive abilities as a function of level of general intelligence: A latent variable approach. *Multivariate Behavioral Research*, *36*(4), 589–609. https://doi.org/10.1207/S15327906MBR3604_05

Carroll, J. B. (1992). Cognitive abilities: The state of the art. *Psychological Science*, *3*(5), 266–271. https://doi.org/10.1111/j.1467-9280.1992.tb00669.x

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Colom, R., Juan-Espinosa, M., Abad, F., & García, L. F. (2000). Negligible sex differences in general intelligence. *Intelligence*, *28*(1), 57–68. https://doi.org/10.1016/S0160-2896(99)00035-5

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence*, *23*(2), 105–132. https://doi.org/10.1016/S0160-2896(96)90008-2

Deary, I. J., Irwing, P., Der, G., & Bates, T. C. (2007). Brother–sister differences in the g factor in intelligence: Analysis of full, opposite-sex siblings from the NLSY1979. *Intelligence*, *35*(5), 451–456. https://doi.org/10.1016/j.intell.2006.09.003

Deary, I. J., & Pagliari, C. (1991). The strength of g at different levels of ability: Have Detterman and Daniel rediscovered Spearman's "law of diminishing returns"? *Intelligence*, *15*(2), 247–250. https://doi.org/10.1016/0160-2896(91)90033-A

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13–21. https://doi.org/10.1016/j.intell.2006.02.001

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M., & Whalley, L. J. (2003). Population sex differences in IQ at age 11: the Scottish mental survey 1932. *Intelligence*, *31*(6), 533–542. https://doi.org/10.1016/s0160-2896(03)00053-9

DeFries, J. C., & Gillis, J. J. (1993). Genetics of reading disability. In R. Plomin & G. E. McClearn (Eds.), *Nature, nurture and psychology* (pp. 121–145). Washington, DC: American Psychological Association.

Der, G., & Deary, I. J. (2003). IQ, reaction time and the differentiation hypothesis. *Intelligence*, *31*(5), 491–503. https://doi.org/10.1016/s0160-2896(02)00189-7

Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence*, *13*(4), 349–359. https://doi.org/10.1016/S0160-2896(89)80007-8

Dumont, R., Willis, J. O., & Elliott, C. D. (2008). *Essentials of DAS-II Assessment*. John Wiley & Sons.

Dykiert, D., Gale, C. R., & Deary, I. J. (2009). Are apparent sex differences in mean IQ scores created in part by sample restriction and increased male variance? *Intelligence*, *37*(1), 42–47. https://doi.org/10.1016/J.INTELL.2008.06.002

Edwards, O. W., & Oakland, T. D. (2006). Factorial invariance of Woodcock-Johnson III scores for African Americans and Caucasian Americans. *Journal of Psychoeducational Assessment*, *24*(4), 358–366. https://doi.org/10.1177/0734282906289595

Elliott, C. D. (2007). *Differential Ability Scales, Second Edition*. San Antonio, TX: The Psychological Corporation.

Evans, M. G. (1999). On the asymmetry of g. *Psychological Reports*, *85*, 1059–1069. https://doi.org/10.2466/pr0.85.7.1059-1069

Facon, B. (2004). Are correlations between cognitive abilities highest in low-IQ groups during childhood? *Intelligence*, *32*(4), 391–401. https://doi.org/10.1016/j.intell.2004.06.002

Facon, B. (2006). Does age moderate the effect of IQ on the differentiation of cognitive abilities during childhood? *Intelligence*, *34*(4), 375–386. https://doi.org/10.1016/j.intell.2005.12.003

Geffen, G., Moar, K. J., O'Hanlon, A. P., Clark, C. R., & Geffen, L. B. (1990). Performance measures of 16– to 86-year-old males and females on the auditory verbal learning test. *Clinical Neuropsychologist*, *4*(1), 45–63. https://doi.org/10.1080/13854049008401496

Guilford, J. P. (1959). Three faces of intellect. *American Psychologist*, *14*(8), 469–479. https://doi.org/10.1037/h0046827

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*(3), 179–203. https://doi.org/10.1016/0160-2896(84)90008-4

Halpern, D. F. (2012). *Sex differences in cognitive abilities* (4th ed.). New York: Psychology Press.

Halpern, D. F., Beninger, A. S., & Straight, C. A. (2011). Differences in intelligence. In R. Sternberg & S. B. Kaufman (Eds.), *The Cambridge Handbook of Intelligence* (pp. 253–272). Cambridge University Press.

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review*, *12*(2). https://doi.org/10.1023/A:1009027516424

Harnqvist, K. (1997). Gender and grade differences in latent ability variables. *Scandinavian Journal of Psychology*, *38*(1), 55–62. https://doi.org/10.1111/1467-9450.00009

Hedges, L. V., & Nowell, A. (1995, July). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 41–45.

Henning-Stout, M., & Close-Conoley, J. (1992). Gender: A subtle influence in the culture of the school. In F. J. Medway & T. P. Cafferty (Eds.), *School psychology: A social psychological perspective* (pp. 113–135). Hillsdale, NJ: Erlbaum.

Holdnack, J. A., & Weiss, L. G. (2006). IDEA 2004: Anticipated implications for clinical practice—integrating assessment and intervention. *Psychology in the Schools*, *43*(8), 871–882. https://doi.org/10.1002/pits.20194

Holzinger, K. J., & Harman, H. H. (1938). Comparison of two factorial analyses. *Psychometrika*, *3*(1), 45–60. https://doi.org/10.1007/BF02287919

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. *Woodcock-Johnson Technical Manual*, 197–232.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253–270. https://doi.org/10.1037/h0023816

Hunt, E., & Madhyastha, T. (2008). Recruitment modeling: An analysis and an application to the study of male–female differences in intelligence. *Intelligence*, *36*(6), 653–663. https://doi.org/10.1016/J.INTELL.2008.03.002

Jackson, D. N., & Rushton, J. P. (2006). Males have greater g: Sex differences in general mental ability from 100,000 17- to 18-year-olds on the Scholastic Assessment Test. *Intelligence, 34*(5), 479–486. https://doi.org/10.1016/j.intell.2006.03.005

Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.

Jensen, A. R. (1992). Commentary: Vehicles of g. *Psychological Science*, *3*(5), 275–279. https://doi.org/10.1111/j.1467-9280.1992.tb00671.x

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Johnson, W., & Bouchard, T. J. (2007). Sex differences in mental abilities: g masks the dimensions on which they lie. *Intelligence*, *35*(1), 23–39. https://doi.org/10.1016/J.INTELL.2006.03.012

Johnson, W., Carothers, A., & Deary, I. J. (2008). Sex differences in variability in general intelligence: A new look at the old question. *Perspectives on Psychological Science*, *3*(6), 518–531. https://doi.org/10.1111/j.1745-6924.2008.00096.x

Kane, H. D., & Brand, C. R. (2006). The variable importance of general intelligence (g) in the cognitive abilities of children and adolescents. *Educational Psychology*, *26*(6), 751–767. https://doi.org/10.1080/01443410500390897

Kane, H. D., Oakland, T., & Brand, C. (2006). Differentiation at higher levels of cognitive ability: Evidence from the United States. *The Journal of Genetic Psychology*, *167*(3), 327–341. https://doi.org/10.3200/GNTP.167.3.327-341

Kaufman, A. S., & Harrison, P. L. (1986). Intelligence tests and gifted assessment: What are the positives? *Roeper Review, 8*(3), 154–159. https://doi.org/10.1080/02783198609552961

Keith, T. Z., Low, J. A., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2010). Higher-order factor structure of the Differential Ability Scales-II: Consistency across ages 4 to 17.

Psychology in the Schools, *47*(7), 676–697. https://doi.org/10.1002/pits.20498

Keith, T. Z., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2008). Sex differences in latent cognitive abilities ages 6 to 59: Evidence from the Woodcock–Johnson III tests of cognitive abilities. *Intelligence*, *36*(6), 502–525. https://doi.org/10.1016/j.intell.2007.11.001

Keith, T. Z., Reynolds, M. R., Roberts, L. G., Winter, A. L., & Austin, C. A. (2011). Sex differences in latent cognitive abilities ages 5 to 17: Evidence from the Differential Ability Scales—Second Edition. *Intelligence*, *39*(5), 389–404. https://doi.org/10.1016/J.INTELL.2011.06.008

Kelley, T. L. (1928). *Crossroads in the mind of man*. Stanford University Press.

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*(4), 457–474. https://doi.org/10.1007/BF02296338

Legree, P. J., Pifer, M. E., & Grafton, F. C. (1996). Correlations among cognitive abilities are lower for higher ability groups. *Intelligence*, *23*(1), 45–57. https://doi.org/10.1016/S0160-2896(96)80005-5

Lynn, R. (1992). Does Spearman's g decline at high IQ levels? Some evidence from Scotland. *The Journal of Genetic Psychology*, *153*(2), 229–230. https://doi.org/10.1080/00221325.1992.10753717

Lynn, R., & Cooper, C. (1993). A secular decline in Spearman's g in France. *Learning and Individual Differences*, *5*(1), 43–48. https://doi.org/10.1016/1041-6080(93)90025-N

Lynn, R., Raine, A., Venables, P. H., Mednick, S. A., & Irwing, P. (2005). Sex differences on the WISC-R in Mauritius. *Intelligence*, *33*(5), 527–533. https://doi.org/10.1016/j.intell.2005.05.001

Mackintosh, N. J. (1996). Sex differences and IQ. *Journal of Biosocial Science*, *28*(4), 558–571. https://doi.org/10.1017/S0021932000022586

Maslowsky, J., Jager, J., & Hemken, D. (2015). Estimating and interpreting latent variable interactions. *International Journal of Behavioral Development*, *39*(1), 87–96. https://doi.org/10.1177/0165025414552301

Maxwell, A. E. (1972). The WPPSI: A marked discrepancy in the correlations of the subtests for good and poor readers. *British Journal of Mathematical and Statistical Psychology*, *25*(2), 283–291. https://doi.org/10.1111/j.2044-8317.1972.tb00498.x

Muthén, L. K., & Muthén, B. O. (2017). Mplus User's Guide. Los Angeles, CA: Muthén & Muthén.

Nyborg, H. (2005). Sex-related differences in general intelligence g, brain size, and social status. *Personality and Individual Differences*, *39*(3), 497–509. https://doi.org/10.1016/J.PAID.2004.12.011

Reynolds, C. R., & Lowe, P. A. (2009). The problem of bias in psychological assessment. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (4th ed., pp. 332–374). New York: John Wiley.

Reynolds, M. R. (2013). Interpreting the g loadings of intelligence test composite scores in light of Spearman's law of diminishing returns. *School Psychology Quarterly*, *28*(1), 63–76. https://doi.org/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 the level of g? *School Psychology Quarterly*, *26*(4), 275–289. https://doi.org/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*(3), 267–281. https://doi.org/10.1016/j.intell.2006.08.002

Reynolds, M. R., Keith, T. Z., Fine, J. G., Fisher, M. E., & Low, J. A. (2007). Confirmatory factor structure of the Kaufman Assessment Battery for Children--Second Edition: Consistency with Cattell-Horn-Carroll theory. *School Psychology Quarterly*, *22*(4), 511–539. https://doi.org/10.1037/1045-3830.22.4.511

Rosén, M. (1995). Gender differences in structure, means and variances of hierarchically ordered ability dimensions. *Learning and Instruction*, *5*(1), 37–62. https://doi.org/10.1016/0959-4752(95)00002-K

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Scheiber, C. (2016). Is the Cattell–Horn–Carroll-based factor structure of the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V) construct invariant for a representative sample of African–American, Hispanic, and Caucasian male and female students ages 6 to 16 ye. *Journal of Pediatric Neuropsychology*, *2*(3–4), 79–88. https://doi.org/10.1007/s40817-016-0019-7

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*(1), 162–173. https://doi.org/10.1037/0022-3514.86.1.162

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* (pp. 99–144). New York: Guilford Press.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *The*

American Journal of Psychology, *15*(2), 201. https://doi.org/10.2307/1412107

Spearman, C. (1927). *The abilities of man, New York: MacMillan*. New York: MacMillan.

Spinath, F. M., Spinath, B., & Plomin, R. (2008). The nature and nurture of intelligence and motivation in the origins of sex differences in elementary school achievement. *European Journal of Personality*, *22*(3), 211–229. https://doi.org/10.1002/per.677

Stern, W. (1914). *The psychological methods of testing intelligence, volume 1*. Baltimore: Warwick & York.

Stumpf, H., & Jackson, D. N. (1994). Gender-related differences in cognitive abilities: Evidence from a medical school admissions testing program. *Personality and Individual Differences*, *17*(3), 335–344. https://doi.org/10.1016/0191-8869(94)90281-X

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Torres, A., Gómez-Gil, E., Vidal, A., Puig, O., Boget, T., & Salamero, M. (2006). Gender differences in cognitive functions and the influence of sex hormones. *Actas Españolas de Psyiquiatria*, *34*, 408–415.

Trundt, K. M. (2013). *Construct bias in the Differential Ability Scales, Second Edition (DAS-II): A comparison among African-American, Asian, Hispanic, and White ethnic groups*. The University of Texas at Austin.

Trundt, K. M., Keith, T. Z., Caemmerer, J. M., & Smith, L. V. (2017). Testing for construct bias in the Differential Ability Scales, Second Edition. *Journal of Psychoeducational Assessment*, 73428291769830. https://doi.org/10.1177/0734282917698303

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology*, *45*(4), 1097–118. https://doi.org/10.1037/a0015864

Valencia, R. R., & Suzuki, L. A. (2001). *Intelligence testing and minority students: Foundations,* *performance factors, and assessment issues*. Thousand Oaks, CA: Sage.

van der Sluis, S., Posthuma, D., Dolan, C. V., de Geus, E. J. C., Colom, R., & Boomsma, D. I. (2006). Sex differences on the Dutch WAIS-III. *Intelligence*, *34*(3), 273–289. https://doi.org/10.1016/j.intell.2005.08.002

Volkman, F., Szatmari, P., & Sparrow, S. (1993). Sex differences in pervasive developmental disabilities. *Journal of Autism and Developmental Disabilities*, *23*, 579–591. https://doi.org/10.1007/bf01046103

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270. https://doi.org/10.1037/0033-2909.117.2.250

Wolfle, D. (1940). *Factor analysis to 1940*. Chicago: University of Chicago Press.