Journal of School Psychology 93 (2022) 1–27

Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

Examination of differential effects of cognitive abilities on reading and mathematics achievement across race and ethnicity: Evidence with the WJ IV

Daniel B. Hajovsky[a,*], Steven R. Chesnut[b]

[a] Texas A&M University, USA
[b] University of Missouri-Kansas City, USA

A R T I C L E I N F O

Editor: Craig A. Albers

Keywords:
Race/ethnicity
Academic achievement
Test bias
Latent variable
Cognitive ability

A B S T R A C T

There has been little research investigating the predictive validity of modern intelligence tests for racially and ethnically diverse students. The validity of test score interpretation within educational and psychological assessment assumes that test scores predict educationally relevant phenomena equally well for individuals, regardless of group membership (American Educational Research Association et al., 2014; Messick, 1995; Warne et al., 2014). We used multiple group latent variable structural equation modeling (SEM) to investigate Cattell-Horn-Carroll general (*g*) and broad cognitive abilities on reading and mathematics achievement and whether these differed between racial (African American, Asian, and Caucasian) and ethnic (Hispanic, non-Hispanic) children and adolescents within the Woodcock-Johnson IV norming sample ($N = 3127$). After establishing construct equivalence across racial and ethnic groups, supporting the consistent calculation of composite scores regardless of group membership, we then examined the predictive validity of intelligence on achievement. After controlling for parent education, findings suggested two instances of differential predictive relations: (a) general intelligence had larger influences on basic reading skills for Caucasians when compared to Asian peers, and (b) comprehension-knowledge had larger influences on basic reading skills for Asians when compared to Caucasian peers. The overall pattern of findings suggests there is little to no predictive bias with the WJ IV. However, the findings indicate that when latent mean differences exist (after establishing strong factorial invariance), then bias will be introduced into the estimation of regression parameters used to identify differential predictive validity. Thus, even when measurement invariance is supported, differential prediction bias is inevitable when there are mean differences in the scores used as predictors. Future test bias research should consider latent ability differences and how that may impact findings of bias, and possibly, socioeconomic status-related indicators when assessing for measurement or prediction bias in intelligence and achievement tests.

Cognitive ability tests are often included in diagnostic psychological assessments in school and clinical settings (Benson et al., 2019; Jewsbury et al., 2017). As predicting academic achievement continues to be used in educational protocols for guiding the selection,

* Corresponding author at: Texas A&M University, Department of Educational Psychology, 712 Harrington Tower, College Station, TX 77843, USA.

E-mail addresses: dhajovsky@tamu.edu (D.B. Hajovsky), schesnut@umkc.edu (S.R. Chesnut).

https://doi.org/10.1016/j.jsp.2022.05.001
Received 19 July 2021; Received in revised form 28 March 2022; Accepted 30 May 2022

diagnosis, and placement of students in special education programs (Konold & Canivez, 2010; Maki et al., 2015), there is a need to assess the accuracy of modern intelligence tests in predicting achievement for members of different groups to ensure equitable assessment practices. Although there is ample research on cognitive-achievement relations (CARs; e.g., Flanagan et al., 2006; McGrew & Wendling, 2010; Zaboski et al., 2018), there has been relatively less research on the extent that race and ethnicity influence CARs. When analyzing CARs in racially and ethnically diverse children, one needs to consider contextual variables to avoid spurious findings, potentially confounded by systematic differences in social and educational opportunities. For example, children from families of lower socioeconomic status (SES) tend to show disproportionate representation in special education, which can compound issues for children of color who are already marginalized in educational contexts (Morgan et al., 2017; Shifrer et al., 2011; Sullivan & Bal, 2013). Although the reasons for disproportionate representation are likely multifaceted and complex (e.g., environmental living conditions, opportunities, resources), the persisting problems with disproportionality in special education, along with the continued widespread use of modeled discrepancies in CARs to identify individuals with learning disabilities (Benson et al., 2020; Maki et al., 2015), makes it imperative to examine the predictive validity of cognitive abilities on academic achievement between racial and ethnic groups while controlling for differences in SES-related contextual variables.

1.1. Test bias research in intelligence and achievement for children and adolescent youth

The topic of bias in psychological assessment has been a popular subject within intelligence research (e.g., Gould, 1981; Herrnstein & Murray, 1994; Jensen, 1980; Reynolds & Lowe, 2009). There has been considerable and controversial debate around the accuracy of the measurement of intelligence as well as the use of intelligence tests in predicting educational outcomes for members of different groups. This controversy remains despite a general pattern of evidence demonstrating consistent construct measurement (with some minor differences) across different groups and a lack of systematic differences in predictive CARs between majority and minority groups (see Edwards & Oakland, 2006; Konold & Canivez, 2010; Reynolds & Lowe, 2009; Scheiber, 2016a, 2016b; Scheiber & Kaufman, 2015; Trundt et al., 2018; Warne et al., 2014; Weiss & Saklofske, 2020; Woods Jr. et al., 2021). Test bias research can generally be divided into two main areas, one focusing on internal measurement bias (e.g., factor structure, item difficulty) and the other focusing on external criterion prediction bias (e.g., slopes, intercepts; see Reynolds & Lowe, 2009, for a review).

1.1.1. Bias in measuring cognitive ability

The Cattell-Horn-Carrell (CHC) theory of cognitive abilities is a hierarchical model that taxonomizes purported cognitive abilities across three strata, ranging from narrow to general (Schneider & McGrew, 2018). The first stratum encompasses a wide range of narrow cognitive abilities (e.g., Lexical Knowledge, Working Memory, Deductive Reasoning) that are observed or measured variables. The second stratum represents broad (also referred to as group factors) cognitive abilities (e.g., Auditory Processing, Processing Speed) that are identified by the aggregation of associated narrow cognitive abilities from the first stratum (e.g., General Information and Lexical Knowledge narrow abilities are measures of the broad ability Comprehension-Knowledge). The third and final stratum is the broadest level of representation for cognitive ability and is commonly referred to as the *g* factor, or general intelligence. The CHC framework has been instrumental for studying human intelligence, which has been used as a system of cognitive abilities to explain individual differences in reading and mathematics achievement (Caemmerer et al., 2020; Hajovsky et al., 2014; Niileksela et al., 2016; Villeneuve et al., 2019).

The broad cognitive abilities that are purported to be represented at the second stratum, as measured using the Fourth Edition of the Woodcock-Johnson Tests of Cognitive Abilities (WJ IV Cognitive), include Fluid Reasoning (G*f*), Comprehension-Knowledge (G*c*), Visual-Spatial Processing (G*v*), Auditory Processing (G*a*), Short-Term Working Memory (G*wm*), Long-Term Storage (G*lr*), and Cognitive Processing Speed (G*s*; McGrew et al., 2014; Schneider & McGrew, 2018; Schrank, McGrew, & Mather, 2014). Studies using the WJ IV Cognitive have led to mixed results concerning the hypothesized factor structure and its alignment with CHC theory. In one study testing the latent factor structure, Dombrowski et al. (2017) conducted an exploratory factor analysis using the standardization sample correlation matrices and suggested the WJ IV Cognitive constructs are best structured as G*s*, G*wm*, G*c,* and a factor resembling perceptual reasoning, with the strongest evidence for the measurement of *g* at the third stratum (Dombrowski et al., 2017). In a subsequent study using confirmatory factor analyses, Dombrowski et al. (2018) suggested the WJ IV Cognitive is best represented as four factors at the second stratum (i.e., Verbal, Working Memory, Perceptual Reasoning, and Processing Speed), arguing the WJ IV Cognitive should be operationalized according to the prior Wechsler conceptualization of intelligence. Researchers have also examined a CHC structure for invariance for use in predicting reading that was composed of as few as five first-order CHC factors (G*a*, G*c*, G*s*, G*sm*, and G*f*) and a second-order *g* factor (Woods Jr. et al., 2021). Similarly, Niileksela et al. (2016) reported on the WJ IV CHC in predicting achievement across age levels using higher-order factor models consisting of eight different CHC-based broad abilities.

An important step in testing for bias in the measurement of cognitive ability is first establishing whether the constructs are measured similarly across different groups or whether there are inconsistencies in information provided by the cognitive test about the latent variable to be measured (Keith & Reynolds, 2018). Tests of construct bias (i.e., factorial or measurement invariance) involve specifying increasingly restrictive equality constraints on parameters within a latent factor model (e.g., equal factor structures, loadings, intercepts; Keith, 2019). When the factor structure, factor loadings, and item intercepts are proportionally the same across different racial and ethnic groups, then factorial invariance is tenable. Findings of factorial invariance across different groups suggest the calculation of composite scores representative of theoretical constructs are not confounded by differences due to measurement artifacts; thus, supporting the conclusion that scoring is equally meaningful for individuals from different groups. Confirming factorial invariance is a critical step in establishing a lack of construct bias, as different scores for different groups may not necessarily indicate

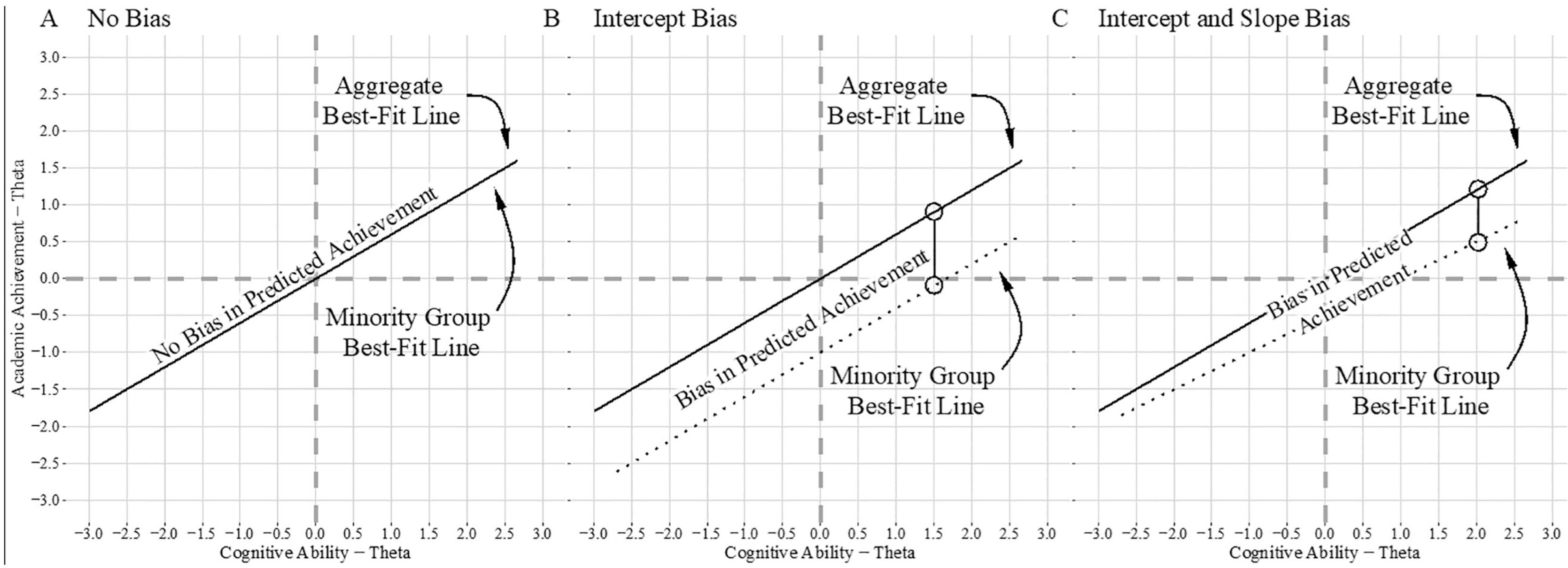

Fig. 1. Examples of alignment with CAR regression trends between groups.

differences in performance, but rather differences in the measurement of the construct. The general pattern of findings within research on modern intelligence tests (e.g., DAS-II, KABC-II, WJ IV, WISC-V) suggests they demonstrate factorial invariance across different racial and ethnic groups (e.g., Scheiber, 2016a, 2016b; Trundt et al., 2018; Woods Jr. et al., 2021), although some research suggests partial invariance when using non-standardization samples (e.g., Picture Span and Figure Weights WISC-V subtests are non-invariant by race; Graves Jr. et al., 2021). These results provide structural validity evidence and instill confidence that scores can be interpreted similarly regardless of group identification by school psychologists. It is worthy of mention that some invariance findings reported in the literature may indicate that an inadequate model did not significantly differ across various groups and therefore was equally inadequate across specified groups (e.g., due to too low or high parameter estimates of Stratum II factors, correlated residuals, or other modifications; Dombrowski et al., 2017, 2018).

1.1.2. Bias in predictive validity of CARs

For scores obtained from educational and psychological assessments to be valid, they must not only measure what they purport to measure but also predict theoretically and empirically associated phenomena consistently for all individuals, regardless of group membership (American Educational Research Association et al., 2014; Messick, 1995; Warne et al., 2014). The issue of predictive validity has been argued to be one of the most important forms of validity for the practical use of tests; the evaluation of predictive bias in CARs has been a focus in modern intelligence research (e.g., Keith, 1999; Scheiber & Kaufman, 2015; Woods Jr. et al., 2021). When examining predictive bias, researchers primarily examine differences in intercepts and slopes in CARs as a function of group membership (Potthoff, 1978; Reynolds & Lowe, 2009). Intercept bias occurs when the CAR is the same for all groups, but the *y*-intercept of the best-fit line for at least one of the groups is significantly different than the others (Konold & Canivez, 2010; Scheiber & Kaufman, 2015). Intercept bias is common; these findings suggest that the regression formula for the reference group (e.g., normative sample) would either consistently under- or over-predict at least one group's performance on the outcome variable (e.g., basic reading skills) along the predictor continuum (e.g., IQ). Alternatively, slope bias occurs when the CAR is not the same for all groups (Keith, 2019; Warne et al., 2014). In slope bias, different groups have different slopes (i.e., rate of change in outcome scores associated with changes in the predictor variable) such that the slope for one group may be steeper than another, leading to under- or over-prediction when groups are combined. A distinguishing feature of slope bias, compared to intercept bias, is that the magnitude of under- or over-prediction depends on the interaction point (i.e., it is not consistent across the continuum of scores on the predictor variable; Keith, 2019). For example, slope bias reflects differential predictive validity coefficients for various groups and that for lower predictive validity coefficients there are greater amounts of prediction error (and thus larger standard errors of estimate). Although slope bias is more difficult to correct, it has occurred less frequently in predictive CARs research (Konold & Canivez, 2010; Scheiber & Kaufman, 2015; Warne et al., 2014; Woods Jr. et al., 2021). Fig. 1 illustrates three examples of bias in CARs. Fig. 1A exemplifies no bias in the regression model. Fig. 1B exemplifies intercept bias by group membership. Fig. 1C exemplifies intercept and slope bias by group membership.

Prior research has suggested mixed findings regarding differential CARs due to race and ethnicity (e.g., Konold & Canivez, 2010; Naglieri et al., 2005, 2007; Scheiber & Kaufman, 2015; Weiss et al., 1993; Weiss & Prifitera, 1995; Woods Jr. et al., 2021). For example, Scheiber and Kaufman (2015) examined the predictive validity of three KABC-II global IQ scores in predicting reading, writing, and math achievement for Caucasian, African American, and Hispanic children. They found a consistent pattern of intercept bias (over-prediction) for African American and Hispanic children's achievement at middle school grade levels. Weiss and Prifitera (1995) found Hispanic students' reading scores were over-predicted when using a common regression equation in the WISC-III and WIAT; however, these were limited to intercept bias with small effect sizes. Conversely, Konold and Canivez (2010) reported findings of slope bias with over-prediction observed for written language for Hispanic students scoring at the lower end of the general ability index (GAI; full scale IQ), and under-prediction for higher scores on the GAI on the WISC-IV and WIAT-II; they noted small effect sizes around the bivariate center with larger effect sizes only at approximately two standard deviations above the mean for the GAI. One potential limitation in these studies is that they used the raw data, failing to control for the effects that social and financial factors have shown to have on both cognitive ability and academic achievement test scores (Weiss & Saklofske, 2020). In addition, prior research examining group differences in cognitive test scores has not always examined measurement invariance prior to examining the predictive validity of IQ scores on achievement (e.g., Naglieri et al., 2005, 2007; Scheiber & Kaufman, 2015). Furthermore, studies using current cognitive and achievement measures to examine differential CARs due to race or ethnicity are limited (Woods Jr. et al., 2021).

G*c* is often one of the strongest correlates of reading comprehension and math problem solving (although G*c* and several other broad abilities contains conflated *g* variance) with evidence of a developmental shift in increased importance over time (Decker & Roberts, 2015; Hajovsky et al., 2014; Villeneuve et al., 2019). The demands of higher-order reading require increased use of linguistic comprehension skills, background knowledge, and vocabulary, whereas math problem solving involves a shift from learning symbol associations and procedural sets of operations towards more language-based reasoning. G*a* has been linked with both basic reading skills and math calculation skills, which is likely due to the phonological component of language that overlaps with learning the associated sounds of letter names and math symbols (Grigorenko et al., 2020). G*s* shows decreasing influences on basic reading skills as children age (Evans et al., 2002). Both G*s* and G*a* appear to be more important in the beginning stages of reading when children are building schemas for phoneme-grapheme correspondence and decoding skills that then leads to successful orthographic mapping skills in older readers (Ehri, 2014; Kilpatrick, 2016). In addition, G*f* and G*s* have been shown to influence mathematics (Floyd et al., 2003; Taub et al., 2008). G*wm* tends to show a more expansive domain-general influence across different academic skills including reading and mathematics academic skills (Evans et al., 2002; Floyd et al., 2012; Grigorenko et al., 2020; Hajovsky et al., 2014; Villeneuve et al., 2019). G*lr* has also been linked with early reading skills (Evans et al., 2002; Hajovsky et al., 2014), whereas G*v* has shown effects on math problem solving skills (Villeneuve et al., 2019). General intelligence (IQ) has shown large effects on reading and mathematics achievement (Caemmerer et al., 2018; Lewno-Dumdie & Hajovsky, 2020; Villeneuve et al., 2019), with meta-analytic work supporting the use of general intelligence scores to diagnose reading and mathematics difficulties (Zaboski et al., 2018).

Although researchers have reported consistencies with CHC broad ability influences on mathematics and reading achievement across studies (with typically small to medium effect sizes; Keith, 2019), the literature is not immune from limitations. For example, despite comprehensive summaries of CARs research (e.g., McGrew & Wendling, 2010), Benson et al. (2016) called on researchers to examine effect sizes and the incremental prediction of CHC broad abilities on academic achievement after accounting for the variance due to *g*. The issue of effect sizes is, by methodological necessity, affected by the type of statistical model employed (an issue pertinent to measurement bias) and the subjective interpretation of effect sizes, whereas the incremental prediction of CHC broad abilities beyond *g* (an issue pertinent to prediction bias) depends on the theoretical orientation of the researcher. Benson et al. (2016) suggested researchers should examine the effect sizes of CHC broad abilities on achievement in comparison to *g* once *g* has been residualized from the broad abilities, and only if they provide an incremental contribution to achievement above and beyond *g*. Zaboski et al. (2018) conducted a meta-analysis of CARs for reading and mathematics and found *g* accounted for a majority of the variance in achievement, with G*c* accounting for the most variance in all measured achievement skills relative to other broad abilities. Canivez (2013) examined the incremental prediction of broad abilities after accounting for the FSIQ (WAIS-IV, WIAT-II, and WIAT-III) and found all broad abilities together explained small to moderate amounts of variance in achievement. In a study employing both observed factor index scores as well as latent ability constructs in predicting reading and mathematics achievement (WISC-IV and WIAT-II), observed scores provided statistically significant incremental prediction above and beyond the FSIQ but the effect sizes were too small to be of practical utility (Glutting et al., 2006). When latent ability constructs were used, *g* and G*c* were significant influences on reading and mathematics achievement (Glutting et al., 2006).

An additional concern is how the hierarchical nature of intelligence is represented within factor models, which provide structural evidence and guide the interpretation of scores (American Educational Research Association et al., 2014; Dombrowski et al., 2018). In a higher-order model, *g* is positioned as an overall factor representing general cognitive ability that manifests due to the shared variance among the broad abilities (e.g., G*c*, G*wm*). In this model, the effects of *g* on individual subtests are fully mediated by the broad abilities (Benson et al., 2016). Alternatively, the bifactor model posits all subtests load on a general factor as well as broad ability factors that remain after the general factor has been partialled from the subtests. In the bifactor model the general and broad ability factors are assumed to be orthogonal. The bifactor model is a useful technique for variance partitioning, which allows for a comparison of effect sizes once *g* has been partialled out from the broad abilities and the residualized broad abilities can then be used to examine whether they provide incremental prediction of achievement above and beyond *g* (Beaujean, 2015; Benson et al., 2016). The higher-order model can also be applied in this respect by determining how much of the relation between a broad cognitive ability and achievement is due to the unique variance left in that broad cognitive ability after extracting variance associated with the second-order factor, *g* (i.e., the residual variance of the broad ability factor). It should be noted there are other theoretically plausible models that have been employed in the extant literature. For example, one plausible model is a correlated first-order factor model, or a model where *g* is considered an emergent property that is caused by the correlations between broad abilities (Kovacs & Conway, 2016; see Gottfredson, 2016, for a different view). Although there are theoretical differences between different hierarchical models of intelligence, examining CARs using both higher-order models and estimates obtained from the first-order residual factor variances of the higher-order model (Caemmerer et al., 2018) helps to address concerns with estimating the incremental prediction of broad cognitive abilities on achievement after controlling for *g* (Benson et al., 2016; Canivez, 2013).

1.2. Social and demographic factors in intelligence and achievement

Although racial and ethnic group differences in mean IQ test scores are known, the reasons for these differences have been widely debated (see Reynolds & Lowe, 2009; Weiss & Saklofske, 2020). Some researchers have argued that race and ethnicity identifiers function as proxies of less tangible factors associated with specific cultural and environmental opportunities that ultimately influence test score performance (Helms et al., 2005). In a review of racial and financial effects in the WISC-IV, WISC-V, and WAIS-IV standardization samples, Weiss and Saklofske (2020) reported on the effects of SES-related mediators (e.g., parental education and income) on IQ test score differences between racial and ethnic groups. Their findings suggested that SES accounted for a much larger percentage of the variance in IQ test score performance between groups than racial or ethnic group endorsements alone, with more variance accounted for by SES in IQ test score differences between Hispanic and White children than between African American and White children (Weiss & Saklofske, 2020). Furthermore, their results suggested race and ethnicity may act as proxies that partially reflect group differences in social and educational opportunities that influence cognitive development during childhood.

The effects of SES on achievement are also well supported. Meta-analytic work has demonstrated a medium to large positive relation between SES and achievement that is moderated by minoritized status (Sirin, 2005). Prior research has shown that deficiencies in reading and mathematics achievement may be accounted for by indicators of SES, such as poverty status, with the achievement gap widening disproportionately for children from racial and ethnic minorities across a 20-year period (Paschall et al., 2018). Findings have also shown that SES moderates the association between IQ and academic performance, with differences between low and high SES groups being half a grade level, on average, from ages 7 to 16 years (von Stumm, 2017). Despite the substantial correlation between global intelligence and achievement reported in the literature (e.g., mean r estimates $= 0.83$; Kaufman et al., 2012), it is important to recognize that these associations do not exist in isolation. Instead, performance on standardized measures of cognitive ability and academic achievement is reflective of the multifaceted, complex interactions between individuals and their environments.

The biopsychosocial ecological model is a theoretical framework that situates human development within the context of multiple, interactive systems (e.g., genetics, personal characteristics, behaviors, social contexts) that together influence cognitive and psychological functioning (Bronfenbrenner & Morris, 2006; Kranzler et al., 2020; Sameroff, 2010). Within these models, genetic influences on psychological functioning are influenced by contextual factors that include the family environment and broader social contexts. These contextual factors within the environment can be partitioned into two sub-components consisting of shared and non-shared environmental influences. Individual differences in developmental trajectories combined with specific life events (unique or shared) lead to outcomes that are of educational significance (e.g., intellectual flexibility, academic achievement, absenteeism). For example, although research has shown that approximately 50% of individual differences in intelligence may be accounted for by genetics (Plomin, 2018), social inequality can have significant impacts on achievement (Selita & Kovas, 2019). Stated differently, the extent that genetics influence child developmental trends and academic achievement is different based on environments in which the individual is engaged (e.g., family, neighborhood, school; Burchinal et al., 2011; Dupere et al., 2010). A meta-analysis of gene x SES interactions on intelligence and academic achievement by Tucker-Drob and Bates (2016) also supports findings of environmental influences on intellectual development. Thus, CARs are not only an outcome of genetic predispositions but also are likely to be influenced by the environmental conditions that an individual is exposed to throughout their lifetime.

Despite the substantial correlations observed between measures of cognitive ability, academic achievement, and SES, there have been rigorous attempts to untangle this complex relationship (for a review, see Lubinski, 2004). For example, one study used a longitudinal, quasi-experimental design to examine the relative influences of ability and SES of origin on a variety of social outcomes in 1074 pairs of biologically related siblings within the National Longitudinal Survey of Youth (NLSY; see Gottfredson, 1997; Murray, 1998). After being assessed as young adults and tracking their differential trajectories along IQ while allowing for a quasi-experimental control of SES, findings indicated the strong influence of ability on social outcomes 15 years later. Likewise, in a comparison of the full NLSY sample ($N = 12{,}686$) against a subsample from which participants in the bottom quartile of income were excluded (to eliminate the potential effects of extreme poverty and limitations in educational opportunity), the analysis indicated a high degree of similarity in mean years of education and percent obtaining a bachelor's degree that suggested general cognitive ability may influence social mobility and opportunities (see Lubinski, 2004; Murray, 1998). These findings are noteworthy as they highlight difficulties with accounting for indicators of SES and help clarify potential limitations associated with inferences drawn from study results.

Parent education has been used within test bias research as an indicator of SES (e.g., Konold & Canivez, 2010; Weiss & Saklofske, 2020). Previous research suggests parent education can act as a proxy to environmental factors critical to child development (e.g., Magnuson et al., 2009) and is associated with both genetic and environmental factors that influence cognitive ability (Lange et al., 2010). Parent education levels have been shown to predict intelligence regardless of family income (Lemos et al., 2011), and parent education explains variability in measures of cognitive ability and achievement (Tucker-Drob & Harden, 2012; Weiss & Saklofske, 2020) with significant correlations observed between global IQ and parent education levels ($r = 0.43$; Weiss et al., 2006).

1.3. Addressing measurement and predictive validity biases with latent variable models

Given our focus on predictive validity, it is essential to determine whether differential CARs between racial and ethnic groups is due to systematic differences in the measurement of constructs across groups (i.e., measurement bias) or differences in the predictions of achievement across groups (i.e., prediction bias). Latent variable models are methodologically advantageous in the examination of predictive bias in CARs research for several reasons. First, latent variable models of cognitive ability can be explicitly tested for measurement invariance, ensuring that the constructs are being measured the same for different groups of individuals (Dimitrov, 2010; Little, 1997). Second, latent variable models of cognitive ability rely on capturing and representing the shared variance among the subtests purported to measure the broad cognitive abilities, reducing the extent that error from individual subtests influences the predictive models (Benson et al., 2016; Keith, 2019). Finally, latent variable models facilitate the examination of complex predictive CARs while maintaining the hierarchical factor structure of cognitive ability.

Once strong measurement invariance is supported (i.e., constrained subtest loadings and intercepts across racial and ethnic groups), differences in latent means and predictive CARs between groups can be examined (Keith, 2019). As mentioned previously, bias in the predictive validity of CARs can occur at the intercept and the slope. In the context of CARs, intercept differences emerge when the groups being compared have divergent averages on the outcome variable (i.e., achievement). Latent mean comparisons are critical within predictive bias research, as differential slopes between groups may not be indicative solely of slope bias but instead differences in latent ability (i.e., the latent mean; Wicherts & Millsap, 2009). However, predictive CAR models must consider latent mean differences, as constraining latent means may bias the magnitude and significance of group slope deviations from the reference group.

Borsboom et al. (2008) conducted a simulation study and demonstrated that when there are latent mean differences, and measurement invariance holds, prediction bias is inevitable. As classification decisions occur at the individual level, the implications of predictive bias are arguably more problematic if they arise from measurement bias related to group membership rather than true differences in latent ability. With measurement bias, it would be unfair to use the test when making special education eligibility determinations; however, with predictive bias, errors would need to be addressed. How these errors are dealt with (e.g., using different cut points based on group membership) may be seen as unfair or of questionable legality. Test bias research may be misdirected if it is left undetermined whether there are true differences in measured ability or if these differences reflect measurement bias. Many of these findings are largely absent from the extant literature but may help guide future research on test bias. Moreover, there has been a paucity of research using latent variable models to examine the effects of CHC-based cognitive abilities on academic achievement across racial and ethnic groups (Keith, 1999; Woods Jr., 2021), and these studies did not control for indicators of SES known to influence performance on cognitive and achievement batteries (Weiss & Saklofske, 2020).

1.4. Purpose of the study

The purpose of the present study was two-fold. First, this study aimed to add empirical evidence to the literature concerning the extent that a seven-factor, higher-order model, informed by CHC-theory, invariantly represents cognitive ability for children and adolescents from racially and ethnically diverse backgrounds. Second, this study aimed to clarify the nature by which predictive CARs differ as a function of race or ethnicity. To accomplish these goals, we leveraged multiple group latent variable structural equation modeling to examine the (a) factorial invariance of a seven-factor, higher-order latent variable representation of cognitive ability between racially and ethnically diverse children and adolescents, and (b) intercept and slope bias in predictive CARs between the latent variable representation of cognitive ability and reading and mathematics achievement. To address limitations in prior research, the current study aimed to control for proxies of SES and opportunity by controlling for parent education at the indicator level (i.e., individual subtest level). To address common trends associated with modeling *g* and broad cognitive abilities in predictive CARs, the current study aimed to examine the incremental predictions of broad cognitive abilities on reading and mathematics achievement, after extracting variance associated with *g* from both the broad cognitive abilities and the achievement outcomes. Importantly, we examined whether latent ability differences account for any observed predictive bias in CARs, which has not been considered in prior research using a current and widely used cognitive test.

2. Method

2.1. Data and sample

To answer our research questions, we analyzed previously collected data used in the norming sample of the WJ IV, provided by Riverside Insights (formerly known as Houghton Mifflin Harcourt; McGrew et al., 2014). The WJ IV Cognitive (Schrank, McGrew, & Mather, 2014) and WJ IV Achievement (Schrank, Mather, & McGrew, 2014) data include measures from individuals between Grades 1–12, sampled from various census regions in the USA (i.e., Northeast, Midwest, South, and West), community types (i.e., metropolitan, micropolitan, and rural), and demographic composition of the population per the 2010 census projections using a stratified random sampling approach. Given the breadth of data to be collected from each participant, data were collected using a planned missing data design, wherein participants were randomly assigned to subgroups of assessments, reducing the overall length of data collection protocols and ensuring that data not collected could be reliably recovered due to them being missing completely at random (see also McGrew et al., 2014; Rhemtulla & Little, 2012). Assessments were administered by trained professional examiners and submitted to the company where multistep, multiple imputation procedures were used (see McGrew et al., 2014).

The norming sample consisted of 3569 observations of children and adolescents ranging in age from 5 to 19 years, with an average of 11.8 years of age ($SD = 3.39$). Most of the participants identified as Caucasian (78%; $n = 2792$), with fewer identifying as African American (14%; $n = 504$), Asian (5%; $n = 165$), American Indian (1%; $n = 24$), or Other (2%; $n = 84$). In addition to racial group endorsements, approximately 18% identified as Hispanic ($n = 644$). There was nearly an even split between gender groups with 50.6% of the participants identifying as female ($n = 1809$), with the remaining identifying as male. Each grade level was nearly equally represented with the percent of students in each grade ranging from 6% (Grade 12; $n = 208$) to 10% (Grade 1; $n = 340$) of the total sample. Most of the participants in the sample lived with parents who had an education level greater than high school (56.7%; $n = 2022$), with fewer living with parents who had a high school equivalency (30.3%; $n = 1081$), and less than a high school education (12.7%; $n = 454$). Table 1 disaggregates parent education level by child endorsed race and ethnic groups.

2.2. Measures

The WJ IV norming data contain multiple measures of cognitive ability and academic skills (McGrew et al., 2014). The test batteries were administered individually and the raw scores from these assessments were converted to age-referenced standardized scores, with means of 100 and standard deviations of 15.

2.2.1. Cognitive ability

In the present study, the WJ IV scoring model was utilized and included *g* and seven CHC broad cognitive abilities (McGrew et al., 2014). Specifically, to represent cognitive ability, we utilized a higher-order factor structure to model *g* as a second-order latent variable, represented by seven broad cognitive abilities. These broad cognitive abilities, identified as Comprehension-Knowledge (G*c*), Fluid Reasoning (G*f*), Short-Term Working Memory (G*wm*), Cognitive Processing Speed (G*s*), Auditory Processing (G*a*), Long-Term

Table 1
Distribution of parent education levels overall, by ethnicity, and by race.

	Sample	By Ethnicity		By Race		
Education Level	Overall	Non-Hispanic	Hispanic	Asian	African American	Caucasian
LTHS	442 (12.8%)	323 (11.2%)	119 (21.3%)	12 (7.3%)	94 (18.8%)	336 (12.1%)
HSEQ	1043 (30.3%)	862 (29.9%)	181 (32.4%)	34 (20.6%)	183 (36.6%)	826 (29.7%)
GTHS	1960 (56.9%)	1701 (58.9%)	259 (46.3%)	119 (72.1%)	223 (44.6%)	1618 (58.2%)

Note. LTHS = less than high school; HSEQ = high school equivalent; GTHS = greater than high school.

Retrieval (G*lr*), and Visual Processing (G*v*), were first-order latent variables represented by the subtests identified in the WJ IV Technical Manual in accordance with CHC factors (McGrew et al., 2014). We modeled the standardized scores of the Oral Vocabulary and General Information subtests as indicators of G*c*, the Concept Formation and Analysis-Synthesis (in place of Number Series) subtests as indicators of G*f*, the Verbal Attention and Numbers Reversed subtests as indicators of G*wm*, the Letter-Pattern Matching and Pair Cancellation subtests as indicators of G*s*, the Phonological Processing and Nonword Repetition subtests as indicators of G*a*, the Story Recall and Visual-Auditory Learning subtests as indicators of G*lr*, and the Visualization and Picture Recognition subtests as indicators of G*v* (see Fig. 3 for visual representation). The internal consistencies for the subtests used to create these latent variables ranged from 0.77 to 0.99 for individuals between the ages of 5–19 years, except for Picture Recognition, which ranged from 0.61 to 0.88 (McGrew et al., 2014). More details concerning the development, validity, and reliability of these cognitive ability measures are available in the WJ IV Technical Manual (McGrew et al., 2014).

2.2.2. *Academic skills*

To represent achievement in reading and mathematics, we modeled four first-order latent variables, identified as Basic Reading (BR), Reading Comprehension (RC), Math Problem Solving (MP), and Math Calculation (MC) using relevant subtests detailed in the WJ IV Technical Manual (McGrew et al., 2014). We modeled the standardized scores of the Letter-Word Identification and Word Attack subtests as indicators of BR, the Passage Comprehension and Reading Recall subtests as indicators of RC, the Applied Problems and Number Matrices subtests as indicators of MP, and the Calculation and Math Fluency subtests as indicators of MC. The internal consistencies for the subtests used to create these latent variables ranged from 0.81 to 0.98 for individuals between the ages of 5–19 years (McGrew et al., 2014). More details concerning the development, validity, and reliability of these achievement measures are available in the WJ IV Technical Manual (McGrew et al., 2014).

2.3. *Protocol for data analysis*

We utilized *R* (R Core Team, 2021, *ver.* 4.1.2) to prepare these data and conduct preliminary analyses and assumption tests, including checks for linearity and normality. Given our focus on predictive bias in CARs as a function of race and ethnic group membership, we retained only the children and adolescents who identified with a racial or ethnic group large enough for analytic comparison (i.e., Caucasian, African American, Asian, Hispanic/non-Hispanic). This reduced our sample size from 3569 to 3461 ($\Delta N = 108$). Eleven did not have data for parent education levels and 17 more were missing scores on at least one of the subtests to be used. We removed these 28 individuals from data prior to analysis, resulting in a pre-analytic sample of 3433.

All latent variable models, including the measurement models and structural equation models (SEM), were specified and estimated with the *lavaan* package (Rosseel, 2012, *ver.* 0.6–9) in *R*. Given the well-documented influence of parent education, as a proxy of SES, on variability in measures of intelligence and achievement (e.g., Tucker-Drob & Harden, 2012; Weiss & Saklofske, 2020) and the statistical issue of confounding effects due to race and education, we controlled for variability in subtest performance that was attributable to differences in the levels of parent reported education. With prior support for collapsing parent education into a college/non-college binary variable (e.g., Hajovsky, Oyen, et al., 2020) and nearly balanced split between those with greater than high school education ($N = 1953$, 56.89%) and those with equal to or less than a high school education ($N = 1480$, 43.11%) in the pre-analytic sample, we dichotomized parent education in the current study. Dichotomizing the parent education variable in the current study also created more robust groups, removing the relatively underpowered sample sizes for the *less than high school* classification (i.e., $N = 441$). Preliminary analyses examining differences in the ability to predict variance in the subtests between the three-group and two-group representations of parent education yielded little difference, with the three-group representation of parent education being able to explain 0%–0.9% more of the variance across subtests ($M = 0.45\%$, $SD = 0.23\%$). Ultimately, the empirical precedent for dichotomization and greater statistical power emerging from aggregating the parent education variable drove our decision. Appendix Table 1 summarizes the effect of the college education binary variable on the subtests of cognitive ability and achievement.

2.3.1. *Measurement model and factorial invariance*

Prior to examining the nature of predictive bias in CARs as a function of racial and ethnic group membership, we first established factorial invariance for the first-order latent variable representations of achievement (i.e., BR, RC, MP, and MC) and the higher-order latent variable representation of *g* and its broad cognitive abilities (i.e., G*c*, G*f*, G*wm*, G*s*, G*a*, G*lr*, and G*v*). The latent variables were identified using the subtests described previously. Residual correlations between four subtests were allowed based on support from prior research. For example, residuals from the Numbers Reversed and Letter-Pattern Matching indicators, and Verbal Attention and Nonword Repetition indicators were allowed to correlate based on common specific variance beyond the latent factors (McGrew et al., 2014; Niileksela et al., 2016; Woods Jr. et al., 2021). The scale of the latent factors was set using the effects coding method to model the latent factors and the relations between these latent factors on the same scale as the domains were measured (Keith, 2019; Little, 2013).

Factorial invariance for the first-order latent variables was established in sequential models. First, configural invariance was established in which all parameters were freely estimated for each race or ethnic group. Second, weak invariance was established, constraining the loadings across race and ethnic groups. Finally, strong invariance was established, constraining the item intercepts across race and ethnic groups. Prior to the moderated SEMs, we examined latent mean differences in comparison to the reference group (i.e., Caucasian for racial groups, non-Hispanic for ethnic groups). Factorial invariance was assessed using changes to the χ^2 and CFI model fit indices. More specifically, $\Delta\chi^2$ that is not statistically significant and/or $\Delta\text{CFI} \leq 0.01$ indicate acceptable fit change between levels of constraint (Cheung & Rensvold, 2002). Additionally, the significance of the $\Delta\chi^2$ between constrained and unconstrained models was used to determine if the difference in latent variable means between race and ethnic groups was statistically significant ($p < .05$). Additionally, latent variable mean differences were modeled as defined parameters in the unconstrained SEM model, with respective confidence intervals and *p*-values.

As we represented *g* as a second-order latent variable, further constraints were applied to ensure that *g* was factorially invariant across all race and ethnic groups prior to the predictive bias analyses. With the broad cognitive abilities (i.e., G*c*, G*f*, G*wm*, G*s*, G*a*, G*lr*, and G*v*) demonstrating strong factorial invariance, we further constrained the loadings of the broad cognitive abilities on *g* to represent weak factorial invariance of the second-order factor, *g*. We then constrained the intercepts of the broad cognitive abilities to represent strong factorial invariance of the second-order factor, *g*. We used the same criteria for ensuring factorial invariance of *g* as we did for the latent variables representing broad cognitive abilities. Likewise, we examined standardized differences in *g* between race and ethnic groups prior to constraining *g* across all race and ethnic groups for the predictive bias analyses. Overall, model fit to data was examined with the χ^2 statistic, root mean square error of approximation (RMSEA), comparative fit index (CFI), and standardized root mean square residual (SRMR). Based on prior findings from the extant literature, we determined RMSEA estimates less than 0.10, CFI estimates greater than 0.90, and SRMR estimates less than 0.10 to be indicative of acceptable model fit to data (Cangur & Ercan, 2015; Iacobucci, 2010; Little, 2013).

2.3.2. SEM models for analytic goal 1: g directly predicting achievement

To address the first part of our main analytic goal to examine intercept and slope bias in predictive CARs between the latent variable representation of cognitive ability and reading and mathematics achievement, we conducted an eight-condition experimental simulation study to determine the best protocol for accurately capturing the bias in predictive CARs, without artificially creating bias due to constraints across groups. More specifically, we designed a $2 \times 2 \times 2$ experimental simulation that manipulated (a) the equality of the latent variable mean of the predictor variable across groups (i.e., equal vs unequal), (b) the equality of the latent variable mean of the outcome variable across groups (i.e., equal vs unequal), and (c) the equality of the strength of the relation between the predictor and outcome latent variables across groups (i.e., equal vs unequal). The analysis of data from these eight conditions relied on sequential models that (a) first established factorial invariance, (b) then constrained the latent variable means of the independent variable, (c) then constrained the latent variable means of the dependent variable (i.e., intercept), and (d) finally constrained the regression parameter of the relationship between the predictor and the outcome. In summary, the results indicated that when the latent variable means of the predictor were equal across groups, the protocol did not introduce any bias in the estimation of intercept and slope bias in the regression. However, when there was a significant difference in the latent variable mean of the predictor, constraining it across groups introduced significant bias in the estimation of intercept and slope parameters in the regression. Systematically altering the order of the protocol (e.g., constraining the slope parameter prior to constraining the latent variable means of the predictor and outcome latent variables) under these conditions did not improve the accuracy in capturing bias in the intercept and slope of the regression model.

With differences in the latent variable mean of the predictor, an iterative process was required to accurately estimate the intercept and slope bias in a latent variable regression model. First, iterative constraints on the (a) latent variable mean of the predictor, (b) latent variable mean of the outcome (i.e., intercept), and (c) relation between predictor and outcome (i.e., slope) were specified with $\Delta\chi^2$ tests conducted to act as omnibus tests of differences between groups. With statistically significant $\Delta\chi^2$ results, comparisons of the coefficients estimated for each pair of groups is justified. These analyses can be accomplished in *lavaan* with defined parameters (e.g., "d.lv := lv.grp1 – lv.grp2"). To ensure that the differences in these estimates are not likely to occur by chance, both the overall *p*-value of the estimates for the defined parameters and their confidence intervals from bootstrapped models were examined. Given that prior studies have demonstrated differences in observed IQ scores as a function of group membership (e.g., Edwards & Oakland, 2006; Scheiber, 2016a, 2016b; Scheiber & Kaufman, 2015), with procedural support from our experimental simulation study, we approached the examination of bias in the intercept and slope of predictive CARs assuming there would be differences in the latent variable mean of *g* between race and ethnic groups.

2.3.3. SEM models for analytic goal 2: g predicting achievement via broad cognitive abilities

To accomplish the second part of our main analytic goal, we sequentially modeled the indirect effects of *g* on achievement through the broad cognitive abilities (i.e., G*c*, G*f*, G*wm*, G*s*, G*a*, G*lr*, and G*v*). To consider the arguments in the literature surrounding the extent that broad cognitive abilities incrementally predict achievement beyond *g*, the first analytic model estimated a regression path from *g* to achievement. With *g* as the predictor of achievement, broad cognitive abilities were then iteratively added to the model to examine the extent that each broad cognitive ability contributed to the prediction of achievement above and beyond *g*. Broad cognitive abilities that did not fall within the exclusion criteria were retained and aggregated into a final analytic model. The criteria for broad cognitive abilities to be excluded from the final model included (a) being a negative predictor of achievement in the presence of *g*, (b) causing *g* to become a negative predictor of achievement, and (c) not being a statistically significant predictor of achievement.

Once these final models were obtained, we examined the extent of difference in the regression intercept and slopes as a function of race and ethnic group identification using an adapted protocol from the prior analyses for goal 1. More specifically, we iteratively constrained latent variable means, regression intercepts, and regression slopes across race and ethnic groups, while testing overall change in model fit via $\Delta\chi^2$ tests. We then examined differences in estimates of regression intercepts and regression slopes, examining the overall *p*-values and confidence intervals from bootstrapped models to determine which groups had statistically significantly different estimates. Additionally, we evaluated the predictive CARs via (a) traditional direct effects of *g* and indirect effects of *g* via the broad cognitive abilities on achievement (higher-order model) and (b) by multiplying the standardized residual of the broad cognitive abilities with the standardized regression coefficient from the broad cognitive ability to achievement (i.e., used the residuals of the broad ability factors as predictors of achievement). This latter approach has been used in prior research (e.g., Caemmerer et al., 2018) to approximate the results one might obtain using a bifactor approach, in which variance in the broad cognitive ability that is attributable to *g* is partialled out (Hajovsky, Villeneuve, Reynolds, et al., 2018; Reynolds & Keith, 2017).

3. Results

Results from the assumption tests revealed 306 multivariate outliers spread across the diverse groups comprising our pre-analytic sample of 3433 children and adolescents. More specifically, gender proportions were similar with the multivariate outliers being comprised of 51.63% male. Racial group proportions were similar with the multivariate outliers being comprised of 72.22% Caucasian, 21.24% African American, and 6.54% Asian. Ethnic group proportions were similar with the multivariate outliers being comprised of 16.01% Hispanic. Finally, parent education group proportions were similar with the multivariate outliers being comprised of 54.25% having an education greater than high school, 29.74% having an education equivalent to high school, and 16.01% percent having an education less than high school. Given the proportional similarity between the pre-analytic sample and the multivariate outliers, we determined there to be little concern for systematic bias and removed them prior to further analyses ($N = 3127$).

Prior to analyzing data, we examined the cross-group distribution of race and ethnicity in the final sample of 3127 children and adolescents. Although nearly 16% of our analytic sample identified as Hispanic, this percentage was not equitable across the racial groups. Instead, of those who identified as Hispanic, 96.06% also identified as Caucasian ($N = 488$), 1.97% also identified as African American ($N = 10$), and 1.97% also identified as Asian ($N = 10$). Given that the Hispanic representation was almost solely nested within those identifying as Caucasian, we subsetted the analytic sample for ethnicity-oriented analyses to include only those individuals in the sample that identified as Caucasian, so that confounding issues with race would not interfere in our interpretation of ethnic bias in the predictive CAR models. This subsetting reduced the sample for ethnicity-oriented analyses to a final sample of 2548. We did not adjust the final sample for race-oriented analyses (i.e., 3127).

3.1. Factorial invariance of g and achievement

Results demonstrated strong factorial invariance for the higher-order representation of *g* across racial groups. More specifically, constraining the loadings of the subtests for the first-order broad cognitive abilities met the $\Delta\chi^2$ and ΔCFI criteria for weak factorial invariance ($\Delta\chi^2 = 16.93$, $p = .260$, ΔCFI = 0.000). Constraining the intercepts of the subtests for the first-order broad cognitive abilities met the $\Delta\chi^2$ and ΔCFI criteria for strong factorial invariance ($\Delta\chi^2 = 7.87$, $p = .896$, ΔCFI = 0.000). Additionally, constraining the loadings of the broad cognitive abilities for second-order *g* met the ΔCFI criterion for weak factorial invariance (ΔCFI = 0.000). Finally, constraining the intercepts of the broad cognitive abilities for the second-order *g* met the ΔCFI criterion for strong factorial invariance (ΔCFI = 0.001). Standardized loadings of the subtests on the broad cognitive abilities ranged from 0.449 to 0.940. The final model, demonstrating strong factorial invariance of *g* across racial groups, also demonstrated acceptable model fit with an RMSEA = 0.073, 90% CI [0.070, 0.076], CFI = 0.907, and SRMR = 0.045. Table 2 summarizes the model fit indices for each model. Additionally,

Table 2
Results of factorial invariance tests of cognitive ability and academic achievement.

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	p ($\Delta\chi^2$)	CFI	ΔCFI	RMSEA	90% CI	SRMR
Factorial Invariance of Cognitive Ability Across Racial Groups										
Configural	1608.58	204				0.908		0.081	[0.078, 0.085]	0.042
BA – Weak	1625.50	218	16.93	14	0.260	0.908	0.000	0.079	[0.075, 0.082]	0.043
BA – Strong	1633.37	232	7.87	14	0.896	0.908	0.000	0.076	[0.073, 0.080]	0.043
g – Weak	1658.58	244	25.21	12	0.014	0.908	0.000	0.075	[0.071, 0.078]	0.045
g – Strong	1681.48	256	22.90	12	0.029	0.907	0.001	0.073	[0.070, 0.076]	0.045
Factorial Invariance of Cognitive Ability Across Ethnic Groups										
Configural	1270.40	136				0.908		0.081	[0.077, 0.085]	0.042
BA – Weak	1278.08	143	7.68	7	0.362	0.908	0.000	0.079	[0.075, 0.083]	0.043
BA – Strong	1286.45	150	8.37	7	0.301	0.908	0.000	0.077	[0.073, 0.081]	0.043
g – Weak	1290.46	156	4.01	6	0.675	0.908	0.000	0.076	[0.072, 0.079]	0.044
g – Strong	1314.38	162	23.93	6	< 0.001	0.906	0.002	0.075	[0.071, 0.078]	0.045
Factorial Invariance of Academic Achievement Across Racial Groups										
Configural	262.34	42				0.985		0.071	[0.063, 0.079]	0.018
Weak	265.07	50	2.73	8	0.950	0.985	0.000	0.064	[0.057, 0.072]	0.019
Strong	278.52	58	13.45	8	0.097	0.985	0.000	0.060	[0.053, 0.068]	0.020
Factorial Invariance of Academic Achievement Across Ethnic Groups										
Configural	234.03	28				0.982		0.076	[0.067, 0.085]	0.019
Weak	237.51	32	3.48	4	0.481	0.982	0.000	0.071	[0.063, 0.080]	0.021
Strong	245.47	36	7.96	4	0.093	0.982	0.000	0.068	[0.060, 0.076]	0.022

Note. BA = broad cognitive abilities, *g* = higher order representation of *g*.

Appendix Table 2 summarizes the first- and second-order loadings for the measurement model of cognitive ability.

Results demonstrated strong factorial invariance for the higher-order representation of *g* across ethnic groups. More specifically, constraining the loadings of the subtests for the first-order broad cognitive abilities met the $\Delta\chi^2$ and ΔCFI criteria for weak factorial invariance ($\Delta\chi^2 = 7.68$, $p = .362$, ΔCFI = 0.000). Constraining the intercepts of the subtests for the first-order broad cognitive abilities met the $\Delta\chi^2$ and ΔCFI criteria for strong factorial invariance ($\Delta\chi^2 = 8.37$, $p = .301$, ΔCFI = 0.000). Additionally, constraining the loadings of the broad cognitive abilities for second-order *g* met the ΔCFI criterion for weak factorial invariance ($\Delta\chi^2 = 4.01$, $p = .675$, ΔCFI = 0.000). Finally, constraining the intercepts of the broad cognitive abilities for the second-order *g* met the ΔCFI criterion for strong factorial invariance (ΔCFI = 0.002). Standardized loadings of the subtests on the broad cognitive abilities ranged from 0.612 to 0.964. The final model, demonstrating strong factorial invariance of *g* across racial groups, also demonstrated acceptable model fit with an RMSEA = 0.075, 90% CI [0.071, 0.078], CFI = 0.906, and SRMR = 0.045. Table 2 summarizes the model fit indices for each model. Across racial and ethnic groups, parent education (represented as having a college education) was a consistently positive predictor of performance across all subtests modeled as manifest indicators of the broad cognitive abilities with standardized regression coefficients ranging from 0.102 to 0.221 ($p < .05$ or less).

Results demonstrated strong factorial invariance for the first-order representations of achievement across racial groups. In testing factorial invariance of achievement, all constructs were modeled simultaneously (i.e., BR, RC, MP, and MC) for each racial and ethnic group. Constraining the loadings of the subtests for the each first-order latent variables of achievement met the $\Delta\chi^2$ and ΔCFI criteria for weak factorial invariance ($\Delta\chi^2 = 2.73$, $p = .950$, ΔCFI = 0.000). Constraining the intercepts of the subtests for the first-order latent variables of achievement met the $\Delta\chi^2$ and ΔCFI criteria for strong factorial invariance ($\Delta\chi^2 = 13.45$, $p = .097$, ΔCFI = 0.000). Standardized loadings of the subtests on the measures of achievement ranged from 0.542 to 0.934. The final model, demonstrating strong factorial invariance of the four latent variables of achievement, also demonstrated acceptable model fit with an RMSEA = 0.060, 90% CI [0.053, 0.068], CFI = 0.985, and SRMR = 0.020. Table 2 summarizes the model fit indices for each model. Additionally, Appendix Table 3 summarizes the first-order loadings for the measurement models of each achievement domain.

Results demonstrated strong factorial invariance for the first-order representations of achievement across ethnic groups. In testing factorial invariance of achievement, all constructs were modeled simultaneously (i.e., BR, RC, MP, and MC) for each racial and ethnic group. Constraining the loadings of the subtests for the each first-order latent variables of achievement met the $\Delta\chi^2$ and ΔCFI criteria for weak factorial invariance ($\Delta\chi^2 = 3.48$, $p = .481$, ΔCFI = 0.000). Constraining the intercepts of the subtests for the first-order latent variables of achievement met the $\Delta\chi^2$ and ΔCFI criteria for strong factorial invariance ($\Delta\chi^2 = 7.96$, $p = .093$, ΔCFI = 0.000). Standardized loadings of the subtests on the measures of achievement ranged from 0.622 to 0.926. The final model, demonstrating strong factorial invariance of the four latent variables of achievement, also demonstrated acceptable model fit with an RMSEA = 0.068, 90% CI [0.060, 0.076], CFI = 0.982, and SRMR = 0.022. Table 2 summarizes the model fit indices for each model. Across racial and ethnic groups, parent education was a consistently positive predictor of performance across all subtests modeled as manifest indicators of achievement with standardized regression coefficients ranging from 0.139 to 0.228 ($p < .05$ or less).

3.2. *Differences in g and achievement as a function of race and ethnicity*

When considering differences in latent means between racial groups, results suggested significant differences in *g* and the four measures of achievement. More specifically, constraining the latent mean of *g* across racial groups resulted in statistically significant changes in model fit ($\Delta\chi^2_g = 44.685$, $p < .001$), suggesting significant differences in *g* between racial groups. Likewise, independently

Table 3
Means, standard deviations, and pairwise comparisons of g and achievement latent variables.

	g		BR		RC		MP		MC	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
By Race										
Caucasian	98.19	7.92	97.56	12.47	98.09	11.21	98.05	10.64	98.15	11.89
African American	94.03	7.83	92.41	12.26	91.68	11.79	91.87	10.78	91.87	12.49
Asian	98.60	8.14	100.21	13.75	98.10	11.07	98.58	11.36	102.22	11.74
By Ethnicity										
Non-Hispanic	98.78	7.72	98.64	12.13	99.49	11.11	98.80	10.37	98.81	11.73
Hispanic	96.48	8.36	94.53	13.45	94.03	10.98	95.58	11.38	95.74	12.35
Pairwise Comparisons (Score Differences and Probabilities)										
	Est	*p*-val	Est	*p*-val	Est	*p*-val	Est	*p*-val	Est	*p*-val
Ca – AA	4.16	< 0.001	5.15	< 0.001	6.41	< 0.001	6.18	< 0.001	6.29	< 0.001
Ca – As	–0.41	0.767	–2.66	0.232	0.00	0.999	–0.53	0.789	–4.06	0.046
As – AA	4.57	0.002	7.81	0.001	6.42	0.003	6.71	0.001	10.35	< 0.001
NonHisp – Hisp	2.30	< 0.001	4.11	< 0.001	5.45	< 0.001	3.22	< 0.001	3.07	0.001

Note. *g* = second order latent variable of cognitive ability; BR = basic reading; RC = reading comprehension; MP = math problem solving; MC = math calculation; Ca = Caucasian; AA = African American; As = Asian; NonHisp = Non-Hispanic; Hisp = Hispanic; Est = Unstandardized difference estimate from unconstrained latent variable model; *p*-val = probability of unstandardized estimate occurring by chance.

constraining the latent means of achievement (i.e., BR, RC, MP, and MC) across racial groups resulted in statistically significant changes in model fit with $\Delta\chi^2_{BR} = 32.819$ ($p < .001$), $\Delta\chi^2_{RC} = 43.804$ ($p < .001$), $\Delta\chi^2_{MP} = 48.845$ ($p < .001$), and $\Delta\chi^2_{MC} = 50.452$ ($p < .001$). Constraining means across pairs of racial groups (i.e., Caucasian-African American, Caucasian-Asian, African American-Asian) resulted in mixed findings, with some pairwise constraints resulting in significant changes to model fit and other pairwise constraints not resulting in significant changes to model fit. Table 3 summarizes the latent means and standard deviations of *g* and the four measures of achievement (i.e., BR, RC, MP, and MC) across racial groups, along with results of racial pairwise comparisons. Overall, differences in *g* and the four measures of achievement between Caucasian and African American children and adolescents ranged from 4.16 to 6.41 standard score points (all $p < .001$). Likewise, differences between Asian and African American children and adolescents ranged from 4.57 to 10.35 points (all $p < .010$). Differences between Caucasian and Asian children and adolescents ranged from -4.06 to 0.00 points, with only the difference on MC (i.e., -4.06) being statistically significant ($p = .046$).

When considering differences in latent means between ethnic groups, results suggested significant differences in *g* and the four measures of achievement. More specifically, constraining the latent mean of *g* across ethnic groups resulted in statistically significant changes in model fit ($\Delta\chi^2_g = 13.004$, $p < .001$), suggesting significant differences in *g* between ethnic groups. Likewise, independently constraining the latent means of achievement (i.e., BR, RC, MP, and MC) across ethnic groups resulted in statistically significant changes in model fit with $\Delta\chi^2_{BR} = 17.521$ ($p < .001$), $\Delta\chi^2_{RC} = 34.788$ ($p < .001$), $\Delta\chi^2_{MP} = 12.958$ ($p < .001$), and $\Delta\chi^2_{MC} = 10.746$ ($p < .001$). Table 3 summarizes the latent means and standard deviations of *g* and the four measures of achievement (i.e., BR, RC, MP, and MC) across ethnic groups, along with results of ethnic group pairwise comparisons. Overall, differences in *g* and the four measures of achievement between Non-Hispanic and Hispanic children and adolescents ranged from 2.30 to 5.45 points (all $p < .001$).

3.3. Analytic goal part 1: direct effects of g on achievement

Given the results from the prior analyses suggesting significant differences in the latent variable means of *g* and achievement between race and ethnic groups, we adhered to our established protocol for analyzing intercept and slope bias in predictive CAR models to reduce the bias that would emerge via constraining the latent variable means of *g*. More specifically, we iteratively constrained (a) the latent variable mean of *g*, (b) the latent variable mean of the achievement measure (i.e., intercept), and (c) the relation between *g* and achievement (i.e., slope). Constraints that resulted in significantly worse fitting models based on a $\Delta\chi^2$ test were followed up with pairwise analyses coded within the SEMs. However, as the latent variable mean of *g* was examined in the prior section, we only report on the constraints and pairwise tests associated with the intercept and slope. Table 4 summarizes the intercept and slope results from the CAR models, with coefficients estimated for disaggregated subgroups and combined. Appendix Table 4 provides additional details about the path coefficients. The results from these models are discussed in the following paragraphs.

Results from the SEM models examining the influence of *g* on BR demonstrated noticeable differences in the intercept coefficients between racial groups, with nearly a 12-point difference between Asian children and adolescents and their African American peers. Although constraining the unstandardized intercepts of all racial groups to an averaged -11.98 resulted in significantly worse model fit ($\Delta\chi^2 = 515.03$, $p < .001$), none of the pairwise comparisons were statistically significant. The differences in the $\Delta\chi^2$ and pairwise tests were likely due to larger standard errors around the individual racial groups' intercepts. Constraining the unstandardized slopes to an averaged 1.12 did not result in significantly worse model fit indices. Aside from the CFI, estimates of model fit for the SEM examining the CAR between *g* and BR between racial groups were acceptable with RMSEA = 0.081, 90% CI [0.078, 0.084], CFI =

Table 4
Intercept and slope coefficients of g predicting achievement by race and ethnicity.

	BR		RC		MP		MC	
Group	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
By Race								
Caucasian	−12.07	1.12	−9.04	1.09	−24.39	1.25	−10.91	1.11
African American	−17.88	1.17	−14.93	1.14	−25.93	1.25	−10.59	1.09
Asian	−6.00	1.08	−14.44	1.14	−36.65	1.37	−9.78	1.14
Aggregate	−11.98	1.12	−12.80	1.12	−28.99	1.29	−10.43	1.11
By Ethnicity								
Non-Hispanic	−12.47	1.12	−9.12	1.10	−24.12	1.25	−9.36	1.10
Hispanic	−14.36	1.13	−8.51	1.06	−24.85	1.25	−16.05	1.16
Aggregate	−13.42	1.13	−8.82	1.08	−24.49	1.25	−12.70	1.13
Pairwise Comparisons (Estimates of Differences in Intercepts and Slopes, and Probabilities)								
Ca – AA	5.81	−0.05	5.90	−0.05	1.54	−0.01	−0.32	0.02
Ca – As	−6.08	0.04	5.40	−0.05	12.27	−0.13	−1.13	−0.03
As – AA	11.89	−0.09	0.50	0.01	−10.73	0.12	0.81	0.05
NonHisp – Hisp	1.89	−0.01	−0.62	0.04	0.73	0.00	6.70	−0.06

Note. No pairwise comparisons for intercepts or slopes were statistically significant ($p < .05$); BR = basic reading; RC = reading comprehension; MP = math problem solving; MC = math calculations; Ca = Caucasian; AA = African American; As = Asian; NonHisp = Non-Hispanic; Hisp = Hispanic; Pairwise comparisons of intercepts and slopes based on unconstrained models.

0.883, and SRMR = 0.052. Results from the same model applied to ethnic groups (i.e., Hispanic, Non-Hispanic) demonstrated similar trends. With a nearly 2-point difference between ethnic groups, constraining the unstandardized intercepts to an averaged −13.42 did not result in significantly worse model fit indices. Likewise, constraining the unstandardized slopes to an averaged 1.13 did not result in significantly worse model fit indices. Aside from the CFI, estimates of model fit for the SEM examining the CAR between *g* and BR between ethnic groups were acceptable with RMSEA = 0.082, 90% CI [0.079, 0.085], CFI = 0.884, and SRMR = 0.051.

Results from the SEM models examining the influence of *g* on RC only demonstrated a noticeable difference for Caucasian children and adolescents, with their intercept being nearly 6 points higher than their African American and Asian peers. Constraining the unstandardized intercepts to an averaged −12.80 did not result in significantly worse model fit indices. Constraining the unstandardized slopes to an averaged 1.12 did not result in significantly worse model fit indices. Aside from the CFI, estimates of model fit for the SEM examining the CAR between *g* and RC between racial groups were acceptable with RMSEA = 0.089, 90% CI [0.086, 0.092], CFI = 0.856, and SRMR = 0.051. Results from the same model applied to ethnic groups demonstrated similar trends. With nearly a half-point difference between ethnic groups, constraining the unstandardized intercepts to an averaged −8.82 did not result in significantly worse model fit indices. Likewise, constraining the unstandardized slopes to an averaged 1.08 did not result in significantly worse model fit indices. Aside from the CFI, estimates of model fit for the SEM examining the CAR between *g* and RC between ethnic groups were acceptable with RMSEA = 0.091, 90% CI [0.088, 0.094], CFI = 0.855, and SRMR = 0.051.

Results from the SEM models examining the influence of *g* on MP only demonstrated a noticeable difference for Asian children and adolescents, with their intercept being nearly 12 points above their Caucasian peers and 11 points above their African American peers. Constraining the unstandardized intercepts to an averaged −28.99 did not result in significantly worse model fit indices. Although constraining the unstandardized slopes to an averaged 1.29 resulted in significantly worse model fit ($\Delta\chi^2 = 1684.6, p < .001$), none of the pairwise comparisons were statistically significant. Aside from the CFI, estimates of model fit for the SEM examining the CAR between *g* and MP between racial groups were acceptable with RMSEA = 0.079, 90% CI [0.077, 0.082], CFI = 0.887, and SRMR = 0.047. Results from the same model applied to ethnic groups demonstrated similar trends. With nearly a 1-point difference between ethnic groups, constraining the unstandardized intercepts to an averaged −24.49 did not result in significantly worse model fit indices. Likewise, constraining the unstandardized slopes to an averaged 1.25 did not result in significantly worse model fit indices. Aside from the CFI, estimates of model fit for the SEM examining the CAR between *g* and MP between ethnic groups were acceptable with RMSEA = 0.081, 90% CI [0.078, 0.084], CFI = 0.885, and SRMR = 0.047.

Results from the SEM models examining the influence of *g* on MC did not demonstrate noticeable differences in the intercepts between the racial groups. Constraining the unstandardized intercepts to an averaged −10.43 did not result in significantly worse model fit indices. Although constraining the unstandardized slopes to 1.11 resulted in significantly worse model fit ($\Delta\chi^2 = 7835.8, p <$ .001), none of the pairwise comparisons were statistically significant. Aside from the CFI, estimates of model fit for the SEM examining the CAR between *g* and MC between racial groups were acceptable with RMSEA = 0.089, 90% CI [0.086, 0.092], CFI = 0.860, and SRMR = 0.056. Results from the same model applied to ethnic groups demonstrated similar trends. Although there was nearly a 7-point difference between ethnic groups, constraining the intercepts to an averaged −12.70 did not result in significantly worse model fit indices. Likewise, constraining the unstandardized slopes to an averaged 1.13 did not result in significantly worse model fit

Table 5
Predicted achievement scores across a continuum of g for different racial groups.

	Continuum of LV Scores on *g*				
	70	85	100	115	130
For Basic Reading					
Caucasian	66.1	82.9	99.6	116.4	133.1
African American	63.9	81.5	99.0	116.6	134.1
Asian	69.6	85.8	102.0	118.2	134.4
For Reading Comprehension					
Caucasian	67.3	83.7	100.1	116.4	132.8
African American	64.5	81.5	98.6	115.6	132.6
Asian	65.4	82.5	99.6	116.7	133.8
For Math Problem Solving					
Caucasian	62.9	81.6	100.3	119.0	137.7
African American	61.8	80.6	99.4	118.2	137.0
Asian	59.5	80.1	100.6	121.2	141.8
For Math Calculation					
Caucasian	66.9	83.5	100.2	116.9	133.5
African American	65.7	82.1	98.4	114.8	131.1
Asian	69.9	87.0	104.0	121.1	138.2

Note. Predicted scores calculated from intercepts and slopes from Table 4. For predicted outcome $\hat{y} = b_{intercept} + g(b_{slope})$. g = higher order representation of *g*.

indices. Aside from the CFI, estimates of model fit for the SEM examining the CAR between *g* and MC between ethnic groups were acceptable with RMSEA = 0.090, 90% CI [0.087, 0.093], CFI = 0.086, and SRMR = 0.056.

To improve our explanation of the differences in relations between *g* and four measures of achievement (i.e., BR, RC, MP, and MC) at contextually relevant scores, we applied the unique regression formulas (i.e., intercept and slope) for each racial group to create predicted levels of achievement from 5 levels of *g* (i.e., 70, 85, 100, 115, and 130). Table 5 reports these predicted achievement scores for each racial group. We visualized the relation between *g* and MC for each racial group in Fig. 2A. Although the prior results indicated no statistically significant intercept or slope bias in the predictive CAR, there was a marked difference in MC between Asian children and adolescents and their African American peers. More specifically, a 4.2-point difference existed at a *g* of 70, increasing to 5.6 points at a *g* of 100 and 7.1 points at a *g* of 130.

To improve our explanation of the differences in relations between *g* and four measures of achievement (i.e., BR, RC, MP, and MC) at contextually relevant scores, we applied the unique regression formulas (i.e., intercept, slope) for each ethnic group to create predicted levels of achievement from 5 levels of *g* (i.e., 70, 85, 100, 115, and 130). Table 6 reports these predicted achievement scores for each ethnic group. We visualized the relation between *g* and RC for each ethnic group in Fig. 2B. Although the prior results indicated no statistically significant intercept or slope bias in the predictive CAR, there is an observable difference in RC between Hispanic children and adolescents and their non-Hispanic peers. More specifically, a 1.8-point difference existed at a *g* of 70, increasing to 2.9 points at a *g* of 100 and 3.9 points at a *g* of 130.

3.4. Analytic goal part 2: *g* predicting achievement via broad cognitive abilities

Prior to examining the direct effects of *g* and the indirect effects of *g* via the broad cognitive abilities on achievement, we iteratively modeled the additive predictive strength of broad cognitive abilities on achievement after controlling for variance attributable to *g*. Broad cognitive abilities were retained for the final models if they did not (a) become a negative predictor of achievement in the presence of *g*, (b) cause *g* to become a negative predictor of achievement, or (c) significantly predict variance in achievement. In the models disaggregated by racial group, G*c* was a statistically significant, positive predictor of BR, RC, MP, and MC, even in the presence of *g*. Additionally, G*s* was a statistically significant, positive predictor of MC. These findings were consistent with the models disaggregated by ethnic group. Fig. 3 exemplifies the models analyzed in this section with *g* predicting BR directly and indirectly via G*c*.

Results from the SEM models examining the direct and indirect effects of *g* on BR suggested potential differences in the predictive relation between *g*, G*c*, and BR as a function of racial group identification. Table 7 and Appendix Table 5 summarize the results from these SEM models. Constraining the slope between *g* and BR to an averaged 0.45 across racial groups resulted in significantly worse model fit indices ($\Delta\chi^2 = 924.09$, $p < .001$). Likewise, constraining the slope between G*c* and BR to averaged 0.47 resulted in significantly worse model fit ($\Delta\chi^2 = 8.58$, $p < .05$). The most notable differences in these slopes emerged in the model for Asian children and adolescents. The slope between *g* and BR for this racial group was quite smaller than in the models for the other racial groups (i.e., $b_g =$

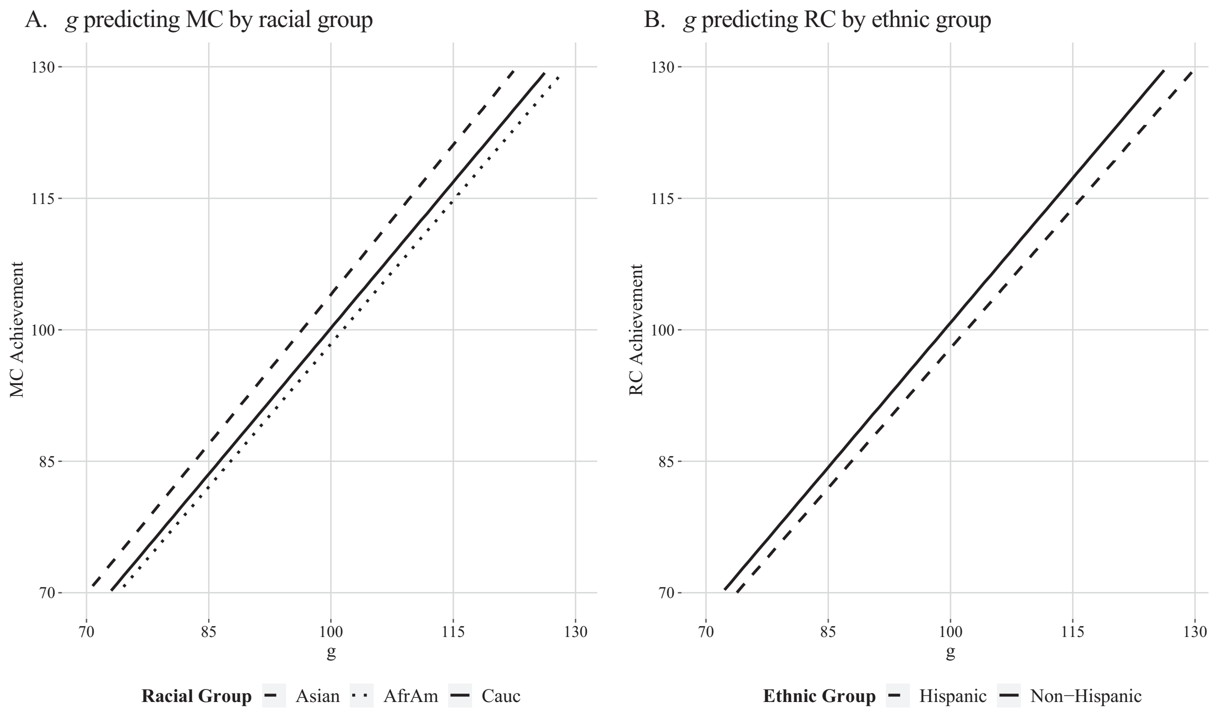

Fig. 2. Regression lines for CARs across racial and ethnic groups.

Table 6
Predicted achievement scores across a continuum of g for different ethnic groups.

	Continuum of LV Scores on *g*				
	70	85	100	115	130
For Basic Reading					
Non-Hispanic	66.2	83.1	99.9	116.8	133.6
Hispanic	64.8	81.8	98.7	115.7	132.7
For Reading Comprehension					
Non-Hispanic	67.8	84.3	100.8	117.3	133.7
Hispanic	66.0	81.9	97.9	113.9	129.8
For Math Problem Solving					
Non-Hispanic	63.0	81.7	100.4	119.1	137.7
Hispanic	62.5	81.2	100.0	118.7	137.4
For Math Calculation					
Non-Hispanic	67.3	83.7	100.1	116.6	133.0
Hispanic	65.1	82.5	99.8	117.2	134.6

Note. Predicted scores calculated from intercepts and slopes from Table 4. For predicted outcome $\hat{y} = b_{intercept} + g(b_{slope})$. g = higher order representation of g.

0.19), with a statistically significant difference from the slope estimated in the model for Caucasian children and adolescents (Δb_g = 0.46, $p < .05$). The slope between G*c* and BR for this racial group was larger than in the models for other racial groups (i.e., $b_{Gc} = 0.63$), with a statistically significant difference from the slope estimated in the model for Caucasian children and adolescents ($\Delta b_{Gc} = -0.30$, $p < .05$). Considering the indirect effect of *g* on BR through G*c*, the effect was strongest for Asian children and adolescents (i.e., b_{ind} = 0.70) and weakest for Caucasian children and adolescents (i.e., $b_{\text{ind}} = 0.36$). There was a statistically significant difference in the indirect effects between Asian children and adolescents and their Caucasian peers ($\Delta b_{\text{ind}} = -0.34$, $p < .05$). Given that these are unstandardized estimates, we can interpret the indirect estimate for Asian children and adolescents as one standard score increase in *g* would result in 0.70 standard score increase in BR via G*c*. These results are corroborated by the estimates of unique influence of G*c* on BR, after residualizing out the variance attributable to *g* (i.e., Residual * β). After controlling for the variability attributable to *g*, G*c* has a greater standardized influence on BR for Asian children and adolescents (β = 0.33, large effect; Keith, 2019) than their African American (β = 0.23, moderate effect) and Caucasian (β = 0.16, moderate effect) peers. Although there was some variability in the models between ethnic groups, there were not any noticeable or significant differences in the estimates.

Results from the SEM models examining the direct and indirect effects of *g* on RC suggested little difference in the predictive relation between *g*, G*c*, and RC as a function of racial or ethnic group identification. Table 8 summarizes the results from these SEM models. Although there was some minor variability in the estimates for each group, iteratively constraining the slopes did not result in significantly worse model fit for any comparative models (i.e., race and ethnicity). Overall, G*c* was a statistically significant mediator in the predictive CAR between *g* and RC. By racial group, indirect effects ranged from 0.44 to 0.51. By ethnic group, indirect effects ranged from 0.36 to 0.43. Concerning the unique influence of G*c* on RC, after residualizing out the variance attributable to *g*, standardized estimates ranged from moderate to large (β = 0.20 to 0.26) across racial groups and moderate (β = 0.19 to 0.22) across ethnic groups.

Results from the SEM models examining the direct and indirect effects of *g* on MP suggested little difference in the predictive relation between *g*, G*c*, and MP as a function of racial or ethnic group identification. Table 9 and Appendix Table 6 summarize the results from these SEM models. Although there was some minor variability in the estimates for each group, iteratively constraining the slopes did not result in significantly worse model fit for any comparative models (i.e., race and ethnicity). Overall, G*c* was a statistically significant mediator in the predictive CAR between *g* and MP. By racial group, indirect effects ranged from 0.18 to 0.30. By ethnic group, indirect effects ranged from 0.15 to 0.22. Concerning the unique influence of G*c* on MP, after residualizing out the variance attributable to *g*, standardized estimates ranged from small to moderate (β = 0.09–0.18) across racial groups and small to moderate (β = 0.08–0.14) across ethnic groups.

Results from the SEM models examining the direct and indirect effects of *g* on MC suggested little difference in the predictive relation between *g*, G*c*, G*s*, and MC as a function of racial or ethnic group identification. In determining the final model, there was some discrepancy in the broad cognitive abilities to be included. Specifically, in the model disaggregated by race, using both G*c* and G*s* as additive predictors of MC did not result in estimates that should have been excluded (i.e., broad cognitive ability becoming negative predictor, *g* becoming negative predictor, and broad cognitive ability not statistically significant). However, in the model disaggregated by ethnicity, adding both G*c* and G*s* as additive predictors resulted in *g* becoming a negative predicator of MC. As such, we retained the strongest predictor of the two (i.e., G*s*) for just the model examining ethnic group differences. Table 10 summarizes the results from these SEM models. Although there was some minor variability in the estimates for each group, iteratively constraining the slopes did not result in significantly worse model fit for any comparative models (i.e., race, ethnicity). Overall, G*c* and G*s* were

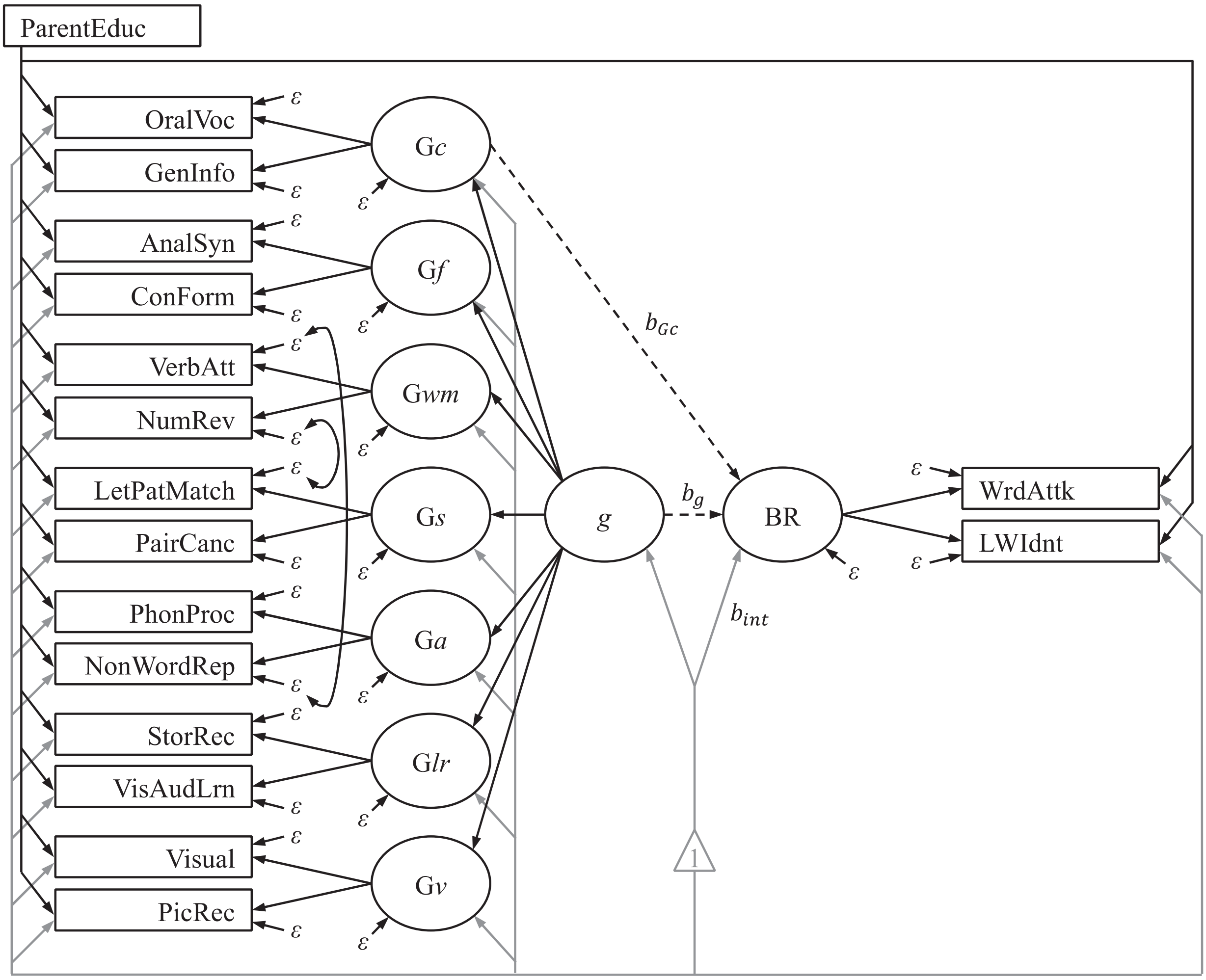

Fig. 3. SEM for g predicting basic reading achievement directly and via *Gc*.
Note. G*c* = comprehension-knowledge; G*f* = fluid reasoning; G*wm* = short-term working memory; G*s* = cognitive processing speed; G*a* = auditory processing; G*lr* = long-term storage; G*v* = visual-spatial processing; *g* = general intelligence; BR = basic reading. Tables 1-1 and 1-3 in McGrew et al. (2014) provide full subtest names for each achievement and cognitive ability domain.

statistically significant mediators in the predictive CAR between *g* and MC. By racial group, results concerning the indirect effects and unique influence both demonstrated that G*s* was a strong predictor of MC than G*c*. Across racial groups, indirect effects ranged from 0.41 to 0.46 for G*s* and from 0.19 to 0.32 for G*c*. Similarly, after residualizing out the variance attributable to *g*, the unique influence of G*s* ranged from 0.31 to 0.37 (large effect) and G*c* ranged from 0.09 to 0.17 (small to moderate effect). By ethnic group, the indirect effects of G*s* ranged from 0.54 to 0.57. Concerning the unique influence of G*s* on MC, after residualizing out variance attributable to *g*, standardized estimates ranged from 0.28 to 0.32 (large effect).

4. Discussion

Test bias research continues to be an important avenue of research to inform equitable assessment practices within school psychology. Measures of cognitive ability and academic achievement are frequently employed within psychological assessment practices and are often used in disability identification models (Benson et al., 2019; Jewsbury et al., 2017). In the current study we examined the construct equivalence and predictive validity of the Woodcock-Johnson Tests of Cognitive Ability, Fourth Edition on reading and mathematics achievement across children and adolescents from diverse racial and ethnic backgrounds using latent variable structural equation modeling. To address limitations in prior research and improve the accuracy of our results, we controlled for parent education, as it is considered a proxy to SES and has theoretical and empirical associations with cognitive ability and standardized achievement scores (e.g., Weiss & Saklofske, 2020). As an important extension of prior research, we examined whether latent mean differences in *g* may account for any observed differences in prediction bias, which have not been considered in prior test bias research. Simulation research suggests that when latent mean differences exist between groups, and measurement invariance is supported, then

Table 7
Results of direct and indirect relations between g and BR via broad abilities.

	Direct Effects (*b*)			Indirect Effects (*b*)	Broad Ability Effects (Residual*β)
Group	b_{Int}	b_g	b_{Gc}	G*c*	G*c*
By Race					
Caucasian	1.76	0.65***	0.32***	0.36***	0.16
African American	2.09	0.51***	0.45***	0.51***	0.23
Asian	22.24	0.19***	0.63***	0.70***	0.33
Aggregate	8.70	0.45***	0.47***	0.52***	
By Ethnicity					
Non-Hispanic	0.41	0.70***	0.30***	0.33***	0.14
Hispanic	–0.46	0.66***	0.33***	0.37***	0.19
Aggregate	–0.03	0.68***	0.32***	0.35***	
Pairwise Comparisons (Estimates of Differences in Intercepts and Slopes, and Probabilities)					
Ca – AA	–0.33	0.15	–0.13	–0.15	
Ca – As	–20.47	0.46*	–0.30*	–0.34*	
As – AA	20.15	−0.32	0.17	0.19	
NonHisp – Hisp	0.87	0.04	–0.04	–0.04	

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; Ca = Caucasian; AA = African American; As = Asian; NonHisp = Non-Hispanic; Hisp = Hispanic; *g* = higher order representation of *g*; BR = Basic Reading; G*c* = Comprehension-Knowledge; Direct b_{int} is unstandardized intercept, b_g is unstandardized regression slope with *g*, and b_{Gc} is unstandardized regression slope of G*c*; *g* Indirect are unstandardized indirect effects calculated from bootstrapped model; Residual * β are standardized coefficients.

Table 8
Results of direct and indirect relations between g and RC via broad abilities.

	Direct Effects (*b*)			Indirect Effects (*b*)	Broad Ability Effects (Residual*β)
Group	b_{Int}	b_g	b_{Gc}	G*c*	G*c*
By Race					
Caucasian	10.83***	0.50***	0.39***	0.44***	0.20
African American	7.19	0.46***	0.45***	0.51***	0.22
Asian	9.44	0.47***	0.44***	0.50***	0.26
Aggregate	9.16*	0.48***	0.43***	0.48***	
By Ethnicity					
Non-Hispanic	10.16**	0.52***	0.38***	0.43***	0.19
Hispanic	10.02	0.56***	0.32***	0.36***	0.22
Aggregate	10.09**	0.54***	0.35***	0.40***	
Pairwise Comparisons (Estimates of Differences in Intercepts and Slopes, and Probabilities)					
Ca – AA	3.63	0.05	–0.07	–0.08	
Ca – As	1.38	0.03	–0.05	–0.06	
As – AA	2.25	0.01	–0.01	–0.02	
NonHisp – Hisp	0.14	–0.04	0.06	0.07	

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; Ca = Caucasian; AA = African American; As = Asian; NonHisp = Non-Hispanic; Hisp = Hispanic; *g* = higher order representation of *g*; RC = Reading Comprehension; G*c* = Comprehension-Knowledge; Direct b_{int} is unstandardized intercept, b_g is unstandardized regression slope with *g*, and b_{Gc} is unstandardized regression slope of G*c*; *g* Indirect are unstandardized indirect effects calculated from bootstrapped model; Residual * β are standardized coefficients.

differential predictive CARs are likely to occur (Borsboom et al., 2008). Given the lack of research using latent variable models to examine CHC-based cognitive abilities on achievement across racial and ethnic groups, we examined general and broad cognitive abilities on reading and mathematics achievement. In a set of these models, we also examined the incremental prediction of broad cognitive abilities on reading and mathematics achievement, after controlling for *g* in the broad cognitive abilities and achievement as suggested by Benson et al. (2016). Our findings suggested that there was limited evidence of bias observed in the present study. Furthermore, these findings were consistent across theoretical models using general and broad cognitive abilities as predictors of reading and mathematics achievement.

4.1. Summary of findings

Our discussion of findings predominantly focus on the analyses within the latent variable SEM models that examined general and broad cognitive abilities on reading and mathematics achievement, after controlling for parent education. It was first important to

Table 9
Results of direct and indirect relations between g and MP via broad abilities.

	Direct Effects (*b*)			Indirect Effects (*b*)	Broad Ability Effects (Residual*β)
Group	b_{Int}	b_g	b_{Gc}	Gc	Gc
By Race					
Caucasian	–18.66***	1.03***	0.16***	0.18***	0.09
African American	–19.05**	1.01**	0.18**	0.20**	0.10
Asian	–26.32**	1.01***	0.27***	0.30***	0.18
Aggregate	–21.34***	1.01***	0.20***	0.22***	
By Ethnicity					
Non-Hispanic	–18.90***	1.05***	0.14***	0.15***	0.08
Hispanic	–18.29**	0.99**	0.20***	0.22***	0.14
Aggregate	–18.59***	1.02***	0.17***	0.19***	
Pairwise Comparisons (Estimates of Differences in Intercepts and Slopes, and Probabilities)					
Ca – AA	0.39	0.02	–0.01	–0.01	
Ca – As	7.66	0.02	–0.11	–0.12	
As – AA	–7.26	0.00	0.09	0.10	
NonHisp – Hisp	–0.60	0.07	–0.06	–0.07	

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; Ca = Caucasian; AA = African American; As = Asian; NonHisp = Non-Hispanic; Hisp = Hispanic; *g* = higher order representation of *g*; MP = Math Problem Solving; G*c* = Comprehension-Knowledge; Direct b_{int} is unstandardized intercept, b_g is unstandardized regression slope with *g*, and b_{Gc} is unstandardized regression slope of G*c*; *g* Indirect are unstandardized indirect effects calculated from bootstrapped model; Residual * β are standardized coefficients.

Table 10
Results of direct and indirect relations between g and MC via broad abilities.

	Direct Effects (*b*)				Indirect Effects (*b*)		Broad Ability Effects (Residual*β)	
Group	b_{Int}	b_g	b_{Gc}	b_{Gs}	Gc	Gs	Gc	Gs
By Race								
Caucasian	–13.81***	0.45***	0.17***	0.52***	0.19***	0.41***	0.09	0.32
African American	–10.69	0.26**	0.23**	0.59***	0.26***	0.46***	0.11	0.37
Asian	–6.68	0.29**	0.28**	0.53**	0.32***	0.41***	0.17	0.31
Aggregate	–10.39**	0.34***	0.23***	0.55***	0.25***	0.61***		
By Ethnicity								
Non-Hispanic	–16.60***	0.66***		0.51***		0.57***		0.32
Hispanic	–27.34***	0.79***		0.48***		0.54***		0.28
Aggregate	–21.97***	0.73***		0.49***		0.56***		
Pairwise Comparisons (Estimates of Differences in Intercepts and Slopes, and Probabilities)								
Ca – AA	–3.12	0.19	–0.07	–0.07	–0.07	–0.06		
Ca – As	–7.13	0.16	–0.12	–0.01	–0.13	–0.01		
As – AA	4.01	0.03	0.05	–0.06	0.06	–0.05		
NonHisp – Hisp	10.74	–0.13		0.03		0.03		

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; Ca = Caucasian; AA = African American; As = Asian; NonHisp = Non-Hispanic; Hisp = Hispanic; *g* = higher order representation of *g*; MC = Math Calculation; G*c* = Comprehension-Knowledge; G*s* = Processing Speed; Direct b_{int} is unstandardized intercept, b_g is unstandardized regression slope with *g*, b_{Gc} is unstandardized regression slope of G*c*, and b_{Gs} is unstandardized regression slope of G*s*,; *g* Indirect are unstandardized indirect effects calculated from bootstrapped model; Residual * β are standardized coefficients.

establish strong factorial invariance before comparing latent means across groups or examining the predictive validity of intelligence on achievement. We found support for strong factorial invariance across racial and ethnic groups, supporting the same comparison of scores across groups. In essence, the calculation of composite scores used in practice are not confounded by measurement differences across groups, which should give confidence that scoring is equally meaningful for individuals from different groups when using the WJ IV. Findings of construct equivalence of modern intelligence tests across racial and ethnic groups is consistent with results from previous research (e.g., Scheiber, 2016a, 2016b; Trundt et al., 2018; Woods Jr. et al., 2021). Given evidence of strong factorial invariance, we then conducted a simulation to determine whether the presence of latent mean differences in *g* would potentially bias predictive validity estimates. Results showed that when latent *g* means were equal across groups, constraining the latent variable mean across groups introduced no additional bias into the estimation of predictive validity of intelligence on achievement (i.e., the intercept or slope in the regression). However, when latent mean differences in *g* were observed, constraining the latent variable mean across groups introduced bias in the estimation of the intercept or slope in the regression when analyzing the predictive validity of intelligence on achievement. As such, when discussing general and broad cognitive ability effects on reading and mathematics achievement from the latent variable SEM models, we do so when latent *g* means were freed across groups to mitigate any potential bias that could be introduced into the models.

Results showed that *g* was the strongest predictor of achievement in predictive models for basic reading, reading comprehension, and math problem solving; however, G*s* was the strongest predictor of achievement in the predictive model for math calculation. It is noteworthy that G*c* was a consistent positive predictor of achievement, in addition to *g*, in all models. The small to moderate G*c* effects on math calculation and math problem solving and moderate to large G*c* effects on basic reading and reading comprehension remained after partialling out *g* from the broad abilities (the broad ability residual factors) and achievement, which is a consistent finding (Benson et al., 2016; Zaboski et al., 2018). In addition, G*s* had large effects on math calculation above and beyond *g* within the latent variable SEM models (Floyd et al., 2003; Niileksela et al., 2016), but no other broad cognitive abilities were significant after partialling out *g*. When examining predictive validity differences across racial and ethnic groups, we found evidence of two instances of slope bias. Specifically, *g* was a stronger predictor of basic reading for Caucasian children and adolescents than it was for Asian children and adolescents (unstandardized effect of 0.65 vs. 0.19). This result suggests that for every standard score point increase in *g*, it results in a corresponding increase in 0.65 standard score points in basic reading for Caucasian children and adolescents, but a 0.19 standard score point increase in basic reading for Asian children and adolescents. Conversely, the broad cognitive ability G*c* was a stronger predictor for Asian children and adolescents than it was for Caucasian children and adolescents (unstandardized effect of 0.63 vs. 0.32). In other words, in the presence of G*c*, overall *g* is less predictive of achievement for Asian children and adolescents. These findings diverge from prior research (e.g., Keith, 1999; Woods Jr. et al., 2021); however, this prior research did not account for parent education levels or examine bias within Asian children and adolescents. We did not find any other evidence of slope bias and we found no evidence of intercept bias in our study. Although findings of slope bias are rare (e.g., Konold & Canivez, 2010), findings of intercept bias are more common (e.g., Lewno-Dumdie & Hajovsky, 2020; Scheiber & Kaufman, 2015; Weiss & Prifitera, 1995). We believe our findings differ from prior literature in two important ways. First, we controlled for parent education prior to any analyses, which reduced systematic variance between groups that was attributable to SES-indicators (Weiss & Saklofske, 2020). Second, we conducted predictive validity bias analyses using hierarchical models of intelligence that iteratively accounted for *g* and broad abilities simultaneously (Benson et al., 2016). Related to this second point, other studies have used observed scores (e.g., Konold & Canivez, 2010), and our use of latent factor scores may be more representative of the targeted construct of interest (Benson et al., 2016). We hypothesize these analytical decisions likely account for some of the differences reported across studies. Overall, the key takeaway from our findings, at least within this sample data and analytic protocol, is that there is no evidence of construct bias and little evidence of prediction bias with the WJ IV Cognitive and WJ IV Achievement.

4.2. Implications for research and practice

We found with the WJ IV that only *g*, G*c*, and G*s* made important contributions to predicting reading and mathematics. According to our findings, there was largely an absence of evidence of test bias, which is consistent with data obtained with intelligence tests over the past three decades. We specifically found an absence of prediction bias for African American and Hispanic youth. However, we did find evidence of slope bias for *g* and G*c* predicting basic reading skills in relation to Asian children and youth. Furthermore, our findings were observed after controlling for parent education levels and when using a latent variable hierarchical model of intelligence, showing evidence of moderation of CHC-based general and broad cognitive ability effects on basic reading skills. In our view, the higher-order model represents a plausible system of cognitive abilities to examine test bias, but there are alternative conceptualizations (e.g., bifactor model). We also obtained estimates from the first-order residual factor variances to validate the cognitive-achievement relations before examining potential racial or ethnic moderation (Caemmerer et al., 2018). Thus, there are a few implications for changes to research and practice that are warranted according to our study results.

Researchers engaged in test bias research should aim to control for variables that may produce spurious effects. For example, Weiss and Saklofske (2020) identified several SES-related indicators (e.g., parent education, income, occupation, academic expectations) that account for substantial variation in test performance between racial and ethnic groups (more so than racial or ethnic group endorsements alone explain). In our study, when we controlled for parent education levels, systematic group differences were substantially reduced, which may have artificially inflated findings of test bias otherwise. These findings highlight two of our main foci: (1) the need to account for contextual factors that might influence performance on cognitive ability and academic achievement assessments, and (2) bias in CARs due to how the tests and administrations interact with racial and ethnic differences. An additional important consideration that has not received due attention in prior test bias research is how differences in latent dimensions impact test bias findings. According to Borsboom et al. (2008) simulation study as well as our own experimental simulation results, when latent mean differences exist between groups, and factorial invariance is supported, then prediction bias is likely to be observed. Our simulation results suggest that when latent *g* mean differences exist (after establishing strong factorial invariance) and are constrained equal across groups, then bias will be introduced into the estimation of the regression parameters that are used to identify differential predictive validity. These methodological considerations should be accounted for in future test bias research.

Our limited findings of prediction bias did not result from measurement bias, which would likely have resulted in more problematic educational consequences. For example, given school psychologists and multi-disciplinary teams (MDTs) make identification and eligibility decisions on a case-by-case basis, implications of differential predictive validity are more consequential if derived from issues related to measurement bias due to group membership than due to differences in latent ability. With the former there would be a larger and more impactful concern about accurate and equitable measurement of intelligence, which our study showed consistent measurement across different groups. In the latter, errors resulting from differential predictive validity can be seen as unfair and how to address those concerns remains unresolved (e.g., Do you use different cut points based on group membership? Do these different cut points vary by co-normed test battery?). Nonetheless, our findings showed minimal evidence of test bias, but researchers should be aware of implications of not accounting for latent mean differences when evidence of prediction bias is found.

Implications for practice may also be warranted given the current study findings. We found evidence of slope bias where *g* was a stronger predictor of basic reading skills for Caucasian children and adolescents when compared to their Asian peers, and G*c* was a stronger predictor of basic reading skills for Asian children and adolescents (large effects) when compared to their Caucasian peers (moderate effects). Given the majority of referrals in school settings tend to be due to concerns with basic reading, this finding is worth consideration in practice. G*c* represents a culturally sensitive index of background knowledge, language, and vocabulary development (Schneider & McGrew, 2018). One seemingly obvious implication is to encourage the use of more linguistically and culturally neutral measures that reduce language and background knowledge requirements in the assessment of intelligence, at least for Asian children and adolescents with the WJ IV. Although slope bias is generally uncommon, it is more difficult to grapple with when found (Keith, 2019). If CARs systematically differ due to moderating factors, such as race and ethnicity, identification methods that rely on discrepancies from predicted achievement have the potential to over- or under-predict the actual achievement of individuals from subgroups within the normative sample. Although the systematic error in prediction occurs due to not accounting for moderating factors, the mis-prediction could emerge as a result of slope bias. In cases where bias does exist, a fairer solution may be to use predictors with smaller group differences by making comparisons with tests that evidence smaller mean differences between race-based groups (e.g., mean differences for the CAS-2 and KABC-II are approximately half of what is observed with the WJ or the Wechsler scales).

Although most researchers examine bias in CARs with a focus on global IQ, broad cognitive abilities continue to be employed within identification methods despite controversy for doing so (cf. Alfonso & Flanagan, 2018; McGill et al., 2018). Nonetheless, researchers and practitioners should be aware of potential bias in CARs with broad cognitive ability scores, even after controlling for variability due to *g* and parent education levels. The results from our race-moderated CAR model between *g*, G*c*, and basic reading achievement offer an example of differential slopes being the driver of prediction bias. For example, the difference in predicted achievement $(\hat{y})$ between Caucasian children and adolescents and their Asian peers can be observed in the linear formula for their unique best fit lines. More specifically, the best fit line for Caucasian children and adolescents follows the linear formula $\hat{y} = 1.76 + 0.65g + 0.32Gc$, whereas the best fit line for Asian children and adolescents follows the linear formula $\hat{y} = 22.24 + 0.19g + 0.63Gc$. Predictive bias emerges when applying these linear formulae to varying levels of G*c* while holding *g* constant. For example, when holding *g* constant at 85 and varying G*c* between 75, 85, and 95, predicted differences in achievement range from nearly 5 points (G*c* = 75) to nearly 11 points (G*c* = 95), favoring Asian children and adolescents. At higher levels of *g* (e.g., *g* = 115), the predictive bias is reduced with differences ranging from nearly 0 points (G*c* = 105) to nearly 6 points (G*c* = 125), favoring Asian children and adolescents. Maintaining the assumption that CARs are linear, which may be refuted in future research, the fluctuating difference in predicted achievement scores between racial and ethnic groups further complicates research and assessment efforts as it creates a moving target along the broad ability continuum (see also Keith, 2019; Kranzler et al., 1999).

4.3. Limitations and directions for future research

There are several limitations to our study that are worthy of consideration. Although we used latent variable SEM models that align with the WJ IV scoring model (McGrew et al., 2014), there are alternative theoretical and empirical conceptualizations of the WJ IV (Dombrowski et al., 2018; Niileksela et al., 2016; Woods Jr. et al., 2021). Furthermore, we only used two indicators per construct within our multiple group latent variable models. Future research may consider a cross-battery confirmatory factor analysis approach that allows for a larger sampling of indicators for each broad cognitive ability. Additionally, although the residual correlations between four subtests were allowed based on prior research and empirical justification from the current study, considerations should be given to the extent that this affects interpretation of the overall model and the extent that these subtests might be measuring independent subdomains.

Consistent with prior empirical research with the WJ IV, model fit statistics were acceptable for some models (especially for RMSEA and SRMR fit statistics), but mediocre for others (according to CFI fit statistic). Niileksela et al. (2016) hypothesized the use of data imputation methods during the development of the WJ IV norming sample as a plausible explanation for model fit issues, which could have affected the model fit results for the diverse populations examined in the current study. Future research should consider latent variable models with cognitive tests that are also relatively popular to use among school psychologists (e.g., WISC-V; Benson et al., 2019). The use of different IQ and achievement instruments would provide opportunities of replication to support the integrity of results; approximately only 1% of psychological studies are replication studies (Koole & Lakens, 2012; Makel et al., 2012). In addition, local fit considerations should be applied in future research as global model fit is not always adequate for judging model viability (Kline, 2016).

The range and depth of sampling of diverse populations may also have affected our results. The standard errors around the best fit line varied, suggesting there may have been insufficient power to be confident in the estimation of some of the regression parameters.

Future studies should employ larger and more racially diverse respondents to be confident about the best fit line for each respective group, which would tighten the standard errors and reduce Type II error. Although samples that are representative of national demographics can aid in broader generalizations, samples that are predominantly one subgroup (e.g., 80% Caucasian) undermine statistical power when conducting complex comparative analyses with smaller subgroups (e.g., 5% Asian). Additionally, the racial and ethnic variables used in our study are only proxies and do not necessarily tap into more meaningful and contextually relevant indices of cultural experiences and language proficiency (Weiss & Saklofske, 2020). As noted within the WJ IV Technical Manual (McGrew et al., 2014), there is no reference to language requirements throughout data collection process and the sampling of participants does not adequately account for cultural and linguistic differences. Thus, future research should operationalize and measure salient contextual variables and quantify their effects on cognitive ability and achievement directly. Another limitation related to the sampling is data were from the standardization sample and not specific to a special validity sample or disability group. Although these data represent a large, nationally representative sample, it is plausible findings of differential predictive validity may be observed in more academically at-risk samples. As an example, prior research has found that cognitive ability effects on academic achievement differ as a function of achievement level (Hajovsky, Oyen, et al., 2020; Hajovsky, Villeneuve, et al., 2020; Hajovsky, Villeneuve, Mason, & De Jong, 2018), and it is possible under- and over-prediction is more problematic at the margins of achievement. Future research should consider alternative analytical techniques to examine whether general findings of a lack of test bias are also present in different special populations.

Although latent *g* differences between groups may introduce bias into the estimation of prediction bias (when factorial invariance is tenable), especially when the latent mean of *g* is constrained equal across groups, we did not determine the degree to which these latent mean differences impact bias estimations. To conduct a study of this magnitude would require a larger set of analyses that were outside the scope of the current study. Our simulation results overlap with Borsboom et al.'s (2008) simulation results, but future research should examine the different ways latent mean differences can impact test bias results. Our study showed a general pattern of a lack of bias – both in measurement and prediction bias; we urge future researchers to consider and possibly account for latent mean differences in *g* between groups when conducting test bias research with intelligence and achievement constructs. Doing so might help explain findings of slope bias when factorial invariance is supported.

Finally, assessment is an iterative process that involves the ongoing systematic collection, organization, and interpretation of multiple data points (Wilcox & Schroeder, 2015). We only investigated cognitive and achievement scores and how racial and ethnic endorsements affect the relation between these scores. Given the concern with over- and under-representation of students from minoritized racial and ethnic groups within special education, the importance of SES differences across groups and how that may impact environmental opportunities also warrants consideration by school teams. Research has shown that uneven (significant interfactor variability) and even broad cognitive ability profiles do not differentially predict reading and mathematics outcomes (see Kotz et al., 2008), suggesting group differences on measured cognitive abilities may not result in differential predictive validities with reading and mathematics. This hypothesis was largely supported in our study. Any potential bias reported may be attenuated within MDT discussions as other relevant assessment data are incorporated into team decision-making (e.g., degree of language proficiency and level of acculturation effects on school learning; teacher support and school climate for addressing diversity and inclusion). As such, future research should address how MDTs engage the multiple ecological contexts to navigate the use of assessment scores across diverse populations. However, as reported within our study, there is no evidence of construct bias and little evidence of prediction bias.

5. Conclusion

School psychologists use cognitive ability and academic achievement tests to measure child and adolescent functioning within educational settings, which may differentially impact students from racially and ethnically diverse backgrounds. Collectively, our results suggest the WJ IV provides equally meaningful scores for individuals from different racial and ethnic groups with little evidence of differential predictive validity. The culturally sensitive, background knowledge and language-based variable of comprehension-knowledge predicted basic reading skills more strongly for Asian children and adolescents when compared to Caucasian peers, even after controlling for general intelligence and parent education levels. Furthermore, as latent mean differences in ability between groups and systematic differences in SES-indicators of test performance may confound conclusions about group differences on intelligence tests and their accuracy in predicting achievement, these variables should also be considered in research on test bias. Our findings should be replicated with other popular tests that are commonly used in comprehensive evaluations to provide evidence of equitable measurement and assessment practices of youth and adolescent children.

Acknowledgement

Supported by Texas Woman's University Woodcock Institute Research Grant. We have no conflicts of interest to disclose. Standardization data from the *Woodcock-Johnson® IV (WJ IV®)*. Copyright © by Houghton Mifflin Harcourt Publishing Company. All rights reserved. Used with permission of the publisher. The authors wish to thank Kiva Sam for feedback on early drafts. We are very appreciative of Drs. Joel Schneider and Randy Floyd for providing thoughtful and constructive feedback on our manuscript.

Appendix A

Appendix Table 1
Regression coefficients of parent college education on WJ-IV subtest scores from the measurement models for g and achievement.

			Caucasian				African American				Asian			
Parameter			Est	SE	Std	*p*	Est	SE	Std	*p*	Est	SE	Std	*p*
Measurement Model for g														
OrlVoc	on	Col Educ	6.229	0.567	0.208	< 0.001	7.206	1.307	0.234	< 0.001	7.700	2.244	0.225	< 0.001
GenInf	on	Col Educ	5.213	0.575	0.169	< 0.001	4.386	1.227	0.148	< 0.001	5.350	1.953	0.166	0.006
AnlSyn	on	Col Educ	5.137	0.569	0.170	< 0.001	5.055	1.277	0.167	< 0.001	6.029	2.103	0.186	0.004
ConFor	on	Col Educ	4.507	0.567	0.151	< 0.001	5.938	1.283	0.197	< 0.001	7.121	2.212	0.209	< 0.001
VrbAtt	on	Col Educ	5.120	0.560	0.171	< 0.001	6.352	1.204	0.216	< 0.001	4.896	1.963	0.149	0.013
NumRev	on	Col Educ	4.696	0.558	0.158	< 0.001	6.608	1.217	0.222	< 0.001	6.260	1.911	0.201	< 0.001
LtrPtr	on	Col Educ	2.725	0.543	0.095	< 0.001	4.526	1.225	0.148	< 0.001	4.298	1.770	0.140	0.015
ParCan	on	Col Educ	2.552	0.556	0.086	< 0.001	3.973	1.198	0.132	< 0.001	5.482	1.748	0.173	0.002
PhnPro	on	Col Educ	5.446	0.578	0.178	< 0.001	6.794	1.288	0.222	< 0.001	5.317	2.110	0.169	0.012
NwdRep	on	Col Educ	4.253	0.561	0.142	< 0.001	3.609	1.215	0.121	0.003	7.970	1.930	0.231	< 0.001
StyRec	on	Col Educ	4.835	0.566	0.161	< 0.001	3.525	1.276	0.114	0.006	4.480	2.023	0.133	0.027
VisAud	on	Col Educ	3.602	0.574	0.119	< 0.001	4.238	1.278	0.136	< 0.001	3.732	2.007	0.112	0.063
Visual	on	Col Educ	5.155	0.578	0.168	< 0.001	4.231	1.308	0.136	< 0.001	5.210	2.069	0.165	0.012
PicRec	on	Col Educ	2.641	0.569	0.087	< 0.001	3.448	1.221	0.114	0.005	3.608	1.870	0.107	0.054
Measurement Model for Achievement														
WrdAtk	on	Col Educ	5.473	0.580	0.180	< 0.001	6.769	1.305	0.226	< 0.001	5.519	2.515	0.159	0.028
LwIdnt	on	Col Educ	5.909	0.590	0.194	< 0.001	7.301	1.370	0.247	< 0.001	3.881	2.868	0.111	0.176
RdgRec	on	Col Educ	4.828	0.598	0.154	< 0.001	6.891	1.403	0.212	< 0.001	4.329	2.279	0.134	0.057
PsgCmp	on	Col Educ	5.965	0.593	0.194	< 0.001	7.898	1.456	0.249	< 0.001	7.638	2.589	0.231	003
NumMat	on	Col Educ	5.404	0.571	0.179	< 0.001	8.090	1.280	0.266	< 0.001	7.400	2.214	0.225	< 0.001
ApProb	on	Col Educ	5.852	0.574	0.197	< 0.001	7.626	1.361	0.258	< 0.001	7.585	2.630	0.230	0.004
MthFlu	on	Col Educ	3.922	0.577	0.130	< 0.001	6.043	1.356	0.192	< 0.001	2.970	2.247	0.095	0.186
Calcul	on	Col Educ	5.140	0.581	0.172	< 0.001	7.421	1.406	0.245	< 0.001	3.622	2.672	0.110	0.175

Note. g = higher order representation of general intelligence. Tables 1-1 and 1-3 in McGrew et al. (2014) provide full subtest names for each achievement and cognitive ability domain.

Appendix Table 2
Model parameters for first-order and second-order cognitive ability latent factors by race with strong factorial invariance and scaling via effects coding.

			Caucasian				African American				Asian			
Parameter			Est	SE	Std	*p*	Est	SE	Std	*p*	Est	SE	Std	*p*
Latent Factors														
Gc	by	OrlVoc	1.120	0.013	0.939	< 0.001	1.120	0.013	0.897	< 0.001	1.120	0.013	0.935	< 0.001
		GenInf	0.880	0.013	0.718	< 0.001	0.880	0.013	0.735	< 0.001	0.880	0.013	0.781	< 0.001
Gf	by	AnlSyn	0.972	0.016	0.675	< 0.001	0.972	0.016	0.650	< 0.001	0.972	0.016	0.621	< 0.001
		ConFor	1.028	0.016	0.722	< 0.001	1.028	0.016	0.690	< 0.001	1.028	0.016	0.624	< 0.001
Gwm	by	VrbAtt	0.993	0.020	0.643	< 0.001	0.993	0.020	0.599	< 0.001	0.993	0.020	0.636	< 0.001
		NumRev	1.007	0.020	0.658	< 0.001	1.007	0.020	0.600	< 0.001	1.007	0.020	0.679	< 0.001
Gs	by	LtrPtr	1.044	0.022	0.791	< 0.001	1.044	0.022	0.748	< 0.001	1.044	0.022	0.770	< 0.001
		ParCan	0.956	0.022	0.703	< 0.001	0.956	0.022	0.699	< 0.001	0.956	0.022	0.683	< 0.001
Ga	by	PhnPro	1.147	0.021	0.661	< 0.001	1.147	0.021	0.656	< 0.001	1.147	0.021	0.700	< 0.001
		NwdRep	0.853	0.021	0.504	< 0.001	0.853	0.021	0.502	< 0.001	0.853	0.021	0.475	< 0.001
Gl	by	StyRec	1.003	0.022	0.577	< 0.001	1.003	0.022	0.514	< 0.001	1.003	0.022	0.591	< 0.001
		VisAud	0.997	0.022	0.569	< 0.001	0.997	0.022	0.509	< 0.001	0.997	0.022	0.596	< 0.001
Gv	by	Visual	1.172	0.021	0.767	< 0.001	1.172	0.021	0.712	< 0.001	1.172	0.021	0.718	< 0.001
		PicRec	0.828	0.021	0.548	< 0.001	0.828	0.021	0.516	< 0.001	0.828	0.021	0.475	< 0.001
g	by	Gc	1.108	0.024	0.708	< 0.001	1.108	0.024	0.714	< 0.001	1.108	0.024	0.677	< 0.001
		Gf	1.175	0.021	0.901	< 0.001	1.175	0.021	0.926	< 0.001	1.175	0.021	0.991	< 0.001
		Gwm	0.967	0.022	0.801	< 0.001	0.967	0.022	0.867	< 0.001	0.967	0.022	0.802	< 0.001
		Gs	0.776	0.025	0.572	< 0.001	0.776	0.025	0.563	< 0.001	0.776	0.025	0.597	< 0.001
		Ga	1.025	0.021	0.932	< 0.001	1.025	0.021	0.932	< 0.001	1.025	0.021	0.932	< 0.001
		Gl	0.988	0.022	0.917	< 0.001	0.988	0.022	0.990	< 0.001	0.988	0.022	0.865	< 0.001
		Gv	0.961	0.023	0.770	< 0.001	0.961	0.023	0.809	< 0.001	0.961	0.023	0.867	< 0.001

(*continued on next page*)

Appendix Table 2 (*continued*)

Parameter			Caucasian Est	SE	Std	*p*	African American Est	SE	Std	*p*	Asian Est	SE	Std	*p*
Residual Covariance														
NumRev	with	LtrPtr	32.667	2.845	0.353	< 0.001	39.972	7.714	0.355	< 0.001	27.440	11.024	0.316	0.013
VrbAtt	with	NwdRep	34.881	3.340	0.252	< 0.001	34.591	8.080	0.242	< 0.001	54.141	15.168	0.361	< 0.001

Note. G*c* = Comprehension-Knowledge, G*wm* = Short-Term Working Memory, G*s* = Cognitive Processing Speed, G*a* = Auditory Processing, G*lr* = Long-Term Retrieval, G*v* = Visual Processing, *g* = higher order representation of general intelligence. Tables 1-1 and 1-3 in McGrew et al. (2014) provide full subtest names for each achievement and cognitive ability domain.

Appendix Table 3

Model parameters for first-order achievement latent factors by race with strong factorial invariance and scaling via effects coding.

Parameter			Caucasian Est	SE	Std	*p*	African American Est	SE	Std	*p*	Asian Est	SE	Std	*p*
Latent Factors														
BR	by	WrdAtk	0.900	0.010	0.748	< 0.001	0.900	0.010	0.742	< 0.001	0.900	0.010	0.791	< 0.001
		LwIdnt	1.100	0.010	0.915	< 0.001	1.100	0.010	0.919	< 0.001	1.100	0.010	0.964	< 0.001
RC	by	RdgRec	0.899	0.013	0.652	< 0.001	0.899	0.013	0.656	< 0.001	0.899	0.013	0.682	< 0.001
		PsgCmp	1.101	0.013	0.816	< 0.001	1.101	0.013	0.822	< 0.001	1.101	0.013	0.814	< 0.001
MP	by	NumMat	0.860	0.014	0.613	< 0.001	0.860	0.014	0.613	< 0.001	0.860	0.014	0.654	< 0.001
		ApProb	1.140	0.014	0.830	< 0.001	1.140	0.014	0.834	< 0.001	1.140	0.014	0.865	< 0.001
MC	by	MthFlu	0.880	0.011	0.705	< 0.001	0.880	0.011	0.703	< 0.001	0.880	0.011	0.730	< 0.001
		Calcul	1.120	0.011	0.906	< 0.001	1.120	0.011	0.928	< 0.001	1.120	0.011	0.875	< 0.001
Latent Variable Covariance														
BR	with	RC	127.141	4.329	0.910	< 0.001	123.798	10.232	0.857	< 0.001	132.521	18.613	0.865	< 0.001
		MP	91.737	3.757	0.692	< 0.001	86.318	8.614	0.653	< 0.001	95.153	16.806	0.606	< 0.001
		MC	103.105	4.014	0.695	< 0.001	98.164	9.344	0.642	< 0.001	103.629	17.538	0.640	< 0.001
RC	with	MP	87.829	3.647	0.737	< 0.001	88.749	8.948	0.699	< 0.001	99.908	15.593	0.795	< 0.001
		MC	99.931	3.908	0.750	< 0.001	110.244	9.942	0.750	< 0.001	116.715	16.702	0.901	< 0.001
MP	with	MC	118.837	4.023	0.939	< 0.001	118.634	9.565	0.883	< 0.001	128.495	17.186	0.968	< 0.001

Note. BR = Basic Reading, RC = Reading Comprehension, MP = Math Problem Solving, MC = Math Calculation. Tables 1-1 and 1-3 in McGrew et al. (2014) provide full subtest names for each achievement and cognitive ability domain.

Appendix Table 4

Intercept and slope coefficients of g predicting achievement by race.

Parameter	*g* Predict BR Est	SE	Std	*p*	*g* Predict RC Est	SE	Std	*p*	*g* Predict MP Est	SE	Std	*p*	*g* Predict MC Est	SE	Std	*p*
Intercept																
Caucasian	−12.07	3.03	−0.96	< 0.001	−9.04	2.90	−0.79	0.002	−24.39	2.53	−2.28	< 0.001	−10.91	2.90	−0.90	< 0.001
African American	−17.88	6.83	−1.34	< 0.001	−14.93	6.84	−1.12	0.048	−25.93	5.76	−2.26	< 0.001	−10.59	6.80	−0.74	0.173
Asian	−6.00	13.45	−0.43	0.661	−14.44	10.26	−1.14	0.189	−36.65	8.81	−3.24	< 0.001	−9.78	10.86	−0.81	0.385
Pairwise Comparisons (Intercept)																
Ca – AA	5.81	7.44	0.38	0.517	5.90	7.29	0.33	0.526	1.54	6.21	−0.02	0.939	−0.32	7.28	−0.16	0.838
Ca – As	−6.08	13.77	−0.54	0.661	5.40	10.55	0.35	0.665	12.27	9.11	0.96	0.186	−1.13	11.20	−0.10	0.906
As – AA	11.89	15.07	0.91	0.471	0.50	12.22	−0.02	0.996	−10.73	10.48	−0.98	0.269	0.81	12.77	−0.07	0.990
Slope																
Caucasian	1.12	0.03	0.71	< 0.001	1.09	0.03	0.77	< 0.001	1.25	0.03	0.93	< 0.001	1.11	0.03	0.74	< 0.001
African American	1.17	0.07	0.75	< 0.001	1.14	0.07	0.75	< 0.001	1.25	0.06	0.92	< 0.001	1.09	0.07	0.70	< 0.001
Asian	1.08	0.14	0.64	< 0.001	1.14	0.10	0.79	< 0.001	1.37	0.09	1.00	< 0.001	1.14	0.11	0.80	< 0.001
Pairwise Comparisons (Slope)																
Ca – AA	−0.05	0.08	−0.03	0.658	−0.05	0.07	0.02	0.752	−0.01	0.06	0.01	0.826	0.02	0.07	0.04	0.582
Ca – As	0.04	0.14	0.07	0.796	−0.05	0.11	−0.02	0.697	−0.13	0.09	−0.07	0.177	−0.03	0.11	−0.06	0.882
As – AA	−0.09	0.15	−0.10	0.648	0.01	0.12	0.04	0.884	0.12	0.11	0.08	0.192	0.05	0.13	0.09	0.609

Note. Ca = Caucasian, AA = African American, As = Asian, BR = Basic Reading, RC = Reading Comprehension, MP = Math Problem Solving, MC = Math Calculation, *g* = higher order representation of general intelligence.

Appendix Table 5
Results of direct and indirect relations between g and BR via broad abilities by race.

	Direct Effects												Indirect Effects			
	Intercept				*g* Predict BR				G*c* Predict BR				G*c* Predict BR via *g*			
	Est	SE	Std	*p*	Est	SE	Std	*p*	Est	SE	Std	*p*	Est	SE	Std	*p*
Racial Groups																
Caucasian	1.76	3.08	0.15	0.544	0.65	0.05	0.42	< 0.001	0.32	0.03	0.32	< 0.001	0.36	0.03	0.23	< 0.001
African American	2.09	6.60	0.19	0.718	0.51	0.11	0.33	< 0.001	0.45	0.07	0.46	< 0.001	0.51	0.08	0.33	< 0.001
Asian	22.24	12.51	1.60	0.073	0.19	0.19	0.11	0.312	0.63	0.12	0.59	< 0.001	0.70	0.13	0.42	< 0.001
Pairwise Comparisons																
Ca – AA	–0.33	7.22	–0.04	0.944	0.15	0.12	0.09	0.200	–0.13	0.08	–0.14	0.083	–0.15	0.08	–0.10	0.084
Ca – AA	–20.47	12.83	–1.44	0.110	0.46	0.20	0.31	0.019	–0.30	0.12	–0.27	0.012	–0.34	0.13	–0.19	0.012
As – AA	20.15	14.09	1.40	0.156	–0.32	0.22	–0.22	0.161	0.17	0.13	0.13	0.211	0.19	0.15	0.09	0.212

Note. Ca = Caucasian, AA = African American, As = Asian, BR = Basic Reading, RC = Reading Comprehension, G*c* = Comprehension-Knowledge, *g* = higher order representation of general intelligence.

Appendix Table 6
Results of direct and indirect relations between g and MP via broad abilities by race.

	Direct Effects												Indirect Effects			
	Intercept				*g* Predict MP				G*c* Predict MP				G*c* Predict MP via *g*			
	Est	SE	Std	*p*	Est	SE	Std	*p*	Est	SE	Std	*p*	Est	SE	Std	*p*
Racial Groups																
Caucasian	–18.66	2.58	–1.74	< 0.001	1.03	0.04	0.76	< 0.001	0.16	0.02	0.19	< 0.001	0.18	0.03	0.13	< 0.001
African American	–19.05	5.86	–1.70	0.002	1.01	0.10	0.74	< 0.001	0.18	0.06	0.20	0.004	0.20	0.07	0.14	0.004
Asian	–26.32	8.60	–2.28	0.003	1.01	0.13	0.73	< 0.001	0.27	0.08	0.31	< 0.001	0.30	0.08	0.22	< 0.001
Pairwise Comparisons																
Ca – AA	0.39	6.32	–0.04	0.996	0.02	0.10	0.02	0.758	–0.01	0.07	–0.01	0.846	–0.01	0.07	–0.01	0.846
Ca – AA	7.66	8.92	0.54	0.432	0.02	0.13	0.03	0.874	–0.11	0.08	–0.13	0.194	–0.12	0.09	–0.09	0.195
As – AA	–7.26	10.36	–0.59	0.501	0.00	0.16	–0.01	0.948	0.09	0.10	0.11	0.349	0.10	0.11	0.08	0.350

Note. Ca = Caucasian, AA = African American, As = Asian, MP = Math Problem Solving, G*c* = Comprehension-Knowledge, *g* = higher order representation of general intelligence.

References

Alfonso, V. C., & Flanagan, D. P. (Eds.). (2018). *Essentials of specific learning disability identification* (2nd ed.). Wiley.

American Educational Research Association, American Psychological Association, & National Council on Measurement on Education. (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Beaujean, A. A. (2015). John Carroll's views on intelligence: Bi-factor vs. higher-order models. *Journal of Intelligence, 3*(4), 121–136. https://doi.org/10.3390/jintelligence3040121.

Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 national survey. *Journal of School Psychology, 72*, 29–48. https://doi.org/10.1016/j.jsp.2018.12.004.

Benson, N. F., Kranzler, J. H., & Floyd, R. G. (2016). Examining the integrity of measurement of cognitive abilities in the prediction of achievement: Comparisons and contrasts across variables from higher-order and bifactor models. *Journal of School Psychology, 58*, 1–19. https://doi.org/10.1016/j.jsp.2016.06.001.

Benson, N. F., Maki, K. E., Floyd, R. G., Eckert, T. L., Kranzler, J. H., & Fefer, S. A. (2020). A national survey of school psychologists' practices in identifying specific learning disabilities. *School Psychology, 35*(2), 146–157. https://doi.org/10.1037/spq0000344.

Borsboom, D., Romeijn, J.-W., & Wicherts, J. M. (2008). Measurement invariance versus selection invariance: Is fair selection possible? *Psychological Methods, 13*(2), 75–98. https://doi.org/10.1037/1082-989X.13.2.75.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner, & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 793–828). Wiley.

Burchinal, M., McCartney, K., Steinberg, L., Crosnoe, R., Friedman, S. L., McLoyd, V., … NICHD Early Child Care Research Network. (2011). Examining the black-white achievement gap among low-income children using the NICHD study of early child care and youth development. *Child Development, 82*(5), 1404–1420. https://doi.org/10.1111/j.1467-8624.2011.01620.x.

Caemmerer, J. M., Keith, T. Z., & Reynolds, M. R. (2020). Beyond individual intelligence tests: Application of Cattell-Horn-Carroll theory. *Intelligence, 79*, Article 101433. https://doi.org/10.1016/j.intell.2020.101433.

Caemmerer, J. M., Maddocks, D. L. S., Keith, T. Z., & Reynolds, M. R. (2018). Effects of cognitive abilities on child and youth academic achievement: Evidence from the WISC-V and WIAT-III. *Intelligence, 68*, 6–20. https://doi.org/10.1016/j.intell.2018.02.005.

Cangur, S., & Ercan, I. (2015). Comparison of model fit indices used in structural equation modeling under multivariate normality. *Journal of Modern Applied Statistical Methods, 14*(1), 152–167. https://doi.org/10.22237/jmasm/1430453580.

Canivez, G. L. (2013). Incremental criterion validity of WAIS-IV factor index scores: Relationships with WIAT-II and WIAT-III subtest and composite scores. *Psychological Assessment, 25*(2), 484–495. https://doi.org/10.1037/a0032092.

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.

Decker, S. L., & Roberts, A. M. (2015). Specific cognitive predictors of early math problem solving. *Psychology in the Schools, 52*(5), 477–488. https://doi.org/10.1002/pits.21837.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*(2), 121–149.

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017). Exploratory and hierarchical factor analysis of the WJ-IV cognitive at school age. *Psychological Assessment, 29*(4), 394–407. https://doi.org/10.1037/pas0000350.

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018). An alternative conceptualization of the theoretical structure of the Woodcock-Johnson IV tests of cognitive abilities at school age: A confirmatory factor analytic investigation. *Archives of Scientific Psychology, 6*(1), 1–13. https://doi.org/10.1037/arc0000039.

Dupere, V., Leventhal, T., Crosnoe, R., & Dion, E. (2010). Understanding the positive role of neighborhood socioeconomic advantage in achievement: The contribution of the home, child care, and school environments. *Developmental Psychology, 46*(5), 1227–1244. https://doi.org/10.1037/a0020211.

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.

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18*(1), 5–21. https://doi.org/10.1080/10888438.2013.819356.

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review, 31*(2), 246–262. https://doi.org/10.1080/02796015.2002.12086154.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. (2006). *The achievement test desk reference (ATDR): A guide to learning disability identification* (2nd ed.). Wiley.

Floyd, R., Meisinger, E., Gregg, N., & Keith, T. (2012). An exploration of reading comprehension across development using models from Cattell-Horn-Carroll theory: Support for integrative models of reading. *Psychology in the Schools, 49*(8), 725–743. https://doi.org/10.1002/pits.21633.

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools, 40*(2), 155–171. https://doi.org/10.1002/pits.10083.

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. *The Journal of Special Education, 40*(2), 103–114.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*(1), 79–132. https://doi.org/10.1016/S0160-2896(97)90014-3.

Gottfredson, L. S. (2016). A *g* theorist on why Kovacs and Conway's process overlap theory amplifies, not opposes, *g* theory. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory, 27*(3), 210–217. https://doi.org/10.1080/1047840X.2016.1203232.

Gould, S. J. (1981). *The mismeasure of man.* Norton.

Graves, S. L., Jr., Smith, L. V., & Nichols, K. D. (2021). Is the WISC-V a fair test for black children: Factor structure in an urban public school. *Contemporary School Psychology, 25*, 157–169. https://doi.org/10.1007/s40688-020-00306-9.

Grigorenko, E. L., Compton, D. L., Fuchs, L. S., Wagner, R. K., Willcutt, E. G., & Fletcher, J. M. (2020). Understanding, educating, and supporting children with specific learning disabilities: 50 years of science and practice. *American Psychologist, 75*(1), 37–51. https://doi.org/10.1037/amp0000452.

Hajovsky, D. B., Oyen, K. A., Chesnut, S. R., & Curtin, S. J. (2020). Teacher-student relationship quality and math achievement: The mediating role of teacher self-efficacy. *Psychology in the Schools, 57*(1), 111–134. https://doi.org/10.1002/pits.22322.

Hajovsky, D. B., Reynolds, M. R., Floyd, R. G., Turek, J. J., & Keith, T. Z. (2014). A multigroup investigation of latent cognitive abilities and reading achievement relations. *School Psychology Review, 43*(4), 385–406. https://doi.org/10.1080/02796015.2014.12087412.

Hajovsky, D. B., Villeneuve, E. F., Mason, B. A., & De Jong, D. A. (2018). A quantile regression analysis of cognitive ability and spelling predictors of written expression: Evidence of gender, age, and skill level moderation. *School Psychology Review, 47*(3), 291–315. https://doi.org/10.17105/SPR-2017-0110.V47-3.

Hajovsky, D. B., Villeneuve, E. F., Reynolds, M. R., Niileksela, C. R., Mason, B. A., & Shudak, N. J. (2018). Cognitive ability influences on written expression: Evidence for developmental and sex-based differences in school-age children. *Journal of School Psychology, 67*, 104–118. https://doi.org/10.1016/j.jsp.2017.09.001.

Hajovsky, D. B., Villeneuve, E. F., Schneider, W. J., & Caemmerer, J. M. (2020). An alternative approach to cognitive and achievement relations research: An introduction to quantile regression. *Journal of Pediatric Neuropsychology, 6*, 83–95. https://doi.org/10.1007/s40817-020-00086-3.

Helms, J. E., Jernigan, M., & Mascher, J. (2005). The meaning of race in psychology and how to change it: A methodological perspective. *American Psychologist, 60*(1), 27–36. https://doi.org/10.1037/0003-066X.60.1.27.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* Free Press.

Iacobucci, D. (2010). Structural equation modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology, 20*(1), 90–98. https://doi.org/10.1016/j.jcps.2009.09.003.

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

Jewsbury, P. A., Bowden, S. C., & Duff, K. (2017). The Cattell-Horn-Carroll model of cognition for clinical assessment. *Journal of Psychoeducational Assessment, 35*(6), 547–567.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock-Johnson and Kaufman tests. *Intelligence, 40*(2), 123–138. https://doi.org/10.1016/j.intell.2012.01.009.

Keith, T. Z. (1999). Effects of general and specific abilities on student achievement: Similarities and differences across ethnic groups. *School Psychology Quarterly, 14*(3), 239–262. https://doi.org/10.1037/h0089008.

Keith, T. Z. (2019). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (3rd ed.). Taylor & Francis.

Keith, T. Z., & Reynolds, M. R. (2018). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan, & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 853–900) Guilford.

Kilpatrick, D. A. (2016). *Essentials of assessing, preventing, and overcoming reading difficulties.* Wiley.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Konold, T. R., & Canivez, G. L. (2010). Differential relationships between WISC-IV and WIAT-II scales: An evaluation of potentially moderating child demographics. *Educational and Psychological Measurement, 70*(4), 613–627.

Koole, S. L., & Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science, 7*(6), 608–614.

Kotz, K. M., Watkins, M. W., & McDermott, P. A. (2008). Validity of the general conceptual ability score from the differential ability scales as a function of significant and rare interfactor variability. *School Psychology Review, 37*(2), 261–278. https://doi.org/10.1080/02796015.2008.12087899.

Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry, 27*(3), 151–177. https://doi.org/10.1080/1047840X.2016.1153946.

Kranzler, J. H., Floyd, R. G., Bray, M. A., & Demaray, M. K. (2020). Past, present, and future of research in school psychology: The biopsychosocial ecological model as an overarching framework. *School Psychology, 35*(6), 419–427.

Kranzler, J. H., Miller, M. D., & Jordan, L. (1999). An examination of racial/ethnic and gender bias on curriculum-based measurement of reading. *School Psychology Quarterly, 14*(3), 327–342. https://doi.org/10.1037/h0089012.

Lange, N., Froimowitz, M. P., Bigler, E. D., Lainhart, J. E., & Brain Development Cooperative Group. (2010). Associations between IQ, total and regional brain volumes, and demography in a large normative sample of healthy children and adolescents. *Developmental Neuropsychology, 35*(3), 296–317. https://doi.org/10.1080/87565641003696833.

Lemos, G. C., Almeida, L. S., & Colom, R. (2011). Intelligence of adolescents is related to their parents' educational level but not to family income. *Personality and Individual Differences, 50*(7), 1062–1067. https://doi.org/10.1016/j.paid.2011.01.025.

Lewno-Dumdie, B. M., & Hajovsky, D. B. (2020). Do global ability composites differentially predict academic achievement across gender? *Journal of Psychoeducational Assessment, 38*(6), 706–723.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*(1), 53–76. https://doi.org/10.1207/s15327906mbr3201_3.

Little, T. D. (2013). *Longitudinal structural equation modeling.* New York: Guilford Press.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "general intelligence, objectively determined and measured". *Journal of Personality and Social Psychology, 86*(1), 96–111. https://doi.org/10.1037/0022-3514.86.1.96.
Magnuson, K. A., Sexton, H. R., Davis-Kean, P. E., & Huston, A. C. (2009). Increases in maternal education and young children's language skills. *Merrill-Palmer Quarterly, 55*, 319–350. https://www.jstor.org/stable/23096260?seq=1#metadata_info_tab_contents.
Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7*(6), 537–542.
Maki, K. E., Floyd, R. G., & Roberson, T. (2015). State learning disability eligibility criteria: A comprehensive review. *School Psychology Quarterly, 30*(4), 457–469. https://doi.org/10.1037/spq0000109.
McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Cognitive profile analysis in school psychology: History, issues, and continued concerns. *Journal of School Psychology, 71*, 108–121. https://doi.org/10.1016/j.jsp.2018.10.007.
McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual*. Riverside.
McGrew, K. S., & Wendling, B. J. (2010). Cattell-Horn-Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47*(7), 651–675. https://doi.org/10.1002/pits.20497.
Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741.
Morgan, P. L., Farkas, G., Cook, M., Wilson, N., Hillemeier, M. M., & Schussler, D. (2017). Are black children disproportionately overrepresented in special education? A best-evidence synthesis. *Exceptional Children, 83*, 181–198.
Murray, C. (1998). *Income, inequality, and IQ*. American Enterprise Institute.
Naglieri, J. A., Rojahn, J., & Matto, H. C. (2007). Hispanic and non-Hispanic children's performance on PASS cognitive processes and achievement. *Intelligence, 35*(6), 568–579. https://doi.org/10.1016/j.intell.2006.11.001.
Naglieri, J. A., Rojahn, J., Matto, H. C., & Aquilino, S. A. (2005). Black-White differences in cognitive processing: A study of the planning, attention, simultaneous, and successive theory of intelligence. *Journal of Psychoeducational Assessment, 23*(2), 146–160.
Niileksela, C. R., Reynolds, M. R., Keith, T. Z., & McGrew, K. S. (2016). A special validity study of the Woodcock–Johnson IV: Acting on evidence for specific abilities. In D. P. Flanagan, & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 65–106). Elsevier.
Paschall, K. W., Gershoff, E. T., & Kuhfeld, M. (2018). A two decade examination of historical race/ethnicity disparities in academic achievement by poverty status. *Journal of Youth and Adolescence, 47*, 1164–1177. https://doi.org/10.1007/s10964-017-0800-7.
Plomin, R. (2018). *Blueprint: How DNA makes us who we are*. MIT Press.
Potthoff, R. F. (1978). *Statistical aspects of the problem of biases in psychological tests (Institute of Statistics Mimeo Series no 479)*. University of North Carolina, Department of Statistics.
R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/.
Reynolds, C. R., & Lowe, P. A. (2009). The problem of bias in psychological testing. In T. G. Gutkin, & C. R. Reynolds (Eds.), *School psychology handbook* (4th ed., pp. 332–374). Wiley.
Reynolds, M. R., & Keith, T. Z. (2017). Multi-group and hierarchical confirmatory factor analysis of the Wechsler intelligence scale for children – Fifth edition: What does it measure? *Intelligence, 62*, 31–47. https://doi.org/10.1016/j.intell.2017.02.005.
Rhemtulla, M., & Little, T. D. (2012). Planned missing data designs for research in cognitive development. *Journal of Cognition and Development, 13*(4), 425–438. https://doi.org/10.1080/15248372.2012.717340.
Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.6-9 (BETA). *Journal of Statistical Software, 48*(2), 1–36.
Sameroff, A. (2010). A unified theory of development: A dialectic integration of nature and nurture. *Child Development, 81*, 6–22. https://doi.org/10.1111/j.1467-8624.2009.01378.x.
Scheiber, C. (2016a). 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 years? *Journal of Pediatric Neuropsychology, 2*, 79–88. https://doi.org/10.1007/s40817-016-0019-7.
Scheiber, C. (2016b). Do the Kaufman tests of cognitive ability and academic achievement display construct bias across a representative sample of Black, Hispanic, and Caucasian school-age children in grades 1 through 12? *Psychological Assessment, 28*(8), 942–952. https://doi.org/10.1037/pas0000236.
Scheiber, C., & Kaufman, A. S. (2015). Which of the three KABC-II global scores is the least biased? *Journal of Pediatric Neuropsychology, 1*, 21–35. https://doi.org/10.1007/s40817-015-0004-6.
Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan, & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163) Guilford.
Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV tests of achievement*. Riverside.
Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV tests of cognitive abilities*. Riverside.
Selita, F., & Kovas, Y. (2019). Genes and gini: What inequality means for heritability. *Journal of Biosocial Science, 51*(1), 18–47. https://doi.org/10.1017/S0021932017000645.
Shifrer, D., Muller, C., & Callahan, R. (2011). Disproportionality and learning disabilities: Parsing apart race, socioeconomic status, and language. *Journal of Learning Disabilities, 44*, 246–257.
Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453.
von Stumm, S. (2017). Socioeconomic status amplifies the achievement gap throughout compulsory education independent of intelligence. *Intelligence, 60*, 57–62. https://doi.org/10.1016/j.intell.2016.11.006.
Sullivan, A. L., & Bal, A. (2013). Disproportionality in special education: Effects of individual and school variables on disability risk. *Exceptional Children, 79*(4), 475–494.
Taub, G. E., Keith, T. Z., Floyd, R. G., & McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly, 23*(2), 187–198. https://psycnet.apa.org/buy/2008-07564-003.
Trundt, K. M., Keith, T. Z., Caemmerer, J. M., & Smith, L. V. (2018). Testing for construct bias in the differential ability scales, second edition: A comparison among African American, Asian, Hispanic, and Caucasian children. *Journal of Psychoeducational Assessment, 36*(7), 670–683.
Tucker-Drob, E. M., & Bates, T. C. (2016). Large cross-national differences in gene x socioeconomic status interaction on intelligence. *Psychological Science, 27*(2), 138–149.
Tucker-Drob, E. M., & Harden, K. P. (2012). Early childhood cognitive development and parental cognitive stimulation: Evidence for reciprocal gene-environment transactions. *Developmental Science, 15*(2), 250–259. https://doi.org/10.1111/j.1467-7687.2011.01121.x.
Villeneuve, E. F., Hajovsky, D. B., Mason, B. A., & Lewno, B. M. (2019). Cognitive ability and math computation developmental relations with math problem-solving: An integrated, multi-group approach. *School Psychology, 34*(1), 96–108. https://doi.org/10.1037/spq0000267.
Warne, R. T., Yoon, M., & Price, C. J. (2014). Exploring the various interpretations of "test bias". *Cultural Diversity and Ethnic Minority Psychology, 20*(4), 570–582. https://doi.org/10.1037/a0036503.
Weiss, L. G., Harris, J. G., Prifitera, A., Courville, T., Rolfhus, E., Saklofske, D. H., & Holdnack, J. A. (2006). WISC-IV interpretation in societal context. In L. G. Weiss, D. H. Saklofske, A. Prifitera, & J. A. Holdnack (Eds.), *WISC-IV advanced clinical interpretation* (pp. 1–57). Elsevier Science.
Weiss, L. G., & Prifitera, A. (1995). An evaluation of differential prediction of WIAT achievement scores from WISC-III FSIQ across ethnic and gender groups. *Journal of School Psychology, 33*(4), 297–304. https://doi.org/10.1016/0022-4405(95)00016-F.
Weiss, L. G., Prifitera, A., & Roid, G. (1993). The WISC-III and the fairness of predicting achievement across ethnic and gender groups. *Journal of Psychoeducational Assessment*, 35–42. Monograph Series, Advances in Psychoeducational Assessment: Wechsler Intelligence Scale for Children-Third Edition.
Weiss, L. G., & Saklofske, D. H. (2020). Mediators of IQ test score differences across racial and ethnic groups: The case for environmental and social justice. *Personality and Individual Differences, 161*(15), Article 109962. https://doi.org/10.1016/j.paid.2020.109962.

Wicherts, J. M., & Millsap, R. E. (2009). The absence of underprediction does not imply the absence of measurement bias. *American Psychologist, 64*(4), 281–283. https://doi.org/10.1037/a0014992.

Wilcox, G., & Schroeder, M. (2015). What comes before report writing? Attending to clinical reasoning and thinking errors in school psychology. *Journal of Psychoeducational Assessment, 33*(7), 652–661.

Woods, I. L., Jr., Niileksela, C., & Floyd, R. G. (2021). Do Cattell-Horn-Carroll cognitive abilities predict reading achievement similarly for black children as for other racial/ethnic groups? *Contemporary School Psychology, 25*, 183–199. https://doi.org/10.1007/s40688-020-00338-1.

Zaboski, B. A., Kranzler, J. H., & Gage, N. A. (2018). Meta-analysis of the relationship between academic achievement and broad abilities of the Cattell-Horn-Carroll theory. *Journal of School Psychology, 71*, 42–56. https://doi.org/10.1016/j.jsp.2018.10.001.