EXPLAINING THE BLACK-WHITE GAP IN COGNITIVE TEST SCORES:
TOWARD A THEORETICAL MODEL OF ADVERSE IMPACT

BY
JONATHAN COTTRELL

THESIS
Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Adviser:
Associate Professor Daniel Newman
ABSTRACT

In understanding the causes of adverse impact, a key parameter is the Black-White difference in cognitive ability test scores. To advance theoretical understanding of why there exist Black-White cognitive test score gaps, and of how these gaps develop over time, the current paper proposes an inductive explanatory model derived from past empirical findings. According to this theoretical model, Black-White group mean differences in cognitive test scores arise from the following disparate conditions: child birth order, maternal cognitive ability scores, the presence of learning materials in the home, parenting factors (maternal warmth and acceptance, safe physical environment, and maternal sensitivity), child birthweight, and family income. Results from a growth model estimated on children in the Study of Early Child Care and Youth Development from age 4 years through age 15 show significant Black-White cognitive test score gaps throughout children’s early development, but these gaps do not grow significantly over time (i.e., significant intercept differences, but not slope differences). Further, the first four disparate conditions listed above fully account for the relationship between race and cognitive test scores. We conclude by proposing a parsimonious four-channel model that fully explains the racial cognitive test score gap. These results attempt to fill a longstanding need for theory on the etiology of the Black-White ethnic group gap in cognitive test scores, suggesting adverse impact may have developmental origins that begin before applicants even start searching for jobs.
CHAPTER 1
INTRODUCTION

In the study of adverse impact, racial group differences in cognitive ability test scores represent a classic problem (Goldstein, Scherbaum, & Yusko, 2010; Outtz, 2010; Schmitt & Quinn, 2010; Zedeck, 2010). To elaborate, personnel selection practice is a key mechanism by which individuals from different racial backgrounds gain access to jobs. Nonetheless, a major empirical tension has plagued the process of hiring and admissions, as pertains to adverse impact/diversity and job performance (Outtz, 2010; Sackett, Schmitt, Ellingson, & Kabin, 2001). In particular, cognitive ability tests robustly predict job performance across many different job types, and are considered the predictor of choice for achieving maximal job performance (Hunter & Hunter, 1984; Schmidt & Hunter, 1998; 2004). On the other hand, these cognitive tests also show very large Black-White subgroup differences, with an average Cohen’s $d$, or standardized mean difference, of 1.0 (Roth, Bevier, Bobko, Switzer, & Tyler, 2001; cf. Sackett & Shen, 2010). In other words, using cognitive tests for hiring purposes will largely exclude African-American job applicants (Bobko, Roth, & Potosky, 1999; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997), and can lead to large-scale race disparities in occupational attainment. This empirical tension is compounded by meta-analytic findings that Black-White differences in job performance are only one-third as large as Black-White differences on cognitive tests (McKay & McDaniel, 2006), which suggest that cognitive test scores are much more strongly race-loaded than is job performance itself (Outtz & Newman, 2010). This empirical tension is troubling for companies who want to have a workforce that is both diverse and maximally productive (De Corte, Lievens, & Sackett, 2007). The current paper attempts to better inform this empirical tension, by theoretically explaining the origins of race differences in cognitive test scores.
Many possibilities have been suggested for how to resolve this empirical tension between the large criterion validity and large ethnic subgroup differences on cognitive tests (for a review of recommendations, see Ployhart & Holtz, 2008). In the short term, selection systems can give greater weight to non-cognitive tests, such as personality tests, interviews, assessment centers, or situational judgment tests. While these tests do in fact show less adverse impact, they also are less valid as predictors of job performance, and would therefore have lower monetary utility for firms if used as a substitute for cognitive tests (Ployhart & Holtz, 2008; Schmidt & Hunter, 1998; Hunter & Hunter, 1984). Discarding cognitive tests would thus be counterproductive in attempting to maximally predict job performance. Optimal tradeoffs of validity and diversity in hiring and admissions continue to be explored (e.g., De Corte et al., 2007; De Corte, Sackett, & Lievens, 2010).

The current paper takes a much more limited approach, by attempting to offer a theoretical rationale for one side of the empirical tension that undergirds the adverse impact problem—i.e., by explaining the origins of the race gap in cognitive test scores. As such, we hope to provide a more parsimonious theory of adverse impact (cf. Outtz & Newman, 2010). The purpose of this paper is to attempt to fill a hole in the adverse impact literature by examining the conditions that give rise to the Black-White cognitive test score gap. This is an important contribution to the study of adverse impact, because industrial/organizational psychology researchers to date have tended to avoid building theoretical explanations for why Black-White cognitive test score gaps exist. Outtz (2010) notes that few attempts have been made to address adverse impact from a theoretical perspective, or to explore why adverse impact occurs. Understanding how and why adverse impact develops is a critical first step toward allowing scientists and practitioners to address and perhaps prevent such gaps in the future.
In order to understand how adverse impact develops, it is important to establish when cognitive test score gaps begin and to examine important predictors of a child’s cognitive development. The section to follow will be a discussion of when adverse impact can be said to begin. Many other fields, such as economics, educational psychology, and developmental psychology, have already attempted to explore the question of why Black-White test score gaps exist, and we will draw upon these literatures to derive our own, more parsimonious theoretical model. Finally, we discuss each of the mechanisms in our model that we believe can explain Black-White cognitive test score gaps.

**When do Cognitive Test Score Gaps Begin?**

In order to understand how adverse impact develops, one must explore when Black-White cognitive test score differences start. Fryer and Levitt (2004; 2006) report a Black-White gap to exist even as young as 5 years old. Gaps persist throughout childhood, through elementary school (Fryer & Levitt, 2006), and into high school (Yeung & Pfeiffer, 2009). When children and adolescents finish their schooling and attempt to enter the workforce, we contend that these cognitive test score gaps then translate into adverse impact in the hiring process. Investigating why these cognitive test score gaps arise is one key to building a theoretical understanding of adverse impact.

The next section of the paper will examine previous attempts to explain the Black-White test score gap. Strengths and weaknesses of past research designs used to study this topic are discussed. In this review, we hope to summarize what has been done so far to theoretically account for the Black-White gap, as well as to highlight what the current paper uniquely contributes.

**Previous Attempts to Explain the Black-White Cognitive Test Score Gap**
Despite the fact that the Black-White gap in cognitive test scores has been discussed for over 90 years (Popenoe, 1922), surprisingly few studies have attempted to empirically test integrated theoretical models that explain this gap. Nonetheless, there have been a handful of attempts to quantitatively explain the Black-White test score gap, as summarized in Table 1. Note that Table 1 does not include articles that only quantify the size of the gap (e.g., Roth et al., 2001), nor does Table 1 include articles which provide only theoretical explanations for the Black-White gap, with no data (e.g., Garcia Coll et al., 1996; Brooks-Gunn & Markman, 2005). The studies described in Table 1 all attempt to both quantify the size of the Black-White test score gap, and also to empirically test covariates that can be used to explain the gap.

Limitations of Previous Research

Whereas we consider all of the papers in Table 1 to be commendable attempts to specify the reasons for the Black-White cognitive test score gap, we believe each is lacking in particular ways. These particular deficits in past studies are: (a) lack of parsimony in the selection of covariates and failure to report results for all covariates, (b) reliance on cross-sectional analyses or on longitudinal analyses that switching indicators across time without establishing measurement invariance, (c) restricted sampling on income or birthweight, (d) peculiarly-coded covariates, (e) using non-standard cognitive tests, and (f) leaving much of the Black-White cognitive test score gap unexplained. Below we briefly discuss each of these past limitations.

Lack of parsimony and failure to report full results. One important aspect of a theoretical model to explain the Black-White cognitive test score gap is parsimony, or using as few explanatory variables as sufficient to explain the gap (Muliak et al., 1989). Some of the studies in Table 1 use a large number of variables. For example, Yeung and Pfeiffer (2009) used 19 covariates, Burchinal et al. (2011) used 11 covariates, Brooks-Gunn, Klebanov, Smith,
Duncan, and Lee (2003) use 10 covariates, and Fryer and Levitt (2004; 2006) and Mandara, Varner, Greene, and Richman (2009) use 9 covariates each. However, some of these models might in actuality be even less parsimonious than they seem, as explained below.

An interesting practice in some of these studies is to report only statistically significant covariates (e.g., Burchinal et al., 2011), which makes their model seem smaller than it actually is. Further, Brooks-Gunn et al. (2003) only report covariates that explain the Black-White gap for the Peabody Picture Vocabulary Test-Revised (PPVT-R), but do not report any results for the Stanford-Binet or Wechsler tests which were also administered. Thus, their results are unclear and, as mentioned earlier, only reporting significant results masks the lack of parsimony in their models. In sum, two potential problems that plague some of the past studies of the Black-White gap—ineffective reporting practices and lack of parsimony—are difficult to tease apart. To address this limitation, the current study seeks to use an a priori approach to identify a small number of covariates, and then—importantly—to report all results from each model tested.

**Cross-sectional analyses and switching indicators across time.** Another major issue is that certain papers examine only a small portion of the lifetime of children. Brooks-Gunn et al. (2003) only examined children from ages 3-4 and 5-6. Mandara et al. (2009) also examined only two points in time (ages 10-11 and ages 13-14). Yeung and Pfeiffer (2009) had different groups of participants at various time points, but also featured only 2 time points per group. Measuring cognitive ability and other variables at only two time points limits our ability to examine developmental trends in the Black-White gap over time.

Further, when measuring cognitive ability across more than two time points, it is important to use the same tests at every time point, or else to establish that different tests tap the same underlying construct across time. Relatedly, Burchinal et al. (2011) assessed developmental
trends in math and reading ability while operationalizing math and reading using tests that varied over time. Because no evidence was given that the different tests were equivalent measures of math and reading (i.e., no measurement equivalence assessment nor model constraints; Chan, 1998) it is difficult to discern whether ostensible subgroup time trends were due to actual change in ability versus due to switching to a different measure of the construct at later time points.

**Restricted sampling.** In contrast to most other studies of the Black-White gap, Burchinal et al. (2011) conducted a proper longitudinal study on the same participants across 4 time points. One natural consequence of this approach was that their study featured a relatively smaller sample size in comparison to the other studies. However, the small sample size problem was perhaps exacerbated by the authors’ choice to restrict their longitudinal sample to low income children only (defined as income which is 225% of the poverty line or lower), which was done as a way to control for the confounding influence of income. As a result of this decision, only 314 children across the 4 time points were examined and over half of Burchinal et al.’s original sample was deleted (i.e., several hundred participants were left out). Cutting out such a large number of participants nonrandomly can greatly reduce the generalizability of one’s estimate of the Black-White test score gap, as well as one’s inferences about which covariates can explain the gap. Another example of restricted sampling occurs in Brooks-Gunn et al. (2003), where one sample features only low birthweight children, defined as children weighing 2.5 kilograms or less at birth.

**Peculiar coding of covariates.** Some articles feature unusual or unclear methods for how covariates were coded. For example, Fryer and Levitt (2004, 2006) separately code whether the mother was a teenager at first birth and whether the mother was 30 or older at first birth. A clearer method of coding would simply be to create a continuous variable for maternal age.
Yeung and Pfeiffer (2009) also report separate regression coefficients for different levels of income from birth to age 5 (e.g., $15,000-24,999, $75,000+, etc.) as well as for net wealth (split into quartiles), instead of creating one continuous income variable and one continuous wealth variable. Burchinal et al. (2011) insert site of data collection as a covariate, but did not report how they entered the various site locations into their regression equations. Additionally, birth order is dichotomized (firstborn versus not firstborn) instead of using the actual birth order (e.g., firstborn, second born, etc.) in the data. The potential problem with these studies is that peculiar coding of variables can affect the significance of covariates, as well as the overall relationship between race and cognitive test scores.

**Using non-standard cognitive tests.** Several papers have used non-standard cognitive ability tests, or combined several measures with no clear justification. Brooks-Gunn et al. (2003), for example, examine different cognitive tests for different ages, with no discussion about whether these measurements can be meaningfully compared to each other. Fryer and Levitt (2004, 2006) examine cognitive ability tests created exclusively for the Early Childhood Longitudinal Study (ECLS). Such tests are claimed to be based on previously existing and validated instruments, such as the Peabody Picture Vocabulary Test and the Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R). However, the authors do not made clear how comparable these tests actually are to other, more rigorously validated cognitive tests. Such measures need to be thoroughly validated to show that they are psychometrically sound before one can assume that they are equivalent to other well-established cognitive ability tests.

**Much of the gap is left unexplained.** Finally, several past models do not succeed in completely explaining the Black-White cognitive test score gap. Brooks-Gunn et al. (2003) are able to explain around 61% of the gap in their Infant Health and Development Program (IHDP)
sample and around 39% of the gap in their National Longitudinal Study of Youth-Child Supplement (NLSY-CS) sample, which constitutes about half of the gap, on average. Fryer and Levitt (2004, 2006) are able to sufficiently explain the Black-White math gap in kindergarten but cannot completely account for the gap in reading or math at later time points. Yeung and Pfeiffer (2009) find both statistically significant and non-significant Black-White gaps after controlling for covariates, depending on both the cohort as well as which subtest (math or reading) is examined.

**Novelty of current study**

The current paper attempts to integrate the findings from past studies of the cognitive test score gap, by specifying and testing the fit of an intact theoretical model that addressed all of the problems enumerated above that have appeared in previous research designs. The current study uses a relatively small number of variables to explain the entire Black-White cognitive test score gap. Regression results are fully reported. The current paper analyzes data on the same children at five time points, from 54 months to 15 years--more than any previous research design. Tests for measurement equivalence across time are conducted, which are necessary to show that the measures of ability are comparable over time. Thus, cognitive test data can be analyzed in the same children through both childhood and adolescence. Additionally, we do not restrict our sample by eliminating participants with moderate incomes; and we avoid peculiar coding of covariates. Finally, we use psychometrically validated cognitive tests. Thus, with a large sample (over 700 respondents), more time points than previous studies (from 4 years to 15 years), a parsimonious model with full reporting of all model results, as well as a measurement model of cognitive ability tests that establishes measurement equivalence over time, the current paper
attempts to make a novel contribution to both the cognitive development and adverse impact literatures.

**Explanatory Variables identified in Past Research on the Cognitive Test Gap.**

Table 2 summarizes past research, by enumerating the covariates that have repeatedly been found statistically significant in explaining the Black-White cognitive test score gap. In other words, these are the explanatory concepts whose unique statistical significance has been replicated (i.e., been found in more than one past study). These explanatory concepts, each of which describes a set of disparate conditions between Black children and White children, are: birth order, maternal cognitive/achievement test scores, learning materials in the home, parenting factors (maternal sensitivity, warmth and acceptance, physical environment), birthweight, and SES.

The next section of the current paper proposes a theoretical model in which the above-listed concepts/disparate conditions explain the relationship between race and cognitive test scores. These concepts are chosen based on their use in previous research, but are also explicated using the strong theoretical literature available for each set of conditions (e.g., Garcia-Coll et al., 1996). We believe that future adverse impact researchers will benefit from including such variables in their theoretical explanations for the origins of Black-White cognitive test score gaps (e.g., see Outtz & Newman, 2010). Each of these explanatory concepts—described in the next section—is theorized to correlate both with cognitive test scores and with race.

**Six Potential Channels Connecting Race to Cognitive Ability**

*Birth Order.* As in any environment with limited resources, children are often competing with their siblings for their parents’ time and attention. In large families, intellectual and maternal resources may not be shared equally for a variety of reasons. Studies have shown that
earlier born children have higher cognitive test scores than their younger siblings (Black, Devereux, & Salves, 2005; Booth & Kee, 2009). Many theoretical rationales have been traditionally proposed to explain the advantages of older siblings with respect to cognitive development: (a) firstborn and early-born children receive greater parental inputs, especially time, than later born children who have to compete with their older siblings for parental attention (Behrman & Taubman, 1986), (b) children who are born earlier also may gain an advantage because parents can spend a larger share of their income on their only child’s development (Becker & Lewis, 1973), and (c) older and firstborn children may also get a greater share of educational resources, such as books, because they have fewer siblings with whom to share these educational resources (Booth & Kee, 2009).

Another explanation for how birth order affects intellectual development comes from the confluence model, originated by R. Zajonc (Zajonc & Markus, 1975). This model asserts that the intellect of an environment is a limited resource, and it also considers the average intellect of adults and siblings with whom a child is most likely to interact. A firstborn child will, according to this theory, interact more with her/his parents and other adults than will kids born later, who will interact more with their siblings and others closer to their age. Firstborn kids have greater access to their parents than later-born kids, who will throughout their lives need to compete for attention with their other siblings. Because of this, younger siblings more often hear the simpler language of young children, instead of the more complex language of adults. As a result, later born children tend to live in a more “diluted” intellectual environment than older born children (Zajonc & Bargh, 1980a).

An additional important advantage that older siblings have is that they are often tutors and surrogate caregivers for younger children, especially in large families. Young children will
ask their older siblings questions regarding how to deal with various tasks, such as homework problems. This phenomenon has been shown to enhance academic achievement and intellectual development for not only the learner, but for the tutor as well (Bargh & Schul, 1980). The youngest child at any given time does not benefit from being a tutor to other siblings, and is therefore at an intellectual disadvantage relative to older siblings (Zajonc & Mullally, 1997; Zajonc & Sulloway, 2007).

The confluence model also suggests that the spacing of the births of children can be important. Younger children benefit more from having much older siblings than older siblings close to them in age. This is because the sibling who is 5 years older, for example, has had more time to undergo intellectual development than a sibling closer in age, meaning those older siblings can be a greater help to their younger siblings. Additionally, the vocabulary of a much older sibling will be more advanced than a sibling close in age, potentially fostering greater cognitive development in the younger sibling (Zajonc, 2001). A child born into an intellectual environment of much older siblings, for example, might even be born into a better environment than an only child in some cases, because much older siblings are well-developed and are more able to help their new sibling thrive (Zajonc & Bargh, 1980b).

Research has also demonstrated that the birth order phenomenon is not simply a function of family size. Across various family sizes, math and verbal scores have been found to be significantly higher for first born children. Even after controlling for family size, birth order is still significantly related to cognitive development (Black, Devereux, & Salvanes, 2005). This effect persists into adulthood when considering earnings, such that a firstborn’s educational advantage leads her/him to have higher earnings than those who were later born (Kantarevic & Mechoulan, 2006). First born children generally have greater parental inputs, as mentioned
earlier, due to having sole possession of parental resources for at least some portion of their lives. According to the theory, firstborn siblings may also accelerate their own development by teaching younger siblings (Zajonc & Markus, 1975; Heiland, 2009).

Further, the birth order phenomenon does seem to be at least partially a function of income. Travis and Kohli (1995) showed that birth order is negatively related to educational attainment mainly for middle class families. By contrast, birth order was not significantly related to educational attainment in upper and lower class families in their sample. This could be because middle class families have enough resources for intellectual development (unlike poor families) but still have to be concerned about resource distribution (unlike wealthy families). This finding suggests that resource allocation and limited resources are another potential reason why birth order affects cognitive development.

Birth order may also be related to race. Black families in the United States, on average, tend to have 20% more of their own children under the age of 18 in their household than White families (N = 78.8 million households; United States Census Bureau, 2010). This suggests that Black families may be particularly vulnerable to the effects of birth order on cognitive development, because Black children tend, on average, to have a greater number of older siblings. Thus, in this section, we have contended that Black-White cognitive test score gaps will be partly attributable to differences in birth order.

Hypothesis 1: Birth order will partially account for the Black-White race gap in cognitive ability test scores.

Maternal cognitive test scores. The cognitive ability of one’s mother has been associated with the cognitive ability of children (Bennett, Bendersky, & Lewis, 2008). There are many possible reasons for this. One factor is the heritability of cognitive ability, where heritability is
defined as the fraction of observed phenotypic variance caused by differences in heredity (Lush, 1940). Two types of studies are generally used to quantify the heritability of cognitive ability: adoption studies, which compare unrelated individuals in ostensibly the same environment; and twin studies, which compare monozygotic and dizygotic twins who are raised either together or separately. Studies like these attempt to estimate the proportion of variance in cognitive ability, or any other individual difference variable, that can be attributed to genetic effects (labeled $h^2$, the heritability coefficient), the common/shared environment of twins and/or adopted siblings (labeled $c^2$), and the non-shared environment of twins and/or adopted siblings (labeled $e^2$). For example, by using an estimation model that compares the correlation of cognitive test scores of monozygotic twins reared apart (who share identical genes; $r = h^2$) against the correlation of cognitive test scores of monozygotic twins reared together, (who share both identical genes and a common/shared environment; $r = c^2 + h^2$), it is possible to examine how much the common/shared environment influences the correlation of cognitive test scores between monozygotic twins.

There is some controversy as to how to estimate heritability in general, with many studies using twin studies (Plomin, Pedersen, Lichtenstein, & McClearn, 1994) and other more recent studies using DNA genotyping evidence from essentially unrelated individuals (defined as less than 2.5% shared genes variants; Trzaskowski et al., 2013). There is further controversy regarding how heritable cognitive ability is in particular (Nisbett et al., 2012; Devlin, Daniels, & Roeder, 1997), partly due to difficulties in detecting specific genetic variants associated with cognitive ability (Chabris et al., 2012). With that said, many studies have estimated the percent of variance accounted for by genes, or heritability, of cognitive ability to be between .4 and .8, with some estimates as high as .9 and an average of around .5 (for review, see Nisbett et al.,
2012; e.g., Boomsma, Busjahn, & Peltonen, 2002; Plomin et al., 1994). However, these heritability data are often based on twin studies and therefore would not necessarily imply what an expected mother-child correlation would be. Indeed, Bouchard and McGue (1981) found a wide range of correlations between the cognitive ability of mothers and the cognitive ability of their offspring, ranging from less than .1 to around .8, with an average of around .41.

Interestingly, the heritability of cognitive ability is not entirely stable over the lifespan. The heritability of cognitive ability has been found to increase with the age of the child and is especially high in adults (Bouchard, 2004; Plomin et al., 1994), possibly due to children choosing environments correlated with their genetic propensities, also called genotype-environment correlation (Trzaskowski, Yang, Visscher, & Plomin, 2013). Heritability itself, however, may be related to environmental factors. For instance, Turkheimer, Haley, Waldron, D’Onofrio, and Gottesman, (2003) showed that the heritability of cognitive ability increases as socioeconomic status (SES) increases, and the influence of common/shared environment on cognitive ability tends to decrease as SES increases. This suggests that the mechanisms which affect cognitive development in poor environments and rich environments might not be the same.

Some scholars have posited that twin studies thus overestimate the extent to which genetics influence cognitive development, because twin study samples usually have higher SES than the general population (Nisbett et al., 2012), making them a non-representative sample. Additionally, the education level of parents may affect the heritability of cognitive ability. Specifically, the heritability of verbal IQ in twins of highly educated families was found to be very high ($h^2 = .72$) with little or no contribution of the common/shared environment to cognitive ability ($c^2 = .00$). In contrast, heritability in less-educated families was much lower, with genetic variance and shared/common environment contributing approximately equally to verbal IQ scores ($h^2 = .26$, $c^2$...
Thus, the heritability of cognitive ability might depend upon child’s age, SES, and parents’ education.

Heritability is also not the sole explanation for why maternal cognitive ability and child cognitive ability are related to each other. Children of smart, well-educated mothers tend to learn longer, more complex, and a larger number of words at a young age; likely due to a greater variety and complexity of words used by their mothers (Dollaghan et al., 1999). Schady (2011) showed that mothers with higher vocabulary levels, as measured by the Peabody Picture Vocabulary Test, had children with more advanced vocabulary and higher scores on tests of memory and visual integration. This study also found that the relationship between maternal vocabulary and child vocabulary increased as children got older. One possible theoretical explanation for this result is that parents with higher cognitive ability might have a better understanding of the importance of making a child’s environment more stimulating, which can positively affect cognitive development (Bacharach & Baumeister, 1998). Indeed, previous research has shown that mothers with higher cognitive test scores generally possess more self-esteem, academic aptitude, and higher expectations for both themselves and their children (Magnuson, 2007). Altogether, the above rationale suggests that mothers’ cognitive ability/achievement influences the cognitive development of their children.

Maternal cognitive ability is also related to race. Previous research evidence has robustly shown the Black-White cognitive test score gap to be around one standard deviation in magnitude (Roth et al., 2001), and we assert that this gap generalizes to racial differences in cognitive ability among mothers. Consistent with this generalization, previous studies have found a negative relationship between race and maternal cognitive test scores that is similar in magnitude to the $d = 1.0$ gap identified by Roth et al. (2001; e.g., $d = 1.19$, Mandara et al., 2009;
$d = 1.35$, Yeung & Pfeiffer, 2009). As for a theoretical rationale to explain the origins of race differences in maternal cognitive ability, the current paper—as a whole—is about precisely this issue. That is, the current paper attempts to advance a theoretical model for the origins of race differences in achievement test scores. In this section, we have contended that the Black-White gap in cognitive test scores will be partly attributable to maternal cognitive test scores.

**Hypothesis 2:** Maternal cognitive test scores will partially account for the Black-White race gap in cognitive ability test scores.

**Learning Materials.** In order for children to cognitively develop properly, parents must provide a suitable home environment for them where they can learn and expand their understanding of the world around them. Learning materials are a key aspect of the home environment for children (Watson, Kirby, Kelleher, & Bradley, 1996). The presence of learning materials is significantly positively related to cognitive test scores, as well as negatively related to problematic behaviors such as aggression and delinquency (Linver, Brooks-Gunn, & Kohen, 2002). The learning materials subscale of the Home Observation for the Measurement of the Environment (HOME; Caldwell & Bradley, 1984) has been significantly related to vocabulary, math, and reading tests, especially for younger children (Bradley, Corwyn, Burchinal, McAdoo, & Garcia Coll, 2001b).

Previous attempts to quantify and explain the Black-White cognitive test score gap have examined learning materials as an explanatory variable. For example, Fryer and Levitt (2004, 2006) use the number of books present in a child’s home as an explanatory variable. They found that the number of books in a child’s home was related to cognitive development. Specifically, a one standard deviation increase in the number of children’s books increased reading and math scores by .143 and .115 standard deviations, respectively. Additionally, it was found that the
Black-White gap in cognitive test scores, specifically on the Peabody Picture Vocabulary Test (PPVT-R) and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), were reduced significantly by adding learning materials as a covariate in the regression model, even after controlling for income and maternal verbal test scores (Brooks-Gunn et al., 2003).

Previous studies have shown that there are Black-White differences in learning materials in the home ($d = 1.23$, Brooks-Gunn et al., 2003 IHDP; $d = 1.05$, Thompson Jr. et al., 1998; $d = 1.17$, Bradley & Caldwell, 1984). One possible reason for this is that Black families tend to be poorer than White ones. In the United States in particular, nearly 40% of Black individuals make less than $40,000, and the median income of Black individuals is $24,000 less than that of White individuals (United States Census Bureau, 2009). Having lower incomes and less access to developmental resources might prevent some Black people from gaining the learning materials necessary to provide their children a more educationally stimulating home environment (Linver et. al, 2002). It is also possible that a parent’s perception of the norms for how many books, puzzles, and other learning materials a young child needs, stems partly from one’s own childhood experience. If so, then there might occur intergenerational transmission of norms for how many learning materials should be made available. Under such circumstances, even families with greater financial resources might still not provide a lot of learning materials to their children, because they do not believe a large number of such materials is necessary.

_Hypothesis 3:_ Learning materials will partially account for the Black-White race gap in cognitive ability test scores.

_Parenting Factors: Maternal Sensitivity, Maternal Warmth and Acceptance, and Physical Environment._ The extent to which parents provide a warm and caring environment, and not just learning materials, is also important for cognitive development. Thus, we differentiate
learning materials from aspects of a child’s environment related to mother’s caring and providing both a secure and welcoming environment. We believe these factors are all related to the provision of a safe and caring home for children, which fosters cognitive development.

Maternal sensitivity is generally defined as “a mother’s ability to perceive and interpret accurately her infant’s signals and communications and then respond appropriately” (Ainsworth, Blehar, Waters, & Wall, 1978, as cited in Shin, Park, Ryu, & Seomun, 2008). Early maternal sensitivity has been found to significantly predict cognitive development (Page, Wilhelm, Gamble, & Card, 2010; Lemelin, Tarabulsy, & Provost, 2006). For example, Estrada, Arsenio, Hess, and Holloway (1987) found that maternal sensitivity correlated with cognitive development from age 4 through age 12.

Maternal sensitivity is especially important for cognitive development in very young children. Stams, Juffer, and van Ijzendoorn (2002) showed that the correlation between 12 month maternal sensitivity and 7 year old cognitive development was higher than the correlation between maternal sensitivity at 7 years old and 7 year old cognitive development. Bornstein and Tamis-Lemonda (1997) found that maternal responsiveness at 5 months significantly predicted attention span and symbolic play at 13 months. Additionally, Feldman, Eidelman, and Rotenberg (2004) showed that cognitive development at 1 year old in a sample of triplets, twins, and singletons was statistically significantly related to multiple-birth status, as well as to maternal sensitivity throughout the children’s first year of life. These authors theorized that a triple birth creates a high-stress environment that prevents parents from providing exclusive parenting to each child. This process results in lower maternal sensitivity, which then interferes with infants’ cognitive growth (Feldman et al., 2004).
Pungello, Iruka, Dotterer, Mills-Koonce, and Reznick (2009) found that parents with low maternal sensitivity were often more depressed, which negatively affected the extent to which children acquired language. One potential explanation for this is that depressed parents do not speak to their children as often, decreasing their children’s expressive language and school readiness (Pungello et al., 2009). All of the above research suggests that maternal sensitivity is not only important for cognitive development in early childhood, but that its positive effect will have lasting effects that are maintained throughout a child’s development.

Similarly, the extent to which parents act warmly around their kids and do not punish them harshly for mistakes is related to cognitive test scores (Brooks-Gunn et al., 2003). Parents who tend to be more accepting and who avoid harsh punishments for children’s mistakes have more cognitively advanced children, as measured by the Mental Development Index at 24 months, and the Stanford-Binet IQ test at 36 months (MDI; Bradley et. al., 1989). This effect of maternal acceptance and warmth is significantly related to math and reading scores, particularly among Black participants and poor White participants (Bradley et al., 2001b).

Finally, living in a home that is not overcrowded, is safe, and is relatively bright is positively related to academic achievement test scores (Bradley et al., 1988). Physical environment scores are also statistically significantly related to scores on Acceptance scales \( r = .52 \) from home observation measures (Bradley et al., 1992). This suggests that parents who provide safe and healthy physical environments for children are also generally warm and accepting of their children, and are also involved in helping their child develop cognitively (Bradley et al., 1992).

Further, maternal sensitivity, maternal warmth and acceptance, and physical environment may all be related to race. Previous research has reported sizeable Black-White gaps in maternal
sensitivity (e.g., \(d = .44\), Huang, Lewin, Mitchell, & Zhang, 2012; \(d = .94\), Dotterer, Iruga, & Pungello, 2012; \(d = .63\), Pungello et al., 2009), maternal warmth and acceptance (e.g., \(d = .49\), Bradley & Caldwell, 1984; \(d = .77\); Brooks-Gunn et al., 2003 NLSY-CS), and physical environment (e.g., \(d = .68\), Bradley & Caldwell, 1984; \(d = .41\), Thompson, Jr. et al., 1998). One possible explanation for this phenomenon is that discrimination and prejudice against Black individuals may contribute to Black mothers’ anxiety and depression, which could reduce the quality of the mother-child relationship (Pungello et al., 2009). Additionally, some scholars have posited that Black parents’ having to cope with discrimination, as well as the fact that they tend to live in more impoverished neighborhoods (on average), may contribute to Black-White differences in parenting practices and home conditions (Bradley & Caldwell, 1984; Bradley, Corwyn, Burchinal, McAdoo, & Garcia Coll, 2001a). Thus, we contend that Black-White gaps in maternal sensitivity, maternal warmth and acceptance, and physical environment can partially explain the Black-White cognitive test score gap.

**Hypothesis 4a:** Maternal sensitivity will partially account for the Black-White race gap in cognitive ability test scores.

**Hypothesis 4b:** Maternal acceptance will partially account for the Black-White race gap in cognitive ability test scores.

**Hypothesis 4c:** Physical environment will partially account for the Black-White race gap in cognitive ability test scores.

**Birthweight.** Babies with low birthweight, both those born prematurely and those not born prematurely, tend to be less healthy and are therefore unable to cognitively develop at the same rate as normal weight children, on average. Evidence for this comes from a recent meta-analysis demonstrating that children of very low birthweight showed significantly reduced
volumes of the total brain, gray matter, white matter, cerebellum, hippocampus, and corpus callosum, all of which are related to lower cognitive test scores (De Kieviet, Zotebier, Van Elburg, Vermeulen, & Oosterlann, 2012). In particular, children of low birthweight suffered from deficits in language, memory, and executive functioning (De Kieviet et al., 2012).

As a result of the problems outlined above, previous studies have shown that lower birthweight was associated with lower scores on cognitive tests (Torche & Echevarría, 2011; Dezoete, MacArthur, & Tuck, 2003). Low-birthweight children were also 3 times more likely to need classroom assistance to achieve appropriate grade level performance, as compared to children of normal birthweight (Gross, Mettelman, Dye, & Slagle, 2001). Even low birthweight children without major neurosensory disorders, such as cerebral palsy, still have significantly lower cognitive test scores than children of normal birthweight (Taylor, Klein, Minich, & Hack, 2000). Interventions to assist parents of low birthweight children, such as helping parents feel more confident and comfortable with their children, periodic in-home visits, and parent group meetings can help reduce some of the negative effects of low birthweight on cognitive development (Rauh, Achenbach, Nurcombe, Howell, & Teti, 1988; Brooks-Gunn, Klebanov, Liaw, & Spiker, 1993).

Previous studies have shown that Black children have significantly lower birthweight than White children (e.g., $d = .48$, Lhila & Long, 2012; $d = .33$, Yeung & Conley, 2008). One potential reason for this is that White mothers are often in higher socioeconomic conditions and physically healthier than Black mothers, and therefore tend to have children of higher birthweight (Lhila & Long, 2012). Thus, we contend that racial gaps in birthweight can partially explain the Black-White cognitive test score gap.
Hypothesis 5: Birthweight will partially account for the Black-White race gap in
cognitive ability test scores.

Income/Socioeconomic Status. Socioeconomic status (SES) has been found to be
correlated with cognitive test scores, as well as college grade point average (GPA; Sackett,
Kuncel, Arneson, Cooper, & Waters, 2009). Family income, often used as an indicator of SES,
has been found to be a significant predictor of IQ even as early as ages 2 and 3 (Klebanov,
Brooks-Gunn, McCarton, & McCormick 1998), as well as at later ages (Fryer & Levitt, 2004;
Brooks-Gunn et al., 2003). This effect is especially strong for children of low birthweight, for
whom low SES exacerbates the negative effects of their low birthweight on cognitive test scores
(Torche & Echevarría, 2011). Additionally, being a low income student in a high income school
was negatively related to science and math achievement test scores in school (Crosnoe, 2009).

Poor families have fewer children’s books and are more likely to live in unstable
neighborhoods, limiting the educational resources that a child has access to (Duncan &
Magnuson, 2005). Compared to non-poor children, poor children were rated as lower in learning
and language stimulation, both of which are necessary for fostering cognitive development
(Watson et al., 1996; Yeung, Linver, & Brooks-Gunn, 2002). In addition to affecting a child’s
learning environment, SES is associated with parental warmth and other positive parenting
practices (Greenman, Bodovski, & Reed, 2011). This may be because low SES parents have less
time to spend with their children, and experience greater stress from a lack of resources, which
also influences parents’ warmth and sensitivity (Mistry, Benner, Biesanz, Clark, & Howes, 2010;
Greenman et al., 2011; McLoyd, 1990). Children of low SES families are also more likely to
experience growth retardation, learning disability, and child abuse, compared to children of high
SES families (Brooks-Gunn & Duncan, 1997). Thus, low SES children are greatly disadvantaged in comparison to their higher SES peers.

Income is related to race. Being poor makes providing a stimulating home environment for children much more difficult, and for Black parents this problem is particularly salient. As mentioned earlier, nearly 40% of Black individuals make less than $40,000 annually, and the median income of Black individuals is $24,000 less than that of White individuals (United States Census Bureau, 2009). African Americans are also much more likely to be enrolled in poorer and more overcrowded schools (Condron, 2009). Thus, African Americans have poorer school environments as well as poorer home environments, which can inhibit learning opportunities for children. Thus, we contend that race gaps in income can partially explain the Black-White cognitive test score gap.

*Hypothesis 6:* Family income will partially account for the Black-White race gap in cognitive ability test scores.
CHAPTER 2

METHODS

Participants and Procedure

Participants were families from the Study of Early Child Care and Youth Development (SECCYD). Participants were recruited in or near 10 hospitals around the United States by the National Institute of Child Health and Development (NICHD) from 1991 until 2009. Cognitive tests were administered at five time points (54 months, first grade, third grade, fifth grade, and 15 years old). For various reasons, some families did not continue to participate throughout the entirety of the study. By phase 4 of the study, when participants were in 7th through 9th grades (from 2005 to 2007), only 1,009 families remained in the study. More information about the recruitment and selection procedures used in this study can be found in publications by NICHD (2005) or online (see http://www.nichd.nih.gov/research/supported/seccyd/overview.cfm). Due to missing data, different variables had different sample sizes [N’s for cognitive tests ranged from 954 (Time 1) to 791 (Time 5)]. To reduce missing data bias and error in the longitudinal model parameters, using a covariance matrix estimated via the Expectation Maximization (EM) Algorithm (Dempster, Laird, & Rubin, 1977; Schafer & Graham, 2002; Newman, 2003).

Measures

Cognitive Ability

Cognitive ability was measured using the math, vocabulary, and reading ability facets of the Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R; Woodcock, 1990; Woodcock & Johnson, 1989). Each of these three ability facets was measured for each child at 5 points in time: 54 months of age, first grade, third grade, fifth grade, and 15 years of age. Math was measured using the Applied Problems subtest, which assesses the use of mathematical skills,
such as adding or subtracting, to solve practical problems. Vocabulary was measured using the Picture-Vocabulary subtest, which assesses children’s ability to identify pictured objects by name. Reading for the first 4 time points (through fifth grade) was measured with the Letter-Word Identification subtest, which assesses children’s ability to identify printed letters and words, as well as their ability to match words to pictures. At 15 years of age, the reading subtest was the Passage Comprehension subtest, which assesses children’s ability to identify missing words in a passage, and their ability to match word phrases to pictures. At each point in time, Math, Vocabulary, and Reading subtest scores were used together to reflect general cognitive ability.

**Explanatory Variables**

Our analyses examine several explanatory variables that we expect to account for the relationship between race and cognitive test scores: birth order, maternal cognitive test scores, learning materials, maternal sensitivity, maternal acceptance, physical environment, birthweight, and income.

**Birth Order:** Birth order data were collected during the researchers’ first visit to the family home, which took place when the child was 1 month old. A higher number indicates that the child was born later (1 = firstborn, 2 = secondborn, etc.). This variable ranged from 1 to 7.

**Maternal Cognitive Test Scores:** Maternal cognitive test scores were measured using the Peabody Picture Vocabulary Test-Revised (PPVT-R). This test was administered when the child was 36 months old. The various split-half reliabilities for this measure (from different splits) ranged between .80 and .83.

**Learning Materials:** Information about stimulation in, and quality of, the home environment was collected using the Home Observation for Measurement of the Environment
(HOME; Caldwell & Bradley, 1984). This measure was administered to mothers as a semi-structured interview when the child was 54 months of age. Questions in this interview focused on the types of family experiences, both in and out of the home, that have been theorized to foster social and cognitive development (Bradley et al., 1989). The Learning Materials subscale, an 11 item measure, assesses the extent to which the child has access to learning materials (e.g., “Child has toys which teach color, size, and shapes,” “Child has three or more puzzles,” “Child has at least 10 children’s books”). The internal consistency reliability was $\alpha = .57$ for the Learning Materials subscale.

**Maternal Sensitivity:** Maternal sensitivity ratings were obtained via a videotaped interaction between a mother [or in rare cases (less than 5%), another caregiver such as a father or grandparent] and his or her child, in a laboratory. Children were each asked to complete a discussion task and a planning task with the help of their mother, and maternal sensitivity was coded by multiple raters using 7 point rating scales. Some examples of such tasks include drawing a tree or house with an Etch-a-Sketch, and discussing areas of disagreement such as chores and homework. Parents were rated on a composite of supportive presence, respect for autonomy of the child, and reflected hostility (reverse coded). Data for maternal sensitivity used in this study were collected at several time points: 54 months, first grade, third grade, fifth grade and 15 years old. Internal-consistency reliabilities for these five time points range from $\alpha = .80$ to .85.

**Maternal Warmth and Acceptance:** The Acceptance subscale of the HOME was administered to mothers using a semi-structured interview at 54 months. This subscale measures how the mother interacts with the child and the extent to which the mother accepts imperfect behavior and avoids punishing the child harshly (e.g., “Parent does not scold or derogate child
more than once,” “Parent neither slaps nor spanks child during visit.” “No more than one instance of physical punishment during last week”). The internal consistency reliability was $\alpha = .52$ for this 4 item subscale.

**Physical Environment:** The 7 item Physical Environment subscale (measured as part of the HOME semi-structured interview at 54 months) assesses the extent to which parents provide a safe and clean home environment (e.g., “Building appears safe and free from hazards,” “Rooms are not overcrowded with furniture,” “House is reasonably clean and minimally cluttered”). The internal consistency reliability was $\alpha = .63$ for the Physical Environment subscale.

**Birthweight:** Child birthweight was reported by the mother. Data were collected by research associates in the hospital on the day the child was born. Birthweight data were reported in 100-gram units.

**SES/Income-to-Needs:** We used family income-to-needs as our indicator of socioeconomic status (SES). This variable was measured at 54 months of age, 1st grade, 3rd grade, 5th grade, and at 15 years of age, and was a computed as a ratio of income-to-needs. This ratio is made up of the total household income divided by the federal index for poverty for a given family size. For example, the poverty line in the year 2012 for a family of 2 parents and 2 children was $23,283 (United States Census Bureau, 2012).

**Analyses**

Mplus Version 7 was used to estimate all hypothesized models (Muthén & Muthén, 2012). A sequence of four a priori models was specified. The first model was a latent growth model (LGM) for changes in cognitive ability ($g$) over time (Model 1a in Table 4). At each time point, $g$ was modeled with three reflective indicators: the math test, the vocabulary test, and the reading test. In order to set the scale for the latent $g$ factor and to achieve model identification,
the loadings of the math test onto the overall \( g \) factor at each point in time were fixed to be 1.0.

To specify the growth model, a cognitive ability intercept factor was created onto which the first-order \( g \) factors from each time point loaded with a fixed loading of 1.0 (see Figure 1).

A cognitive ability slope factor was also created, by fixing the loadings of \( g \) to increase linearly with time. That is, the loadings of the cognitive ability \( g \) factors were fixed to -0.1 (T1: 54 month \( g \)), 0.1 (T2: first grade \( g \)), 0.3 (T3: third grade \( g \)), 0.5 (T4: fifth grade \( g \)), and 1.0 (T5: 15 year/tenth grade \( g \)). Additionally, uniquenesses for each of the indicators of \( g \) (e.g., for math test scores) were allowed to correlate over time [i.e., we freed all autoregressive error covariances, within each indicator (Singer & Willett, 2003)].

Model 1b, like Model 1a, was a latent growth model (LGM) for changes in cognitive ability \( (g) \) over time. However, as recommended by Chan (1998), when implementing the growth model, we first sought to establish measurement equivalence over time. That is, math, vocabulary, and reading loadings onto \( g \) were constrained to be invariant over time. Thus, each time point of vocabulary had the same loading onto \( g \) as each other vocabulary time point, and each reading time point had the same loading onto \( g \) as each other reading time point (see Figure 2).

The third model (Model 2 in Table 4) is an LGM for \( g \), which is identical to Model 1b, but with the addition of parameters to estimate racial differences in both the cognitive ability intercept and slope. Measurement equivalence over time is still imposed, as in Model 1b. Model 2 was used to test whether race significantly predicted cognitive ability intercept and slopes (see Figure 3; Race was coded 0 = White, 1 = Black).

Our final model (Model 3 in Table 4) includes the explanatory variables that we hypothesized would account for the relationship between race and cognitive ability test scores.
Model 3 focuses on race differences in the cognitive ability intercept, because in the current study we did not detect any race differences in the cognitive ability slope (as explained in the Results section). Several explanatory variables were investigated, as depicted in Figure 4. Factors were specified for birth order, maternal cognitive test scores, learning materials, maternal acceptance, physical environment, and birthweight, with each reflected by a single manifest indicator with a fixed loading of 1.0 and uniqueness of zero. For maternal sensitivity and income (which were each measured at 5 points in time), we modeled the time intercept by fixing the loadings of maternal sensitivity and income from each time point to 1.0 (Willett & Sayer, 1994). Although some covariates (maternal sensitivity and income) were measured at five points in time and were specified as time intercepts, other covariates (learning materials, acceptance, and physical environment) used single-time-point measures taken at 54 months of age. We consider the 54-month timing of these covariates to be appropriate for our current purposes; because our objective is to explain Black-White differences in the $g$ intercept, which already existed at 54 months of age (i.e., just before participants entered school).

In order to compare the four a priori models described above, we used several goodness-of-fit indices. These include the Comparative Fit Index (CFI; Bentler, 1990), Non-Normed Fit Index (NNFI; Tucker & Lewis, 1973), and the Root Mean Squared Error of Approximation (RMSEA; Steiger, 1990).
CHAPTER 3

RESULTS

Table 3 presents the correlation matrix among constructs in this study, as well as standardized mean differences (d) by race. Large race gaps in cognitive ability test scores are present at every time point (d ranges from -1.14 to -1.38, p < .05).

Measurement Models for Cognitive Test Scores and Race

Model 1a, shown in Figure 1, is a measurement model for g over time. As shown in Table 4, overall fit for Model 1a is adequate, $\chi^2 (57) = 225.69$, CFI = .98, NNFI = .97, RMSEA = .061. Model 1b, is identical to Model 1a shown in Figure 1, but with measurement equivalence specified over time (Chan, 1998), such that the loadings of each of the three indicators of g (Math, Vocabulary, and Reading tests) are fixed to be equal over time. Fit of this constrained Model 1b (i.e., with measurement equivalence in the measurement of g over time) is also deemed adequate, $\chi^2 (65) = 313.29$, CFI = .98, NNFI = .96, RMSEA = .069. Because the change in CFI between Model 1a and Model 1b was less than .01 (Cheung & Rensvold, 2002), we interpret these results to suggest adequate measurement equivalence over time, enabling subsequent longitudinal modeling.

For Model 2, we added race to the measurement model (i.e., Model 1b) and allowed race to relate to the cognitive ability intercept and slope (see Figure 2). Adding race to the model results in a model with the following fit indices: $\chi^2 (78) = 390.94$, CFI = .97, NNFI = .96, RMSEA = .071; which we judged to be adequate fit. Model 2 allows us to examine the baseline relationship between race and cognitive ability test scores. The race-cognitive ability test intercept path coefficient was large and statistically significant (standardized $\gamma = -0.416$, p < .05).
In contrast, the path coefficient between race and the cognitive ability test time *slope* was not statistically significant (standardized $\gamma = -0.05$, $p > .05$, n.s.).

**Using Explanatory Variables to Account for the Cognitive Test Score Gap**

Model 3, shown in Figure 4, expands Model 2 by adding the following explanatory variables: birth order, maternal cognitive test scores, learning materials, maternal sensitivity, maternal acceptance, physical environment, income, and birthweight. In Model 3 we only include relationships between the explanatory variables and the cognitive ability test *intercept*. The cognitive ability test *slope*, in contrast, was not statistically significantly related to race, and therefore no explanatory variables were needed (because there is no race gap in the cognitive test slope).

Model 3 also displays adequate fit to the data, $\chi^2 (408) = 1434.61$, CFI = .94, NNFI = .93, RMSEA = .056. Table 5 shows the race results from Models 2 and 3 (note again that race was coded as White = 0, Black = 1; therefore a positive relationship between a variable and race means the variable in question has a higher mean for Black participants than for White participants). In Model 3, the effect of race on the cognitive ability intercept was no longer statistically significant (standardized $\gamma = -0.06$, $p > .05$, n.s.) after all explanatory variables were added. Thus, the set of explanatory variables in Model 3 has fully explained the relationship between race and cognitive test scores in the current sample.

In addition to assessing the fit of the sequence of models described above (Models 1a, 1b, 2, and 3), we also attempted to quantify how much of the race-cognitive test score gap was accounted for by each explanatory variable in Model 3. That is, we partitioned the race-cognitive test score gap into components that were attributable to each explanatory variable. In Model 3, race was specified as a predictor of each of the explanatory variables, as well as a predictor of
the cognitive ability intercept. Each explanatory variable was then allowed to predict the
cognitive ability intercept. All explanatory variables were also allowed to correlate with each
other.

In the full model (Model 3), the following explanatory variables were related to race:
birth order (standardized $\gamma = .129, p < .05$; Blacks have higher [later] birth order), maternal
cognitive test scores (standardized $\gamma = -.435, p < .05$), learning materials (standardized $\gamma = -.348, p < .05$), maternal sensitivity (standardized $\gamma = -.442, p < .05$), maternal acceptance (standardized $\gamma = -.267, p < .05$), physical environment (standardized $\gamma = -.307, p < .05$), birthweight
(standardized $\gamma = -.186, p < .05$), and income (standardized $\gamma = -.282, p < .05$). Also in the full
model (Model 3), the following variables are related to the cognitive ability intercept: birth order
(standardized $\beta = -.174, p < .05$), maternal cognitive test scores (standardized $\beta = .330, p < .05$),
learning materials (standardized $\beta = .099, p < .05$), maternal sensitivity (standardized $\beta = .253, p
< .05$), and physical environment (standardized $\beta = .083, p < .05$). In contrast, the following
explanatory variables were not related to the cognitive ability intercept in the full model:
maternal acceptance (standardized $\beta = .033, n.s.$), birthweight (standardized $\beta = .048, n.s.$), and
income ($\beta = .035, n.s.$).

We also tested the indirect effect of each explanatory variable as an explanation for the
relationship between race and cognitive ability. That is, we attempt to estimate the extent to
which each explanatory variable accounts for the race-test score gap. In the sections that follow,
we first present the indirect effects for each explanatory variable tested independently, and then
we present the indirect effects for each explanatory variable from the full model (Model 3), in
which all explanatory variables were tested simultaneously. The full model allows us to partition
the race-test score total effect (total race gap) into portions of the gap that were accounted for by
each explanatory variable. That is, the total effect of race on the cognitive ability intercept (Black-White race gap) can be partitioned into several indirect effects that each operate through the explanatory variables, plus the leftover direct effect from race to the cognitive ability intercept after the explanatory variables have all been accounted for. We thus report the indirect effect size, as well as the percent of the total race effect on cognitive test scores, which is calculated by dividing each indirect effect (e.g., race to maternal sensitivity × maternal sensitivity to cognitive test scores) by the total effect (race to cognitive test scores).

When birth order is considered alone (in the absence of other explanatory variables), the indirect effect from race to cognitive test scores through birth order is statistically significant ($\gamma \times \beta = -0.026$, Sobel test $p < 0.05$; $\gamma =$ path from race to birth order, $\beta =$ path from birth order to cognitive ability intercept), and birth order accounts for 6.1% of the race gap in the cognitive test score intercept (see first two columns of Table 6). In contrast, in the full model (Model 3; with all explanatory variables modeled simultaneously), the indirect effect through birth order is $\gamma \times \beta = -0.022$ (Sobel test $p < 0.05$) and birth order uniquely explains 5.3% of the race gap in cognitive test scores (see last two columns of Table 6).

When maternal cognitive test scores are considered alone, the indirect effect from race to cognitive test scores through maternal cognitive test scores is statistically significant ($\gamma \times \beta = -0.238$, Sobel test $p < 0.05$), and maternal cognitive test scores account for 54.6% of the race gap in cognitive test scores. In contrast, in the full model (Model 3), the indirect effect through maternal cognitive test scores is $\gamma \times \beta = -0.143$ (Sobel test $p < 0.05$), and maternal IQ uniquely explains 33.7% of the race gap in cognitive test scores.

When learning materials are considered alone, the indirect effect from race to cognitive test scores through learning materials is statistically significant ($\gamma \times \beta = -0.126$, Sobel test $p <$
and learning materials account for 29.3% of the race gap in cognitive test scores. In contrast, in the full model (Model 3), the indirect effect through learning materials is \( \gamma \times \beta = -0.034 \) (Sobel test \( p < .05 \)), and learning materials uniquely explain 8.0% of the race gap in cognitive test scores.

When maternal sensitivity is considered alone, the indirect effect from race to cognitive test scores through maternal sensitivity is statistically significant (\( \gamma \times \beta = -0.222 \), Sobel test \( p < .05 \)), and maternal sensitivity accounts for 51.5% of the race gap in cognitive test scores. In contrast, in the full model (Model 3), the indirect effect through maternal sensitivity is \( \gamma \times \beta = -0.111 \) (Sobel test \( p < .05 \)), and maternal sensitivity uniquely explains 26.3% of the race gap in cognitive test scores.

When maternal acceptance is considered alone, the indirect effect from race to cognitive test scores through maternal acceptance is statistically significant (\( \gamma \times \beta = -0.073 \), Sobel test \( p < .05 \)), and maternal acceptance accounts for 17.2% of the race gap in cognitive test scores. However, in the full model (Model 3), the indirect effect through maternal acceptance is not statistically significant (\( \gamma \times \beta = -0.008 \), n.s.) and maternal acceptance uniquely accounts for only 1.9% of the race gap in cognitive test scores.

When physical environment is considered alone, the indirect effect from race to cognitive test scores through physical environment is statistically significant (\( \gamma \times \beta = -0.089 \), Sobel test \( p < .05 \)), and physical environment accounts for 20.7% of the race gap in cognitive test scores. In contrast, in the full model (Model 3), the indirect effect through physical environment is \( \gamma \times \beta = -0.026 \) (Sobel test \( p < .05 \)), and physical environment uniquely accounts for 6.1% of the race gap in cognitive test scores.
When birthweight is considered alone, the indirect effect from race to cognitive test scores through birthweight is not statistically significant \((\gamma \times \beta = -0.013, \text{Sobel test } p > 0.05)\), and birthweight accounts for 3.1% of the race gap in cognitive test scores. In the full model, the indirect effect through birthweight is \(\gamma \times \beta = -0.009\) (n.s.), and birthweight uniquely accounts for 2.2% of the race gap in cognitive test scores.

When income is considered alone, the indirect effect from race to cognitive test scores through income is statistically significant \((\gamma \times \beta = -0.096, \text{Sobel test } p < 0.05)\), and income accounts for 22.3% of the race gap in cognitive test scores. However, in the full model, the indirect effect through income is not statistically significant \((\gamma \times \beta = -0.010, \text{n.s.})\) and income uniquely accounts for only 2.4% of the race gap in cognitive test scores.

In summary, Table 6 (column 3) shows that Hypotheses 1 (birth order), 2 (maternal cognitive test scores), 3 (learning materials), 4a (maternal sensitivity) and 4c (physical environment) were supported. In contrast, Hypotheses 4b (maternal acceptance), 5 (birthweight) and 6 (income) were not supported. Altogether, the set of explanatory variables in Model 3 accounts for 85.8% of the total race gap in cognitive test scores.

**Supplementary Analyses**

Elaborating the pathway from maternal cognitive ability scores to child cognitive ability scores. The goal of the current study is to use a parsimonious set of covariates to explain Black-White race gaps in cognitive test scores, as they develop longitudinally across childhood and adolescence in the general population. As such, we have focused our attentions on proposing reasons why each explanatory covariate should relate to both cognitive development and to race. What we have not done, however, is to build a sophisticated theory of the causal relationships among the various covariates themselves. In this regard, we now take the opportunity to model
one theoretically-important set of relationships—i.e., the possibility that maternal cognitive test scores give rise to child cognitive test scores by way of several other covariates. These additional covariates, which we believe might help explain the intergenerational transmission of cognitive test scores, include: birth order (Rodgers, Cleveland, van den Oord, & Rowe, 2000; this article uses family size, which is not the same thing as birth order, but large families do have more children with higher birth orders by definition) learning materials (Bennett, Bendersky, & Lewis, 2008), maternal sensitivity (Poe, Burchinal, & Roberts, 2004), warmth and acceptance (Bradley et al., 1992; Mandara et al., 2009), safe physical environment (Bradley et al., 1992), birthweight (Garret, Ng’andu, & Ferron, 1994), income (Bacharach & Baumeister, 1998), maternal education (reported by mother when child was 1 month old; Garret, Ng’andu, & Ferron, 1994), and maternal age (reported by mother when child was 1 month old; Bacharach & Baumeister, 1998). We propose that all of the above-listed covariates are higher among mothers with higher cognitive ability test scores, except for birth order (which should be lower among mothers with high cognitive ability scores, because these mothers have fewer children). To examine whether these variables could partially account for the relationship between maternal cognitive test scores and child cognitive ability test scores, we estimated an additional model, depicted in Figure 5. The model fit for this model is $\chi^2 (258) = 843.576$, CFI = .96, NNFI = .94, RMSEA = .054.

As seen in Figure 5, the following variables were significantly predicted by maternal cognitive test scores as predicted: birth order ($\beta = -.110, p < .05$), learning materials ($\beta = .454, p < .05$), maternal sensitivity ($\beta = .590, p < .05$), maternal acceptance ($\beta = .359, p < .05$), physical environment ($\beta = .313, p < .05$), birthweight ($\beta = .157, p < .05$), income ($\beta = .472, p < .05$), maternal education ($\beta = .630, p < .05$), and maternal age ($\beta = .491, p < .05$). The following explanatory variables were in turn related to the cognitive ability intercept: birth order ($\beta = -.194$,}
learning materials ($\beta = .094, p < .05$), maternal sensitivity ($\beta = .262, p < .05$), physical environment ($\beta = .089, p < .05$), and birthweight ($\beta = .059, p < .05$). These explanatory variables together accounted for approximately 49% of the direct effect of maternal cognitive test scores on the cognitive test score intercept, the remaining direct effect from maternal to child cognitive test scores was $\beta = .326 (p < .05)$. This suggests the relationship between maternal cognitive ability scores and child cognitive ability scores can be partially and uniquely accounted for by lower birth order, greater availability of learning materials, higher maternal sensitivity, safer physical environment, and higher birthweight.

**A modified, ‘Four-Channel Model’ of the origins of the race gap in cognitive test scores.** We finally note that—even though the explanatory model in Figure 4 (i.e., Model 3) is already more parsimonious than alternative models that have been offered to explain the Black-White gap in cognitive test scores—it could be made more parsimonious still. That is, not all of the specified explanatory variables are needed to explain the gap. As such, we next offer a modified post hoc model that can even more parsimoniously explain the gap. Whereas post hoc models risk capitalizing on chance and thus need future replication (MacCallum, Roznowski, & Necowitz, 1992), we believe that the modified model in Figure 6 has utility in helping future readers to clearly recall which explanatory variables are necessary (vs. unnecessary) to explain the race gap. That is, given that this topic area is politically controversial, and give that past, non-empirical theoretical models are rife with explanations for cognitive ability that are based upon SES (see Sackett et al., 2009 and Sackett et al., 2012 for a description of this literature), our opinion is that it will be handy for future theorists to understand that the full race gap can be uniquely explained using a small handful of covariates that *do not* include income nor maternal education (i.e., the explanatory model does not need SES—or more precisely, SES is only a
distal indicator of the more proximal and direct explanations shown in Figure 6). The goodness-of-fit for the model depicted in Figure 6 is $\chi^2 (255) = 829.77$, CFI = .96, NNFI = .95, RMSEA = .053, which we deem to be adequate fit.

In Figure 6, we see the full ‘Four-Channel Model’ that explains the race gap in cognitive test scores. We note that this model is based upon past empirical work (see Table 2), but that we further incorporated all of these explanatory constructs into a single, integrated theoretical model. The effects of each construct shown in Figure 6 are unique effects, which each account for the roles of all of the other explanatory variables that are in the model simultaneously. The four channels (or pathways) that can be used to uniquely explain the relationship between race and cognitive test scores are: (a) birth order, (b) maternal cognitive ability scores, (c) learning materials, and (d) parenting factors (maternal sensitivity, acceptance, and physical environment).
CHAPTER 4
DISCUSSION

The purpose of the current paper was to theoretically explain the origins of adverse impact. We did this by modeling Black-White cognitive test score gaps between 54 months and 15 years of age (i.e., across the majority of the life course before individuals enter the workforce), and by attempting to offer an integrated, parsimonious theoretical model to explain this gap. We quantified the size of the gap over time, examined whether the gap grows over time, and also investigated the extent to which our developmental explanatory variables (birth order, maternal cognitive test scores, learning materials, maternal sensitivity, maternal acceptance, physical environment, birthweight, and income) could account for the relationship between race and cognitive test scores. Finally, in a supplementary analysis, we attempted to examine explanatory variables for the relationship between maternal cognitive test scores and child cognitive test scores.

Our results suggest that Black-White gaps in cognitive test scores are large and pervasive, and are already established at the young age of 54 months. This is indicated by the mean differences in cognitive test scores at each time point (i.e., subgroup d’s range from -1.15 to -1.38 across the time points, see Table 3), as well as the race gap in the cognitive ability intercept (intercept d = -1.33). Further, between 54 months and 15 years of age, this gap did not significantly increase over time, as indicated by the lack of relationship between race and the cognitive ability slope from the LGM.

Figure 6 depicts our four-channel model. The four-channel model explanatory variables (birth order, maternal cognitive test scores, learning materials, and maternal sensitivity/acceptance/physical environment) each uniquely explain significant variance in the
relationship between race and cognitive test scores in our sample. Moreover, the relationship between race on cognitive test scores was no longer statistically significant after accounting for just these explanatory variables, suggesting that the race-test score relationship has been fully accounted for in our sample with a small number of covariates.

Finally, birth order, learning materials, maternal sensitivity, physical environment, and birthweight partially accounted for the relationship between maternal cognitive ability scores and child cognitive ability scores. The implications of this paper are that our four-channel model can fully account for the Black-White cognitive test score gaps over the course of a child’s life. This suggests that adverse impact created by cognitive tests may arise as a result of Black-White differences in these important developmental conditions. The current theoretical model thus contributes to theories about the origins of subgroup differences in cognitive test scores, which has been cited as a major theoretical gap in current models of adverse impact (Outtz, 2010).

The results of this paper also suggest new directions for adverse impact research. Namely, researchers should continue to examine the extent to which the different societal and developmental resources that create cognitive test score gaps (Figure 6) might also create gaps in actual job performance. This is an essential question for personnel selection scientists and practitioners, given than race gaps in cognitive ability tests are approximately three times larger than corresponding race gaps in job performance (McKay & McDaniel, 2006; Outtz & Newman, 2010). Such studies have the potential to develop an even fuller picture of which racial inequalities (or inequities) must be addressed in order to reduce adverse impact in hiring and admissions.

To elaborate, some models of test fairness suggest that the key problem of adverse impact is due to elements of cognitive tests that overlap with race but do not overlap with job
performance (Darlington, 1971; Cole, 1973; Newman, Hanges, & Outtz, 2007). Outtz & Newman, (2010) refer to this as performance irrelevant race-related variance in cognitive test scores. If this aspect of cognitive test scores is large, it implies that when cognitive tests are used for hiring, African-Americans would be excluded from jobs for reasons that have nothing to do with job performance. Because the current study does not include any measures of job performance (i.e., the sample was not old enough to be legally employed), we cannot presently address the development of performance irrelevant race-related cognitive test score variance.

Another potential direction for future research is to change cognitive tests themselves so that they retain their high validity while reducing adverse impact. This could involve changing the way test material is presented (Schmitt & Quinn, 2010) as well as exploring the extent to which cognitive test questions may be race-loaded. For example, technical knowledge tests tend to show much larger Black-White differences than do math tests or cognitive speed tests (Alderton, Wolfe, & Larson, 1997; Hough, Ployhart, & Oswald, 2001; Kehoe, 2002; Outtz & Newman, 2010). This may be because the measure of some facets of cognitive ability also unintentionally measure aspects of socially privileged life experience, as well as one’s familiarity with testing styles and situations (Goldstein, Scherbaum, & Yusko, 2010). Thus, one potential way to reduce Black-White cognitive test score gaps is to create a cognitive test that is unfamiliar to all participants while still being a valid measure of cognitive ability. While this might not eliminate adverse impact altogether, such a strategy could eliminate contamination of the cognitive test due to privilege (Goldstein et al., 2010; Yusko & Goldstein, 2008). A revival of research on the construct of intelligence might help solve these and other fundamental questions regarding the use of cognitive tests in hiring and admissions decisions (see Scherbaum, Goldstein, Yusko, Ryan, & Hanges, 2012).
Limitations

This paper has several limitations. One limitation is that the Study of Early Child Care and Youth Development (SECCYD) is not a strictly random probability sample of the United States population. Families were not eligible for the SECCYD if the mother was under 18 years of age, did not speak English, or had a substance abuse or other serious health problem. Additionally, if the child was hospitalized for more than 7 days after birth, had disabilities, had a twin, or if the family was in a neighborhood that was too dangerous or too far from the study site, they were not eligible to participate. The response rate from those who were eligible was around 58% at the final time point (NICHD Early Child Care Research Network, 1999).

Additionally, the current dataset does not allow us to explore potential interesting research questions brought up in previous research, such as the effects of summer learning versus school learning over time (e.g., Alexander, Entwisle, & Olson, 2007; Downey, von Hippel, & Broh, 2004). Additionally, there is no data on the cognitive development of these participants beyond 15 years of age. Future studies should examine Black-White cognitive test score gaps as individuals continue into the workforce, to assess the possibility that work experience might enhance or ameliorate the cognitive gap for individuals in certain occupations.

Finally, we do not have any employment data on these participants and therefore cannot explicitly explore the extent to which Black and White participants differ on their ability to acquire jobs, as well as how they differ on the types of jobs they acquire as a result of gaps in cognitive test scores, as well as the other variables of our model. Employment data would allow for a fuller connection between racial gaps in cognitive development and adverse impact, possibly showing that cognitive test score differences caused by developmental resource differences in childhood lead to substantially different hiring ratios later in life. Future studies
should utilize longitudinal designs to explore the extent to which Black-White cognitive test scores differences in childhood, as well as gaps in the variables present in our four-channel model, predict success at acquiring jobs.

**Conclusion**

We examined the extent to which specific developmental conditions could account for Black-White gaps in cognitive test scores. We found that Black-White gaps were large at every time point from 54 months to 15 years of age, but that the gap did not grow (nor shrink) over time. Finally, we fully explained the relationship between race and cognitive test scores using our four-channel explanatory model, which features birth order, maternal cognitive test scores, learning materials, maternal sensitivity, maternal acceptance, and physical environment as disparate conditions that give rise to the race gap in test scores. This study therefore pinpoints how cognitive test score gaps can arise due to differences in childhood environments of potential job applicants.
REFERENCES


44


**APPENDIX**

Table 1

*Past Studies Attempting to Explain the Black-White Gap in Cognitive Test Scores*

<table>
<thead>
<tr>
<th>Authors</th>
<th>Sample Type</th>
<th>Sample Size</th>
<th>Ability Measures</th>
<th>Explanatory Variables</th>
<th>Race-g relationship</th>
<th>Statistically Significant Explanatory Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brooks-Gunn, Klebanov, Smith, Duncan, Lee (2003)</td>
<td>Study 1: Two Wave, ages 3 and 5, nationally representative from Infant Health and Development Program (IHDP). Study 2: Cross-Sectional, ages 3-4 and 5-6, low birthweight children from National Longitudinal Study of Youth-Child Supplement (NLCS).</td>
<td>IHDP: N = 627 (312 Black, 315 White) NLSCS: N = 2,220 for 3-4 years old, 1,354 for 5-6 years old</td>
<td>IHDP: Peabody Picture Vocabulary Test–Revised (PPVT-R) ages 3 and 5, Stanford-Binet Intelligence Test age 3, Wechsler Preschool and Primary Scale of Intelligence (WPPSI) age 5.</td>
<td>Birthweight, Gender, Family income, Female head of household, Maternal education, Maternal verbal ability, Maternal age, HOME Learning, HOME warmth</td>
<td>IHDP: Standardized regression coefficient drops from an average of -.49* to -.19* when all covariates are included. NLSY-CS: Standardized regression coefficient drops from an average of -.49* to -.30* when all covariates are included. IHDP PPVT-R Age 5, Income, Maternal education, HOME Learning, HOME Warmth</td>
<td></td>
</tr>
<tr>
<td>Fryer &amp; Levitt (2004)</td>
<td>Longitudinal (4 time points, Fall and Spring of Kindergarten, Spring of First grade, subsample for Fall of First Grade), ECLS nationally representative both public and private, full-time and part time schools and kindergartens</td>
<td>N = 13,290 for Math N= 12,601 for Reading</td>
<td>Math and Reading tests developed exclusively for ECLS. Measurement Equivalence over time not assessed. Slope differences not assessed. Separate regression model at each time point.</td>
<td>Models 4 &amp; 9, p. 451: SES (composite of parental education, occupational status, &amp; household income) Number of children’s books Number of children’s books squared Birthweight Mother over 30 at first birth Teenage mother at first birth Gender Child age at Kindergarten Participation in nutrition program (WIC)</td>
<td>Math: Unstandardized regression coefficient for Black-White gap at Fall of Kindergarten reduced from -.638* to -.094* with covariates included. Reading: Unstandardized regression coefficient for Black-White gap reduced from -.401* to -.117* with covariates included. Spring First Grade gap (Math): b = -.250*, (Reading): b = -.071*</td>
<td>SES, Number of children’s books, Number of children’s books squared, Gender (reading only), Child age at Kindergarten, Birthweight, Mother over 30 at first birth, Teenage mother at first birth, Participant in nutrition program (WIC) Every covariate significant at Fall of Kindergarten was significant at Spring of First Grade</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Sample Size</td>
<td>Variables</td>
<td>Outcomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Fryer & Levitt (2006)                    | Longitudinal (4 time points, Fall of Kindergarten, Spring of Kindergarten, Spring of First Grade, Spring of Third Grade) using data from the Early Childhood Longitudinal Study (ECLS) | Total N: 11,201 for Math, 10,540 for Reading | Math and Reading tests developed exclusively for ECLS based on existing instruments. Measurement Equivalence over time not assessed. Separate regression model at each time point. | • SES (same as Fryer & Levitt (2004))  
  • Number of children’s books  
  • Number of children’s books squared times 1000  
  • Birthweight  
  • Mother over 30 at first birth  
  • Teenage mother at first birth  
  • Gender  
  • Child age at Kindergarten  
  • Participation in nutrition program (WIC)  
  • Math: unstandardized regression coefficient (average over 4 time points) reduced from .76* to .24* with covariates included. Reading: unstandardized regression coefficient (average over 4 time points) drops from .53* to .06* with covariates included.  |
  • Mother’s achievement test scores  
  • Family SES (occupational prestige, poverty status, wealth)  
  • Child decision making  
  • Parental monitoring of children  
  • Child house chores  
  • Arguing about rules  
  • School-oriented home  
  • Maternal warmth  
  • Black-White d = .81 (arithmetic reasoning)  
  • d = .62 (word recognition)  
  • d = .75 (reading comprehension) drops to overall standardized regression coefficient β = -.07* (favoring Blacks).  |
| Yeung & Pfeiffer (2009)                  | Two-Wave, 3 groups. First cohort is grade K in 1997, grades 4-6 in 2003, second cohort is grades 1-3 in 1997, grades 7-9 in 2003, and third group is grades 4-7 in 1997, grades 10-12 in 2003. All are from Panel Study of | N = 1794 (856 Black and 938 White) between the three cohorts | Woodcock Johnson Revised  
  • Applied Problems subtest  
  • Letter Word Identification subtest (Tests are age standardized). Passage Comprehension test for mothers Subtests analyzed separately.  | • Gender  
  • Paternal grandparent education  
  • Maternal grandparent education  
  • Mother received federal aid when child was born  
  • Teenage mother  
  • Low birthweight  
  • Birth order  
  • Cohort 1, Preschool, 1997 Math: Gap drops from unstandardized regression coefficient of -.78* to -.24. Reading: gap drops from -.43* to .02.  
  • Cohort 1, Grades 4-6 2003 Math: Gap drops from -.98* to -.43*.  |
|                                        |                                                                        |             |                                                                           |                                                                           |
| Income Dynamics (PSID) | Measurement equivalence over time not assessed. Slope differences not assessed | Parental education | Parental occupational prestige | Income from birth to age 5 | Average family wealth | Number of children | Family structure | Urbanicity | Parental expectations | Cognitive Stimulation | Emotional support at home | Weekly TV time | Mother’s verbal test score | Reading: gap drops from -.67* to .02. Cohort 2, Grades 1-3, 1997 | Math: Gap drops from -.67* to -.10. Reading: Gap drops from .84* to -.10. Cohort 2, Grades 7-9, 2003 | Math: Gap drops from -1.0* to -.47*. Reading: gap drops from -.77* to -.22. Cohort 3: Grades 4-7, 1997 | Math: Gap drops from -.78* to -.58*. Reading: Gap drops from -.74* to -.40*. Burchinal McCartney Steinberg Crosnoe Friedman McLoyd Pianta and NICHD Early Child Care Research Network (2011) Longitudinal (4 time points: 54 months, first grade, third grade, fifth grade) Low Income Sample Only (2.25 x poverty line and below); Dropped over 400 participants who were above poverty threshold. N = 314 Woodcock-Johnson Revised (WJ-R) • Applied Problems at 54 months and 1st grade • Letter-Word ID at 54 months and First Grade • Broad Reading at Third and Fifth grades • Broad Math at Third and Fifth grades Reading and Math analyzed separately. Reading operationalized differently at T1 and T2 versus T3 and T4 Math operationalized differently at T1 and T2 versus T3 and T4 Tested intercept and slope Reading, Intercept: Unstandardized regression coefficient drops from -12.53* to -3.80. Reading, slope: Unstandardized regression coefficient increases from -.40 to -.66. Math, Intercept: Unstandardized regression coefficient drops from -.97* to .16. Math, slope: Unstandardized regression coefficient drops from .97* to .16. Black-White intercept differences are significant for math and reading in favor of Whites. Black-White slope differences are significant only for Intercept • Parenting Quality composite (+) • Whether child was firstborn (+) (reading only). • Neighborhood disadvantage (-) (math only) • Child-teacher ratio (-) (math, Black only) • Gender (math only) Slope • Two-parent household (+) (reading only) • Classroom quality (+; math, Black only) • Gender (math only)
| Current Paper | Longitudinal (5 time points, 54 months, first grade, third grade, fifth grade, 15 years), full income range, families recruited from United States hospitals | N = 791 | WJ-R  
Applied Problems (math) at all time points  
Letter-Word ID (reading) at 54 months, 1st grade, 3rd grade and 5th grade.  
Passage Comprehension (reading) at 15 years.  
Picture Vocabulary at all time points | Estimated measurement model for Cognitive Ability.  
Assessed measurement equivalence over time.  
Tested cognitive ability intercept and slope differences over time | Overall g gap drops from standardized beta of -.42* to -.06 with covariates included. | Estimated measurement model for Cognitive Ability.  
Assessed measurement equivalence over time.  
Tested cognitive ability intercept and slope differences over time |
|---|---|---|---|---|---|
| differences over time. | student body receiving free or reduced price lunch and non-White proportion of student body.  
• Classroom quality (observer ratings)  
Regression models include interaction terms of every covariate with age.  
Only statistically significant regression coefficients are reported. | math in favor of Blacks. | Birth order  
Maternal cognitive test scores  
Learning materials  
Maternal sensitivity  
Maternal warmth and acceptance  
Physical environment  
Birthweight  
Income | Birth order  
Maternal cognitive test scores  
Learning materials  
Maternal sensitivity  
Physical environment |
Table 2

*Replicated covariates that partly explained the Black-White gap in cognitive test scores (statistically significant across multiple samples)*

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Number of samples where supported</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother’s Cognitive test scores</td>
<td>4</td>
<td>Brooks-Gunn et al. (2003, IHDP), Mandara et al. (2009), Yeung &amp; Pfeiffer (2009) Cohort 1 and 2</td>
</tr>
<tr>
<td>Maternal Sensitivity/Home Warmth, Maternal Acceptance, Physical Environment</td>
<td>3</td>
<td>Brooks-Gunn et al. (2003, IHDP and NLSY-CS), Burchinal et al. (2011)</td>
</tr>
<tr>
<td>• Income</td>
<td>3</td>
<td>• Brooks-Gunn et al. (2003, IHDP and NLSY-CS), Burchinal et al. (2011)</td>
</tr>
<tr>
<td>• Maternal Education</td>
<td>3</td>
<td>• Brooks-Gunn et al. (2003, IHDP and NLSY-CS), Burchinal et al. (2011)</td>
</tr>
</tbody>
</table>
Table 3

*Correlation Matrix among Latent Variables*

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. GT1</td>
<td>32.25</td>
<td>12.50</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. GT2</td>
<td>32.23</td>
<td>11.31</td>
<td>.87*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. GT3</td>
<td>32.20</td>
<td>11.23</td>
<td>.86*</td>
<td>.95*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. GT4</td>
<td>32.18</td>
<td>11.11</td>
<td>.86*</td>
<td>.96*</td>
<td>.97*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. GT5</td>
<td>32.12</td>
<td>12.00</td>
<td>.78*</td>
<td>.87*</td>
<td>.90*</td>
<td>.93*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. g intercept</td>
<td>32.24</td>
<td>11.07</td>
<td>.89*</td>
<td>.98*</td>
<td>.97*</td>
<td>.98*</td>
<td>.88*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. g slope</td>
<td>-0.12</td>
<td>4.33</td>
<td>-.14*</td>
<td>-.07</td>
<td>.00</td>
<td>.08*</td>
<td>.25*</td>
<td>-.11*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Birth Order</td>
<td>1.82</td>
<td>0.94</td>
<td>-.23*</td>
<td>-.25*</td>
<td>-.25*</td>
<td>-.26*</td>
<td>-.26*</td>
<td>-.26*</td>
<td>.03</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Maternal Test Scores</td>
<td>99.02</td>
<td>18.14</td>
<td>.54*</td>
<td>.61*</td>
<td>.62*</td>
<td>.63*</td>
<td>.60*</td>
<td>.62*</td>
<td>.10*</td>
<td>-.11*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Learning Materials</td>
<td>9.39</td>
<td>1.51</td>
<td>.42*</td>
<td>.46*</td>
<td>.47*</td>
<td>.47*</td>
<td>.44*</td>
<td>.47*</td>
<td>.03</td>
<td>-.11*</td>
<td>.45*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Maternal Sensitivity</td>
<td>-0.34</td>
<td>1.85</td>
<td>.53*</td>
<td>.59*</td>
<td>.59*</td>
<td>.60*</td>
<td>.56*</td>
<td>.60*</td>
<td>.01</td>
<td>-.04</td>
<td>.59*</td>
<td>.48*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Maternal Acceptance</td>
<td>4.39</td>
<td>0.80</td>
<td>.34*</td>
<td>.37*</td>
<td>.37*</td>
<td>.37*</td>
<td>.34*</td>
<td>.38*</td>
<td>-.04</td>
<td>-.04</td>
<td>.36*</td>
<td>.37*</td>
<td>.47*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Physical Environment</td>
<td>6.31</td>
<td>1.11</td>
<td>.37*</td>
<td>.40*</td>
<td>.40*</td>
<td>.40*</td>
<td>.35*</td>
<td>.41*</td>
<td>-.07</td>
<td>-.13*</td>
<td>.31*</td>
<td>.37*</td>
<td>.40*</td>
<td>.35*</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Birthweight</td>
<td>34.90</td>
<td>5.05</td>
<td>.14*</td>
<td>.15*</td>
<td>.15*</td>
<td>.15*</td>
<td>.14*</td>
<td>.16*</td>
<td>-.01</td>
<td>.04</td>
<td>.16*</td>
<td>.07</td>
<td>.11*</td>
<td>.07</td>
<td>.10*</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>15. Income</td>
<td>-0.37</td>
<td>3.17</td>
<td>.41*</td>
<td>.44*</td>
<td>.44*</td>
<td>.44*</td>
<td>.39*</td>
<td>.46*</td>
<td>-.10*</td>
<td>-.16*</td>
<td>.47*</td>
<td>.43*</td>
<td>.46*</td>
<td>.30*</td>
<td>.35*</td>
<td>.05</td>
<td>—</td>
</tr>
<tr>
<td>16. Race (r)</td>
<td>0.14</td>
<td>0.35</td>
<td>-.37*</td>
<td>-.42*</td>
<td>-.43*</td>
<td>-.44*</td>
<td>-.42*</td>
<td>-.42*</td>
<td>-.07</td>
<td>.13</td>
<td>-.44*</td>
<td>-.35*</td>
<td>-.44*</td>
<td>-.27*</td>
<td>-.31*</td>
<td>-.19*</td>
<td>-.28*</td>
</tr>
<tr>
<td>(0 = W, 1 = B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Race (d)</td>
<td>-1.15</td>
<td>-1.31</td>
<td>-1.34</td>
<td>-1.38</td>
<td>-1.30</td>
<td>-1.33</td>
<td>-.19</td>
<td>.37*</td>
<td>-1.37*</td>
<td>-1.06*</td>
<td>-1.40*</td>
<td>-.79*</td>
<td>-.92*</td>
<td>-.54*</td>
<td>-.84*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05. Race subgroup d’s are approximate from race r’s, using the formula $d = \frac{r}{\sqrt{1-r^2)(p(1-p))}$ (Lipsey & Wilson, 2001).
Table 4

*Summary of Model Fit*

<table>
<thead>
<tr>
<th>Model</th>
<th>(\chi^2) (df)</th>
<th>CFI</th>
<th>NNFI</th>
<th>RMSEA (90% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a: Cognitive Test LGM (without Measurement Equivalence)</td>
<td>225.69 (57)</td>
<td>.98</td>
<td>.97</td>
<td>.061 (.053, .070)</td>
</tr>
<tr>
<td>1b: Cognitive Test LGM (Measurement Equivalence across Time)</td>
<td>313.29 (65)</td>
<td>.98</td>
<td>.96</td>
<td>.069 (.062, .077)</td>
</tr>
<tr>
<td>2: Race and Cognitive Test LGM, (Measurement Equivalence across Time)</td>
<td>390.94 (78)</td>
<td>.97</td>
<td>.96</td>
<td>.071 (.064, .078)</td>
</tr>
<tr>
<td>3: Race, Explanatory Variables, and Cognitive Test LGM (Measurement Equivalence across Time)</td>
<td>1588.50 (430)</td>
<td>.94</td>
<td>.93</td>
<td>.056 (.053, .060)</td>
</tr>
</tbody>
</table>

*Note.* LGM = Latent Growth Model; CFI = comparative fit index; NNFI = Non-Normed Fit Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean residual; df = degrees of freedom; CI = confidence interval

Table 5

*Structural Equation Modeling Results Involving Covariates*

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>Step 1</th>
<th>Step 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race</td>
<td>-.42*</td>
<td>-.06</td>
</tr>
<tr>
<td>Birth Order</td>
<td>-.17*</td>
<td></td>
</tr>
<tr>
<td>Maternal Cognitive Test Scores</td>
<td>.33*</td>
<td></td>
</tr>
<tr>
<td>Learning Materials</td>
<td>.10*</td>
<td></td>
</tr>
<tr>
<td>Maternal Sensitivity</td>
<td>.25*</td>
<td></td>
</tr>
<tr>
<td>Maternal Acceptance</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Physical Environment</td>
<td>.08*</td>
<td></td>
</tr>
<tr>
<td>Birthweight</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>.04</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* \(N = 791\). DV = Dependent Variable. Coefficients are standardized.

*\(p < .05\)
Table 6

*Indirect Effects and Percent of Total Race Gap Accounted for by Each Explanatory Variable*

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>Indirect Effect Size (each covariate alone)</th>
<th>Percent of Total Gap (each covariate alone)</th>
<th>Indirect Effect Size (full model)</th>
<th>Percent of Total Gap (full model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth Order</td>
<td>-.026*</td>
<td>6.1%</td>
<td>-.022*</td>
<td>5.3%</td>
</tr>
<tr>
<td>Maternal Cognitive Test Scores</td>
<td>-.238*</td>
<td>54.6%</td>
<td>-.143*</td>
<td>33.7%</td>
</tr>
<tr>
<td>Learning Materials</td>
<td>-.126*</td>
<td>29.3%</td>
<td>-.034*</td>
<td>8.0%</td>
</tr>
<tr>
<td>Maternal Sensitivity</td>
<td>-.222*</td>
<td>51.5%</td>
<td>-.111*</td>
<td>26.3%</td>
</tr>
<tr>
<td>Maternal Acceptance</td>
<td>-.073*</td>
<td>17.2%</td>
<td>-.009</td>
<td>1.9%</td>
</tr>
<tr>
<td>Physical Environment</td>
<td>-.089*</td>
<td>20.7%</td>
<td>-.026*</td>
<td>6.1%</td>
</tr>
<tr>
<td>Birth Weight</td>
<td>-.013</td>
<td>3.1%</td>
<td>-.008</td>
<td>2.2%</td>
</tr>
<tr>
<td>Income</td>
<td>-.096*</td>
<td>22.3%</td>
<td>-.010</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

Note: Indirect effect size uses standardized coefficients of path a (race to covariate) and path b (covariate to cognitive test intercept).

*p < .05 based on Sobel test
Figure 1. Cognitive Ability Test LGM (No Measurement Equivalence across Time)

*p < .05

*a Loadings fixed to define latent growth factors.

Note. Coefficients are standardized (unstandardized estimates appear in parentheses).

gT1 = Cognitive Ability at Time X (math loadings fixed at 1.0).
T1 = 54 months, T2 = First Grade, T3 = Third Grade, T4 = Fifth Grade, T5 = 15 years old.
Figure 2. Cognitive Ability Test LGM (Measurement Equivalence Across Time)

*p < .05

* Loadings fixed to define latent growth factors.

Note. Coefficients are standardized (unstandardized estimates appear in parentheses).

gT1 = Cognitive Ability at Time X (math loadings fixed at 1.0). T1 = 54 months, T2 = First Grade, T3 = Third Grade, T4 = Fifth Grade, T5 = 15 years old.
Figure 3. Race and Cognitive Ability Test LGM (Measurement Equivalence Across Time)

*p < .05

*a Loadings fixed to define latent growth factors.

Note. Coefficients are standardized (unstandardized estimates appear in parentheses).

gTx = Cognitive Ability at Time X (math loadings fixed at 1.0).

T1 = 54 months, T2 = First Grade, T3 = Third Grade, T4 = Fifth Grade, T5 = 15 years old.
Figure 4. Cognitive Ability Test LGM with Race and Explanatory Variables (Measurement Equivalence Over Time).

* $p < .05$  a Loadings fixed to define latent growth factors.

Note. gTx = Cognitive Ability at Time X (math loadings fixed at 1.0). T1=54 months, T2 = First Grade, T3 = Third Grade, T4 = Fifth grade, T5 = 15 years old.
Coefficients are standardized (unstandardized estimates appear in parentheses).
Figure 5. Model of the Relationship between Maternal cognitive test scores and Child IQ

Note. *p < .05. Coefficients are standardized.
Figure 6. Four-Channel Explanatory Model

Note. *p < .05. Coefficients are standardized (unstandardized estimates appear in parentheses)