

A META-ANALYSIS OF THE RELATIONSHIP BETWEEN
GENERAL COGNITIVE ABILITY AND FERTILITY

by

Michael David Heeney

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Arts in
Psychology

Charlotte

2018

Approved by:

Dr. Charlie L. Reeve

Dr. George J. Demakis

Dr. Richard D. McAnulty

©2018
Michael David Heeney
ALL RIGHTS RESERVED

ABSTRACT

MICHAEL DAVID HEENEY. A meta-analysis of the relationship between general cognitive ability and fertility. (Under the direction of DR. CHARLIE L. REEVE)

Empirical studies of the relationship between general cognitive ability and fertility span nearly a century. Reported effects generated through previous studies vary widely, and are believed to reflect differences in: 1) sample characteristics (e.g., gender, race, geographic region), and 2) methodological differences (e.g., different assessments of cognitive ability, range restriction, age of sample). The purpose of this study is to conduct a quantitative literature review using meta-analytic procedures. A thorough search identified 17 unique datasets that passed the inclusion criteria. The overall weighted effect was $r = -.11$. Analysis of variation in effect sizes due to methodological factors found significant differences due to age of sample (i.e., whether or not fertility was completed) with larger effects seen when fertility had not yet been completed. Further moderator analyses yielded stronger effects for females compared to males, and for Black and Hispanic groups compared to Whites. For studies conducted within the United States, the effect sizes were stronger among samples collected post Roe vs. Wade than those prior to the legalization of abortion indicating that the dysgenic trend may be stronger when women are afforded that option.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: BACKGROUND	4
2.1. General Cognitive Ability	4
2.2. Dysgenic Fertility	6
2.3. The Flynn Effect and g	9
2.4. Moderators of the g -Fertility Relationship	10
2.4.1. Moderators Due to Methodological Artifacts	10
2.4.1.1. g -score	10
2.4.1.2. Range Restriction	11
2.4.2. Substantive Moderators	12
2.4.2.1. Gender	12
2.4.2.2. Race	13
2.4.2.3. Geographic Region	14
2.4.2.4. Historical Events	14
CHAPTER 3: THE CURRENT STUDY	16
3.1. Hypotheses	16
3.1. Methodology	16
3.2.1. Search Strategy	16
3.2.2. Inclusion/Exclusion Criteria	17

3.2.3. Coding Procedures	18
3.2.4. Meta-Analytic Procedures	18
3.2.5. Analysis for Publication Bias	19
CHAPTER 4: RESULTS	21
CHAPTER 5: DISCUSSION	25
5.1. Limitations	29
REFERENCES	31

LIST OF TABLES

TABLE 1: Study Characteristics	40
TABLE 2: Point Estimate and Moderator Analysis for Methodological Artifacts	42
TABLE 3: Results for Substantive Moderators	43

LIST OF FIGURES

FIGURE 1: Flowchart for Literature Selection Procedure	44
FIGURE 2: Distribution of Effects Scaled to Year When Parent age is 30	45
FIGURE 3: Funnel Plot of Standard Errors by Fisher's Z	46

CHAPTER 1: INTRODUCTION

A large and growing body of empirical literature has documented an inverse relationship between cognitive ability and birth rates in modern society, which has, in turn, resulted in predictions of a declining trend in general intelligence overall (Lynn, 1999; Lynn & Harvey, 2008). This trend has broad implications, impacting occupational, educational, and social outcomes at the individual level, and our ability to collectively tackle societal problems in an increasingly complex world (Neiss et al., 2002; Rindermann, et al., 2009; Rindermann, 2012).

Although most researchers now assume the relation is in fact negative, estimates of this relationship vary widely across studies. This is likely due to a diversity of research designs, measurements, and sample characteristics spanning nearly a century. For example, Lynn & Van Court (2004) found large, statistically significant negative effects (as large as $r = -.53$, $N = 56$) in some African American cohorts for both males and females. Although much less common, other research produced positive effects, one of which came from the same study (Lynn & Van Court, 2004), reporting ($r = .18$, $N=87$) for an African American, male cohort between the ages of twenty-one and thirty-six. One older study (Conrad & Jones, 1932) also reported a relatively large, positive effect ($r = .20$, $N = 45$) for a group of rural, New England males. Other studies included in this paper, most with larger sample sizes, report a range of effects in between, leaving the research community to question, based on the weight of all comparable empirical

evidence, how large the true effects are and what accounts for their differences.

These differences may be due to in part to an evolution in methodological approaches since effects were first reported in the early 20th century. For example, most early studies measured the intelligence of children (as a proxy for parental intelligence) and defined fertility according to the number of siblings (e.g., Burks & Jones, 1935; Burt, 1947; Lentz, 1927). But more recent studies have directly measured the intelligence of each parent, and measured fertility as the number of children born to that person (e.g., Reeve, et al., 2013; Retherford & Sewell, 1989; von Strumm, et al., 2011). Other complexities exist that create variance in estimates of the dysgenic trend as well. For example, some studies measure fertility while people are still in their 20's or 30's (e.g., Reeve, et. al, 2013) whereas others measured fertility later in life to reasonably assure that their reproductive years were complete (Chen, et al., 2013; Wang, et al., 2016; Woodley of Menie, et al., 2015). In addition, early studies were exclusively based in the U.S. and Western Europe, but more recently data from other countries has been reported (Meisenberg, et al., 2006; Chen, et al., 2013; Wang, et al., 2016).

The instruments used to measure intelligence have also varied widely across studies. Differences in measurement validity and reliability make it difficult to compare reported effects across studies. While many used established, psychometrically valid cognitive ability tests such as the Armed Services Vocational Aptitude Battery (Meisenberg, 2010), the Terman Group Intelligence Test (Bajema, 1968) and the Raven's Standard Progressive Matrices test (Meisenberg, et al., 2006), others rely on test batteries designed locally (von Strumm, et al., 2011; Wang, et al., 2016), or composite and proxy variables (Vining, 1982, Vining, 1995). More recently, some researchers (Kanazawa,

2014; Reeve, et al., 2013; Woodley of Menie, et al., 2015) have utilized factor-analytic techniques to extract the systematic variance associated with the underlying construct, general cognitive ability (g), before calculating the effects. Although these more recent efforts reflect a significant methodological improvement, the clear majority of studies that report statistical relationships with fertility rely on raw test scores in various forms. As a result, the amount of variance in observed effects due measurement differences is unknown.

This paper begins with an overview and discussion of the basic concepts and theoretical backdrop relevant to the relationship between intelligence and fertility. Following this, the focus shifts to an empirical analysis that synthesized the results from literature that reports statistical relationships between measures of intelligence and fertility, using systematic approach in study selection and analysis using available meta-analytic tools. The purpose is to identify any commonalities in the findings, and explain any differences by accounting for group-level characteristics, such as race and gender, that are expected to influence the magnitude of reported effects. Moderator analysis was used to achieve this, and to compare studies with known statistical artifacts against those where those artifacts appear to have been adequately addressed. The analysis was conducted in the following sequence: 1) an overall effect size was calculated for the combined studies, 2) an analysis of heterogeneity was conducted to illustrate the extent of dispersion across study-level effects, and 3) moderator analyses were performed to test sample and methodological characteristics. The goal is to enhance our understanding of the relationship between cognitive ability and fertility and thus provide a more stable foundation upon which future research can advance.

CHAPTER 2: BACKGROUND

2.1 General Cognitive Ability

Intelligence is a term that is common to our vocabulary, but one for which a precise meaning defies a consensus view and extends beyond a brief definition (Gottfredson, 1997; Reeve & Bonaccio, 2011). Although intelligence is latent, its properties are indeed measurable, and successful efforts to formulate valid tests have proliferated for more than a century. Models of intelligence have been generally conceptualized as consisting of two broad components: fluid and crystallized intelligence (Cattell, 1943). Fluid intelligence reflects the ability to learn and solve novel problems across a variety of domains. Crystallized intelligence, on the other hand, reflects accumulated knowledge and skills, and is manifest through fluid intelligence coupled with successful efforts to learn over time (Jensen, 1998). Crystallized intelligence is thought to be domain-specific and by its nature is more amenable to training and experience (Reeve & Bonaccio, 2011).

Despite some disagreement about the exact nature of intelligence and its precise structure, empirical research has largely coalesced around a hierarchical structure moving from task-specific skills at the lowest stratum, aggregating up to larger “group” abilities and finally culminating in a single general cognitive ability construct, commonly referred to as “g”. As such, intelligence may be best conceptualized as a collection of closely-related constructs, structured hierarchically in a “tightly-knit nomological network,” with

g reflecting the core ability to learn from and reason with novel information (Reeve & Bonaccio, 2011).

The discovery and measurement of *g* was facilitated through Spearman's development of factor analysis, whereby common variance across sets of test items effectively isolates measures of the underlying construct, while setting aside variance that is unrelated to *g* (Spearman, 1904). The extent to which indicators share this common variance is frequently referred to as "g-saturation", and through continual refinement tests have been developed that measure *g* with increasing precision. This work, and that which followed, provides compelling evidence to support a psychometric structure for intelligence that is hierarchical, with a domain-general "g-factor" at the apex, reflecting, in the words of Spearman (1927), individual differences in "the eduction of relations and correlates."

Empirical research on *g* has expanded across two primary dimensions to form what is commonly referred to as the *g*-nexus (Jensen, 1998). The vertical dimension assumes an evolutionary perspective through a focus on the biological and neurological bases for intelligence. Examples include the study of relationships between IQ and a range of heritable traits including reaction times, evoked potentials of the cerebral cortex, and brain pH (Herrnstein & Murray, 1994; Jensen, 1998). The horizontal dimension examines relationships between *g* and a range of personal, social, educational, occupational and health outcomes. Examples of the horizontal line of inquiry include studies of the relationships between IQ and psychological wellbeing (Lubinski & Benbow, 2000), physical health (Gottfredson, 2004), religiosity (Reeve, 2009; Razmyar & Reeve, 2013), job performance (Gottfredson, 1997; Meisenberg, 2010), and

reproductive behavior (Lynn, 1999; Lynn & Harvey, 2008; Meisenberg, 2010; Peach, et al., 2014; Reeve, et al., 2013; Retherford & Sewell, 1988).

2.2 Dysgenic Fertility

Evolutionary theory has provided a cogent explanation of survival mechanisms through much of human history. Individuals with higher levels of general intelligence evolved to enjoy an enhanced ability to adapt and survive in an increasingly complex world (Gordon, 1997; Gottfredson, 1997; Jensen, 1998; Kanazawa, 2014). It is theorized that this evolutionary advantage prevailed until the early to middle 19th century, when researchers began to observe inverse relationships between intelligence and the number of children born (Herrnstein & Murray, 1994; Lynn & Van Court, 2004; Lynn, 2008; Nyborg, 2012; Woodley & Figueredo, 2013). Over time, this resulted in a rebalancing across the IQ continuum, with the proportion of offspring from lower intelligence individuals increasing with each successive generation. This dysgenic fertility trend has been generally exacerbated by the (younger) age at which lower IQ individuals began having children, thereby shortening the time between generations relative to higher IQ groups (Lynn, 1999).

The timing of this shift has been largely attributed to improved hygiene standards, social services, medical care and general nutrition that occurred largely in tandem with industrialization in the Western world (Lynn, 1996; Nyborg, 2012). Industrialization itself initiated a drift away from the largely hunter-gather and agrarian lifestyle of our ancestors, where large families were an asset for both survival and productivity. As human culture and technology began to fundamentally change our environment, the traditional Darwinian advantage for larger numbers of offspring began to erode. For

example, medicine began to reduce the infant mortality rate (thus, one did not need to have large numbers of children to ensure survival of offspring), and economically speaking, large numbers of offspring began to become a liability rather than a resource.

Although these advances reduced mortality rates and improved general health for industrializing societies at large, they disproportionately benefitted those of lower socioeconomic status for whom these benefits were previously less available which, in turn, facilitated the survival and proliferation of lower IQ groups that may have previously lacked the resources to thrive (Lynn, 1996; Nyborg, 2012). These developments were coined “Internal Relaxation/Reversal of Darwinian Selection” (IRDS) by Nyborg (2012), and are widely cited as the underlying drivers of the dysgenic fertility patterns observed in a large and growing body of empirical literature (Herrnstein & Murray, 1994; Lynn & Van Court, 2004; Lynn, 2008; Nyborg, 2012; Woodley & Figueredo, 2013).

In addition to IRDS theory, differential mating behaviors across the IQ gradient may be explained by r-K Life History theory (aka, differential K theory), which lends insight into dysgenic patterns from both an environmental and an evolutionary perspective. Among different species, those said to be r-selected devote more effort to reproduction and less to rearing offspring, while those more K-selected reverse those priorities. Species are either r or K-selected based upon their surrounding environmental conditions, with r-selected species adapting in relatively unstable and unpredictable environments, and K-selected species surviving under more secure and predictable conditions (Pianka, 1970; Rushton, 1985). Although humans are generally more K-oriented, individuals vary widely across this continuum, with low-K individuals engaging in more impulsive and short term thinking, and high-K individuals maintaining a long-

term orientation conducive to having fewer children and investing heavily in their success as adults (Figueredo et al., 2005; Figueredo et al., 2006). This theory, in the context of human behavior, has been extended to explain the observed relationship between g and fertility in modern society (Rushton, 1985; Rushton, 2004). In short, those with higher intelligence were more likely to disproportionately reduce birth rates through more careful family planning, and enhanced knowledge of and access to contraception (Lynn & Van Court, 2004).

Another theoretical perspective finds its roots in the work of Kanazawa (2004), who argues that “general intelligence evolved as a domain-specific adaptation for the originally limited sphere of evolutionary novelty in the ancestral environment”. Stated differently, intelligence has been shown to influence a range of outcomes applicable to an increasingly complex, “evolutionary novel” world that places demands on our decision making and abstract reasoning that far exceed those of preindustrial times. This is consistent with findings from numerous empirical studies that examine relationships between g and occupational, educational and social outcomes, cited previously, that compose the g -nexus.

Although there remain “evolutionary familiar” elements in modern life, such as copulation, and maintaining social networks, empirical evidence demonstrates that, consistent with theory, g does not well predict these “evolutionarily familiar” behaviors (Kanazawa, 2004). In contrast to copulation, which is evolutionarily familiar, procreation has become an evolutionary novel behavior, thanks to the availability of contraception and other modern changes (explained above). Copulation and reproduction are effectively decoupled. Thus, “the control of reproduction” is now an evolutionary novel behavior

because the individual can decide whether to have children, when to have children, and how many children to have. Often, these decisions compete against other modern-day concerns such as the cost of having children, educational and occupational pursuits. These considerations reflect the “evolutionary novelty” of modern life, and thus reflect behaviors that are highly dependent on g .

This theoretical backdrop has enjoyed a growing body of empirical support. Internal Reversal/Relaxation of Darwinian Selection asserts that improving living conditions set the stage for increasing survival among those with fewer resources. Similarly, Kanazawa’s theory of intelligence as a domain-specific adaptation explains why so many modern behaviors are a function of g (i.e., because most modern behaviors require reasoning ability). Likewise, r-K Selection Theory lends insight into why we see differential reproduction rates (as well as other reproductive behaviors) between high and low intelligence individuals. How we fare in the world, educationally, occupationally, economically, and reproductively are all linked to intelligence, to the degree that they reflect evolutionarily novelty.

2.3 The “Flynn Effect” and g

Evidence suggesting a decline in intelligence may appear at odds with the popularized belief that IQ scores are increasing; the so-called Flynn Effect (Flynn, 1984). While a full review of the “Flynn Effect” is beyond the scope of this paper, recent research suggests the phenomenon is unlikely to reflect a competing hypothesis. Subsequent research has largely explained this apparent paradox by illustrating how secular increases in raw IQ scores are (a) fleeting and inconsistent (Sundet, et al., 2008; Teasdale & Owen, 2008; Williams, 2013), (b) do not reflect changes in g (e.g., Rushton,

1999; Rushton, 2010; te Nijenhuis & van der Flier, 2013; Woodley & Meisenberg, 2013), and (c) reflect a “reeling in” of the lower tail of the distribution through improved environmental conditions, such as nutrition, hygiene and education, in developing countries (Lynn & Harvey, 2008). Recent research has also shown that the Flynn Effect has in fact “reversed” itself in many of the countries in which Flynn observed the initial effect (Lynn & Harvey, 2008; Must, et al., 2009; Teasdale & Owen, 2008). In short, the bulk of current research confirms that the so-called Flynn effect does not reflect an increase in actual cognitive ability, and is thus not directly pertinent to the discussion of dysgenic trends on g .

2.4 Moderators of the g -fertility relationship

Moderators in this study were tested for potential influences on the magnitude of reported effects due to: 1) statistical artifacts, and 2) group-level characteristics (substantive moderators). These moderators are described as follows:

2.4.1 Moderators Due to Methodological Artifacts.

2.4.1.1 g -score. Most of the studies included in the analysis are calculated effects based on the relation between fertility and raw test scores (e.g., Bajema, 1968; Meisenberg, 2006). This creates a potential methodological problem in that the levels of g -saturation will vary across tests/test batteries. Thus, to the degree fertility is associated with g , the observed effect size may be stronger for studies using more g -saturated tests than those using less g -saturated tests. More recent studies tend to apply factor-analytic techniques to extract estimates of g , devoid of any measurement error or systematic variance due to non- g constructs (Kanazawa, 2014; Reeve, et al., 2013; Woodley of Menie, et al., 2015). Thus, it is possible that some of the observed variation in effect sizes

is due to differences in the g -saturation of the tests used, or whether effects were calculated using analytically-derived estimates of g . Unfortunately, many studies failed to provide clear reports of the exact tests used, which otherwise would have facilitated the use of their corresponding g -loadings to correct for this artifact. As an alternative, a dichotomous variable was created to indicate whether each study used raw test scores or g -scores (derived via factor analysis) to calculate their reported effects. A moderator analysis using this variable can indicate whether the use of raw test scores significantly compromises the precision of reported effects.

2.4.1.2 Range restriction. Most of the studies included in the analysis suffer from little or no range restriction in the independent variable (raw test scores or g -scores) because most used sampling strategies that facilitated inclusion of participants with a broad range of cognitive abilities. Although some studies excluded participants with severe cognitive impairments or mental illness (eg. Chen, et al., 2013; Lynn & Van Court, 2004), no study was included in the analysis that focused on samples with a pre-defined cognitive ability, such as intellectually gifted groups. Taken as a whole, we suggest the evidence indicates little to no range restriction in cognitive ability is likely to have occurred in most studies, with the exceptions noted.

The dependent variable (number of children born) suffers two primary range restrictions. First, three studies (Conrad & Jones, 1932; Higgins, et al., 1962; Willoughby, 1928) limit the samples to families with at least one child. By not including participants with no children, this methodology has the potential to attenuate the reported effects. A dichotomous indicator was developed to test for significant differences between these three studies and the remaining studies that included participants with no children.

Several studies also report effects from individuals who are not old enough to reflect a final count of the number of children born (who reflect “incomplete” fertility). Two versions of a full-fertility moderator were developed to test for differences in effect sizes between “full fertility” and “incomplete fertility” samples. Both approaches define the age of the sample using the best measure of central tendency available (mean, and if not reported, the midpoint of the reported range). The first approach simply defines full fertility as samples with a central-tendency age of 45 years old or more. The second approach using something akin to an extreme groups design by removing studies with an “indeterminate” mean age (range 35 to 49 years old). This leaves one group of effects based on “full-fertility” defined as a sample with a mean age of 50 years or older, and the “incomplete fertility” group defined as a mean age less than 35 years old. The expectation is that the effect sizes from the full fertility groups may be somewhat smaller due to delays in childbearing, in favor of school or early career, among those of higher cognitive ability.

2.4.2 Substantive Moderators.

2.4.2.1 Gender. Evolved gender differences in mate preference have been researched extensively, and provide both a theoretical and empirical basis to explain gender differences in the *g*-fertility relationship. In general, women assign relatively more value to potential partners who signal the ability to acquire resources that provide for and protect themselves and their offspring (Trivers, 1972, Trivers & Willard, 1973).

Accordingly, previous studies identified significant gender differences across mate-preference traits such as: socioeconomic status (Buss, 1989; Hopcroft, 2006; Shackelford, et al., 2005; Wiederman & Allgeier, 1992; Wiederman, 1993), education (Hopcroft, 2006;

Shackleford, et al., 2005) and intelligence (Shackleford, et al., 2005).

Female selectivity is reinforced by gender differences in the relative lengths of their typical fertility windows, the levels of biological investment necessary to procreate, and cultural norms that place a disproportionate burden on women during child rearing (Trivers, 1972, Trivers & Willard, 1973). These higher stakes from a biological and cultural perspective are only compounded by the more “evolutionary novel trade-offs” (Kanazawa, 2010) that women face between work, education and raising a family. These differences are supported by empirical studies that have demonstrated more pronounced *g*-fertility gradients for women than men (Lynn, 1999; Reeve, et al., 2013; Retherford & Sewell, 1989).

2.4.2.2 Race. Several studies have reported larger dysgenic effects among African-American populations relative to Whites (Lynn & Van Court, 2004; Vining, 1982; Vining, 1995). A recent study (Meisenberg & Kaul, 2010) also reported larger dysgenic effects among Hispanic participants relative to Whites. Although the reasons behind these patterns remain uncertain, it has been suggested that differences may be attributed to education (Lynn, 1996) and cultural pressures on behavior (Meissenberg & Kaul, 2010). An interesting feature is that the relative rank ordering in these studies seems to be consistent with known patterns of mean differences in *g* across racial and ethnic lines. This may suggest that the dysgenic trend is operating among all racial groups, but is more pronounced among populations with lower average cognitive ability. Which, if any, of these theories are correct has not yet been established. However, for the purpose of the current study, they all lead to the hypothesis that there may be observed racial differences in *g*-fertility gradients.

2.4.2.3 Geographic Region. Although studies of the g-fertility relationship using samples from the United States and Western Europe are relatively abundant, little attention has been applied to such studies in developing nations, or to comparisons across nations (Lynn & Harvey, 2008). Since the timing of the shift towards dysgenic fertility has been largely attributed to nineteenth-century industrialization in the Western world (Lynn, 1996; Nyborg, 2012), this may suggest that differing levels of present-day industrialization would exert an influence on the relationship between intelligence and reproductive behavior, and may also yield preliminary findings that highlight the influence of culture on fertility independent of g. Although exploratory in nature, the intent is to inform future studies that integrate these potentially important influences. From a theoretical perspective (Kanazawa, 2010), populations living in less industrialized nations would experience relatively less evolutionary novelty in the reproductive process than those living in fully industrialized societies. Thus, we would expect the IQ-fertility relationship to be stronger in industrialized societies than non-industrialized societies. However, previous research has also shown that national religiosity has a large effect on fertility rates, and less industrialized nations tend to have high religiosity rates (Reeve, 2009). Thus, this could wash out any potential differences.

2.4.2.4 Historical Events. The studies included in this analysis range from very early (Willoughby, 1928) to modern-day (Wang, et al., 2016; Woodley of Menie, et al., 2016), which provides an alternate approach to exploring the relationship between cognitive ability and fertility over time. Two dichotomous variables were derived to reflect Historical Events: 1) pre and post World War II, and 2) pre and post Roe vs. Wade legislation for U.S. studies only. These moderators will be used to identify any

differences in effect sizes according to these historical events.

CHAPTER 3: THE CURRENT STUDY

The current study collected and analyzed existing empirical findings that relate measures of general intelligence (often referred to as “IQ tests” or “Cognitive Ability Tests [CATs]”) with measures of fertility to estimate: 1) an overall effect size, 2) the degree of heterogeneity of this effect across the collected studies, and 3) potential moderators of these relationships.

3.1 Hypotheses

Although this study is intended to be primarily descriptive in nature, two patterns were expected based on existing theory and previous empirical findings: 1) the inverse relationship between intelligence and fertility is expected to be significantly stronger for women than for men, and 2) the inverse relationship between intelligence and fertility is expected to be significantly weaker for White than for Hispanic and Black groups. The size of the effects for the two Asian studies were compared to those of the other racial groups for descriptive purposes only, as no known study has been performed to facilitate comparisons against this group.

3.2 Methodology

3.2.1 Search Strategy. The following databases were used for searches of available research: PsycINFO, PsycArticles, MEDLINE, Health Source Nursing Edition, Health Source Consumer Edition, Academic Search Complete and ProQuest Dissertation and Theses. Based on an inspection of keywords identified in a subset of studies focusing

on dysgenic fertility trend, variants and combinations of the following keywords were used to obtain an initial pool of articles: "intelligence", "IQ", "cognitive ability", "GMA", "reproduc*", "dysgenic", "fertility" and "birth rate". Following this, forward and backward searches using the Social Science Citation Index were performed to identify studies not found in the previous database searches, followed by manual reviews of the reference sections from the selected studies. Finally, prominent experts in the field were contacted to identify any unpublished study results or any studies currently in press.

3.2.2 Inclusion/exclusion criteria. This review included studies published in English and available through January 2016 that related quantitative measures of cognitive abilities (e.g., Wonderlic, Ravens, etc.) or intellectual achievement tests (e.g., Program for International Student Assessment) with the total number of biological children as the indicator of fertility. Upon collecting the initial pool, studies were screened for relevance and reported effect sizes in the form of bivariate correlations, other effect-size statistics that can be converted to bivariate correlations, or standardized linear regression (beta) weights that can be transformed in keeping with guidelines outlined in Peterson & Brown (2005).

Figure 1 depicts the details of the literature selection procedure. As shown, we began with 737 studies, and through this procedure included seventeen studies in this analysis. Several published articles sourced identical data sources, and in those cases, a single article was selected based on reported effects deemed most conducive to the moderator analyses performed in this paper. Characteristics of the included studies are presented in Table 1.

3.2.3 Coding Procedures. Coded data are maintained in an Excel table that includes the following variables: bivariate correlation or (or convertible effect-size statistics) relating g and measures of fertility, and sample size. Additionally, this table includes moderators such as sample mean age or age range, gender, race, country and other indicators to reflect study and sample characteristics required to perform the analyses for methodological artifacts. The author inputted the data for this study, with 20% of the included articles double coded by an independent researcher to ensure accuracy.

3.2.4 Meta-analytic procedures. The initial analysis estimated the weighted mean effect size and its distribution across studies. Analyses utilized a Random Effects Model because effects are expected to be heterogeneous across studies due to potentially diverse measures of g (i.e., differential g -loadings), sample characteristics, and operationalizations of fertility. These and other potentially relevant characteristics were classified in the database and moderator analyses were conducted to identify any significant differences in effect sizes across studies due to methodological artifacts. Consistent with recommendations of the Cochrane collaborative (Higgins & Green, 2008) the present study adopted an alpha level of .10.

The primary analyses were conducted using Comprehensive Meta-analysis (CMA) Version 3 (Borenstein, et al., 2015). A shifting unit of analysis approach, as prescribed by Cooper (2010), was used where multiple effects are drawn from the same samples, yielding for each sample one effect per category per moderator test, and avoiding the problem of weighting certain samples more heavily based on the number of effects reported. Inverse variance weights to correct for measurement imprecision was

applied in these analyses according to recommendations provided in Lipsey and Wilson (2001). Effect-size statistics that are convertible to bivariate correlations were converted using the Practical Meta-Analysis Effect Size Calculator that accompanies Lipsey and Wilson (2001).

3.2.5 Analysis for publication bias. Publication bias can result when studies yielding null findings fail to achieve publication. To the extent that this occurs, conventional literature searches may overlook these studies causing estimated meta-analytic effect sizes to be overstated and statistically significant relationships to be potentially reported in error (Borenstein, et al., 2009). Although there is no way to account for publication bias directly, there are techniques to estimate the potential for this to occur based on the studies that were identified through the literature search. Rosenthal's Fail-Safe N calculates the number of studies required with null results to render the p-value of the overall effect to be greater than .05. Orwin's Fail-Safe N is considered a more conservative test of publication bias because it does not assume null results; rather, it allows for selection of a trivial effect (Lipsey & Wilson, 2001). The present study chose an effect size equal to less than half of the overall point-estimate as meeting the "trivial" criterion.

A third method of detecting publication bias is through the creation of a funnel plot. Publication bias is evident when the plot depicts an asymmetrical distribution of effects about the overall point-estimate (represented by a vertical line), suggesting that studies of smaller sample sizes (and therefore greater standard error) and with large effects are favored by publishers due to favorable outcomes (Borenstein, et al., 2009). However, in the current meta-analysis, most effect sizes were based on very large

samples (average $N = 21,666$). Thus, the use of the funnel plot in this is somewhat limited as almost all of the observations lie at the tip of the funnel.

CHAPTER 4: RESULTS

Table 2 depicts the results of the analysis of the overall effect from combined the studies, and moderator analyses for methodological artifacts. The overall correlation between reported cognitive ability scores and the number of children born is $r = -.11$, with a 95% confidence interval of $r = -.08$ to $r = -.13$. This supports the hypothesis of a small, but significant dysgenic trend. However, study level effect sizes were found to be heterogeneous, $\chi^2(16) = 292.17, p < .001$, confirming that there is significant variability in the study-level effects. The I^2 index of 94.52%, which can be loosely conceptualized as a ratio of “signal-to-noise” (Borenstein, et al., 2009), supports the conclusion that the observed heterogeneity in study-level effects reflects primarily “true” variation rather than merely measurement error.

We first conducted the analyses of moderators for methodological artifacts. Results are presented in Table 2. First, the results show no meaningful differences in effects between studies that use raw test scores vs. factor-analytically derived measures of g . This suggests that the use of raw test scores is not a methodological concern in terms of estimating the dysgenic trend, however the limited number of studies that used g -scores ($K = 3$) may be a factor behind the non-significant result.

The results do not indicate significant differences in effects between samples that are limited to participants that have at least one child vs. samples where adults with no children are included. However, the weighted effect for the parent-only group ($r(2) =$

-.08, $p < .10$) is observed to be smaller than the group that includes adults with no children ($r(13) = -.11, p < .001$). Due to the limited number of studies that include only parents ($K = 3$), it remains uncertain whether limiting samples in this manner attenuates measured effects.

The results provide evidence that effects are larger amongst samples with incomplete fertility compared to samples who have likely achieved complete fertility. Using both versions of our coding system for this moderator yields significant effects; $\chi^2(18) = 6.25, p < .01$ when incomplete fertility groups are classified as less than 45 years old, and $\chi^2(11) = 26.49, p < .001$ when groups classified as 35-49 years old are excluded. These differences also appeared when data were analyzed for men and women separately; though the difference was not significant among men when the “indeterminate” studies were included. These results suggest that the dysgenic trend is stronger among younger adults and wanes as a generational cohort ages. That is, more intelligent people appear to be delaying reproduction until later in life compared to less intelligent people who begin reproducing earlier in life. Ultimately this indicates that less intelligent groups will have shorter inter-generational windows than more intelligent people.

Analyses of substantive moderators (Table 3) yield results that are generally consistent with both theory and existing empirical findings. The weighted effect for women ($r(13) = -.13, p < .001$) is over double that for men ($r(14) = -.06, p < .01$) and statistically significant ($\chi^2(28) = 5.96, p < .01$), although the effects from samples within the two groups vary widely (women, $\chi^2(13) = 52.70, p < .001$; men, $\chi^2(14) = 362.50, p < .001$). This latter finding suggests that heterogeneity in the overall effect may not be fully accounted for by differences between men and women.

Group effects are also significantly different between each of the four identified race groups ($\chi^2 (15) = 21.87, p < .001$). Consistent with the mean differences in the distributions of intelligence, effects are larger for Blacks ($r (2) = -.017, p < .001$) and Hispanics ($r (0) = -.23, p < .001$) than Whites ($r (9) = -.07, p < .001$). The effect for Asians ($r (1) = -.11, p < .001$) is likewise weaker than for Blacks and Hispanics.

These racial differences appear to be concentrated primarily among women ($\chi^2 (14) = 56.87, p < .001$), with no statistically significant differences across racial groups for men ($\chi^2 (14) = 4.42, p > .10$). This may be due a combination of the stronger dysgenic effect for women in general, combined with limitations in the number of studies available for most of the race groups (Black: $K = 3$, Hispanic: $K = 1$, and Asian: $K = 2$), which may have contributed to the non-significant finding for men.

Although the weighted effect for the two European studies was less than half of those for the other geographic regions ($r (1) = -.05, p > .10$), this moderator analysis resulted in no significant differences across the four groups. This may also be due to limitations in the number of studies using samples for regions outside of the United States (Asia: $K = 2$, Dominica: $K = 1$, Europe: $K = 2$).

Of the two moderator variables that classified historical events, only the pre vs. post Roe vs. Wade moderator demonstrated a statistically significant difference between the two groups ($\chi^2 (13) = 6.90, p < .01$), with a much stronger weighted effect from post Roe vs. Wade studies ($r (5) = -.14, p < .001$). Not surprisingly, further analysis shows the difference is significant among women ($\chi^2 (10) = 19.52, p < .001$). The same analysis for men confirmed the dysgenic effect was relatively stronger post Roe v. Wade ($r (4) = -.10,$

$p < .001$) than prior ($r(5) = -.04, p > .10$). Although this difference is not statistically significant, this is may be due to the limited number of studies.

To further explore the trend in the dysgenic relationship over time, we computed a bivariate correlation between the date of study (scaled according to the age of each sample) with the corresponding effect sizes, yielding a moderate association with trend-level significance ($r = -.37, p = .05$). This produced some evidence to suggest that the more recent samples are displaying a stronger dysgenic trend. Figure 3 reflects the trend of the study-level effect sizes over time.

Rosenthal's Fail-Safe N indicates that 7,552 studies with null findings would be required reduce the overall effect to a level that is non-significant at $p > .05$. Orwin's Fail-Safe N indicates that it would require 36 null studies to reduce the overall effect to less than $r = -.05$, which is less than half of the observed weighted effect. These results lead to the conclusion that publication bias is not a major concern for the current study.

CHAPTER 5: DISCUSSION

The purpose of the current study was to conduct a meta-analysis of the dysgenic trend in modern humans. Based on a $K=17$, we found that there is a small but significant and meaningful dysgenic trend.

Several moderators were tested to identify sources of this variation that may provide evidence to explain the range of effects sizes included in this analysis. The collection of moderators is classified into two groups: 1) moderators that reflect methodological differences at the study level, and 2) moderators that classify sample characteristics. The results of the methodological moderators will assist the research community in gauging the extent to which differences in study-level effects are the result of methodological flaws. The results of the substantive moderator analyses enable us to gauge the degree to which measurable human characteristics may influence the relationship between intelligence and fertility. Taken together, the results from these two moderator groups have the potential to tease apart “true” differences that influence the relationship from those methodological differences that could obscure reported effects.

The moderator to test for significant differences between studies that calculated effects using g -scores vs. raw scores yielded effects for the two groups that were materially the same. This may be due to the fact that raw scores generated through cognitive ability tests, although not a direct measure of g , are by design highly g -saturated. This result also serves to assuage concerns about measurement error in the

independent variable. Reliability estimates were not available in most of the included studies, however *g*-scores are deemed perfectly reliable due to the psychometric techniques applied to derive them. Given this, the moderator analysis effectively tested for differences in effects for those studies that directly corrected for reliability of the independent variable against studies that did not, and likewise produced no evidence of significant differences between the two groups.

The moderator for samples that included only parents with at least one child did not produce a statistically significant difference from studies that included respondents with no children. The weighted effect for the parent only group was noticeably weaker however, suggesting that perhaps restricting samples to only parents may in fact attenuate the reported effect due to statistical restriction in range, and raising the substantive possibility that higher IQ people may be more likely to never reproduce. The absence of a significant difference between the two groups may be the result of a limited number of parent-only studies ($K = 3$).

Of the methodological moderators tested, only the two full fertility indicators produced significant differences between the reported effects, with the full-fertility samples yielding smaller effect sizes compared to the incomplete fertility samples. The more restrictive of the two full-fertility indicators (with the indeterminate age range removed) showed a more pronounced difference than the single-break indicator at 45 years old. These results provide evidence that the size of the relationship in general may be inflated in studies where participants who have not reached the end of their fertility windows are included.

The weaker effects for the full-fertility samples, regardless of gender, are consistent with more intelligent people possibly deferring childbirth in favor of education and career pursuits, especially for women, who tend to be sidelined for longer due to pregnancy and possibly to raise the children in families that assume more traditional gender roles (Retherford & Sewell, 1989). This is also consistent with previous empirical findings that have shown education and SES serve to mediate and moderate the relationship in question (Reeve, et al., 2013).

The differences between these two groups is more pronounced for men, and significantly more so with the more restrictive version of the full-fertility moderator. Defining “full fertility” for men is less exact due to their theoretically boundless fertility windows, and the evidence underscores the importance of measuring this relationship for men in their later years to obtain accurate estimates. Conversely, researchers should take care to not limit the age of respondents to thresholds that are too old, as attrition (through death) may occur, creating a competing methodological concern.

In contrast, measuring dysgenic effects for women is relatively straightforward because their fertility windows generally expire somewhere near 50 years old or earlier. The results of both moderator analyses indicate this through more consistency in the effects between the incomplete and full-fertility groups for women. These results reflect both heightened methodological complexity in measuring effects for men and differences in lifetime reproductive capacities for men versus women.

The gender moderator yielded a significantly higher effect for women that is over double than that for men. This finding is consistent with both theory and empirical findings that have been previously well-established. This difference by gender carries

into each of the identified race categories, with effects from two of the four racial groups for men (Asian and Black) not significantly different from zero, and the remaining two categories (Hispanic and White) still smaller than those for women and significant within only $p < .05$. Except for Black men (who have the weakest effect of all the race-gender combinations), there is consistent rank-ordering in the magnitude of effects by race within each gender group, with Hispanics producing the highest effects, followed by Black, Asian and then White respondents. This pattern between the defined racial groups is consistent with previous research that cites education and socioeconomic status (Reeve, et al., 2013) and cultural norms (Meissenberg & Kaul, 2010) as potentially relevant to the magnitude of the association between cognitive ability and fertility.

Weighted effects by geographic region were relatively uniform except for the European studies that yielded a magnitude of less than half of each of the other three groups. Limitations in the available studies did not permit rigorous analysis according to relative levels of industrialization of the compared geographies, with only one study from Dominica that may be considered less industrialized than the remaining three groups. At present, no known comparable studies exist for groups residing outside of the four groups represented in this analysis, which calls for future studies that examine samples from more diverse locations.

Of the two moderators tested for historical events, only the Roe vs. Wade moderator (for U.S. based studies) yielded significant differences. The weighted effect for the post Roe vs. Wade period is approximately double the effect for the period prior. This suggests that the dysgenic may be stronger when people have full control over their reproductive options than when they are restricted by governments or other powerful

entities (e.g., religious organizations). However, it may also be possible that the stronger effects post *Roe vs. Wade* are merely a reflection of the increasing magnitude of the dysgenic trend over time. Our analysis shows a correlation of $r = -.37$ between the size of the dysgenic effect and date of the study. This may reflect Kanazawa's (2010) hypothesis that the effect of g will become increasingly stronger as the world that we inhabit becomes increasingly complex. Whether these observations are the result of the removal of artificial barriers to choice, or a constant trend due to an increasingly "evolutionary novel" environment, will require further study.

5.1 Limitations

This study attempted to address some of the methodological issues that may influence the reliability of study-level effects through the development and testing of specific moderators. This serves as a substitute for psychometric meta-analytic procedures recommended by Schmidt & Hunter (2015) but for which limitations in the data rendered impractical for this study. The principal methodological challenge appears in the differences between the complete and incomplete fertility groups, which suggests that the inclusion of respondents who have not completed their reproductive lives may inflate the magnitude of reported effect sizes. A correction for this artifact would be difficult to perform accurately due to differences in the fertility windows between men and women, individual differences regarding this personal decision, cultural and other influences that are not yet completely understood.

It is not possible to analyze all the substantive moderators simultaneously due to the limited number of studies and limitations in the number of studies representing different moderator groups. Meta-regression is a common approach when these

limitations are not present, but the presence of consistent differences in the magnitudes of effects for men and women permitted the separation of moderator groups by gender when warranted, providing a reasonable means to delve deeper into observed differences between other moderator groups.

Several moderators (such as religiosity, socioeconomic status, and Education) are not included in this analysis, but were demonstrated in previous research (Reeve, 2009; Reeve, et al., 2013) to influence the relationship between cognitive ability and fertility. These measures were not available from most of the included studies, but their strong correlations with *g* suggest that their absence may not dramatically compromise the results of this study. We were, however, able to account for those influences that explain at least some of the heterogeneity in the overall effect, with some methodological concerns addressed and others that identify and estimate group-level differences in the relationship.

REFERENCES

* References marked with an asterisk indicate sources that were included in the meta-analysis.

- *Bajema, C.J. (1968). Relation of fertility to occupational status, IQ, educational attainment, and size of family of origin: a follow-up study of a male Kalamazoo public school population. *Eugenics Quarterly*, 15, 198-203.
- Borenstein, M., Hedges, L.V., Higgins, J. P. T., & Rothstein H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Borenstein, M., Hedges, L.V., Higgins, J. P. T., & Rothstein H. R. (2015). *Comprehensive Meta-Analysis*, Version 3. Englewood, NJ: Biostat.
- Burks, B. S., & Jones, H. E. (1935). A study of differential fertility in two California cities. *Human Biology*, 7, 539-554.
- Burt, C. (1947). Family size, intelligence, and social class. *Population Studies*, 1(2), 177-186.
- Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *The Behavioral and Brain Sciences*, 12, 1-49.
- *Chen, H., Chen, Y., Liao, Y., & Chen, H. (2013). Relationship of fertility with intelligence and education in Taiwan: A brief report. *Journal of Biosocial Science*, 45(4), 567-571. doi:10.1017/S0021932012000545
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40(3), 153-193. doi:10.1037/h0059973
- *Conrad, H. S., & Jones, H. E. (1932). A field study of the differential birth rate. *Journal of the American Statistical Association*, 27, 153-159.

- Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach (4th ed.)*. Thousand Oaks, CA, US: Sage Publications, Inc.
- Figueredo, A. J., Vásquez, G., Brumbach, B. H., Sefcek, J. A., Kirsner, B. R., & Jacobs, W. J. (2005). The K-factor: Individual differences in life history strategy. *Personality and Individual Differences, 39*(8), 1349-1360.
doi:10.1016/j.paid.2005.06.009
- Figueredo, A. J., Vasquez, G., Brumbach, B. H., Schneider, S. R., Sefcek, J. A., Tal, I. R., & ... Jacobs, W. J. (2006). Consilience and Life History Theory: From genes to brain to reproductive strategy. *Developmental Review, 26*(2), 243-275.
- Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*(1), 29-51. doi:10.1037/0033-2909.95.1.29
- Gordon, R. A. (1997). Everyday life as an intelligence test: Effects of intelligence and intelligence context. *Intelligence, 24*(1), 203-320. doi:10.1016/S0160-2896(97)90017-9
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*, 79–132.
- Gottfredson, L. S. (2004). Intelligence: is it the epidemiologists' elusive fundamental cause of social class inequalities in health? *Journal of Personality and Social Psychology, 86*, 174–199.
- Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York, NY: The Free Press.
- Higgins, J. P., & Green, S. (2008). Identifying and measuring heterogeneity. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0, 1*.

- *Higgins, J. V., Reed, E. W., & Reed, S. C. (1962). Intelligence and family size: A paradox resolved. *Eugenics Quarterly*, 9(2), 84-90.
doi:10.1080/19485565.1962.9987508
- Hopcroft, R. L. (2006). Sex, status, and reproductive success in the contemporary United States. *Evolution and Human Behavior*, 27(2), 104-120.
doi:10.1016/j.evolhumbehav.2005.07.004
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Kanazawa, S. (2004). General intelligence as a domain-specific adaptation. *Psychological Review*, 111(2), 512-523. doi:10.1037/0033-295X.111.2.512
- Kanazawa, S. (2010). Evolutionary psychology and intelligence research. *American Psychologist*, 65, 279–289. doi:10.1037/a0019378
- *Kanazawa, S. (2014). Intelligence and childlessness. *Social Science Research*, 48, 157–170.
- Kane, H., & Oakland, T. D. (2000). Secular Declines in Spearman's g: Some Evidence from the United States. *Journal of Genetic Psychology*, 161(3), 337.
- Lentz, T. (1927). Relation of I.Q to size of family. *Journal of Educational Psychology*, 18, 486–496.
- Lipsey, M. W, & Wilson, D. B (2001). *Practical meta-analysis*. Thousand Oaks, CA, US: Sage Publications, Inc.
- Lubinski, D., & Benbow, C. P. (2000). States of excellence. *American Psychologist*, 55, 137-150.
- Lynn, R. (1996). *Dysgenics: Genetic deterioration in modern populations*. CT., Praeger: Westport.

- Lynn, R. (1999). New evidence for dysgenic fertility for intelligence in the United States. *Social Biology*, 46(1-2), 146-153. doi:10.1080/19485565.1999.9988992
- Lynn, R., & Harvey, J. (2008). The decline of the world's IQ. *Intelligence*, 36(2), 112-120. doi:10.1016/j.intell.2007.03.004
- *Lynn, R., & Van Court, M. (2004). New evidence of dysgenic fertility for intelligence in the United States. *Intelligence*, 32(2), 193-201.
- Meisenberg, G. (2010). The reproduction of intelligence. *Intelligence*, 38(2), 220-230. doi:10.1016/j.intell.2010.01.003
- *Meisenberg, G., & Kaul, A. (2010). Effects of sex, race, ethnicity, and marital status on the relationship between intelligence and fertility. *Mankind Quarterly*, 50, 151-187.
- *Meisenberg, G., Lawless, E., Lambert, E., & Newton, A. (2006). The social ecology of intelligence on a Caribbean island. *Mankind Quarterly*, 46(4), 395-433.
- Must, O., te Nijenhuis, J., Must, A., & van Vianen, A. E. M. (2009). Comparability of IQ scores over time. *Intelligence*, 37, 25-33.
- Neiss, M., Rowe, D. C., & Rodgers, J. L. (2002). Does education mediate the relationship between IQ and age of first birth? A behavioural genetic analysis. *Journal of Biosocial Science*, 34, 259-275.
- Nyborg, H. (2012). The decay of Western civilization: Double relaxed Darwinian selection. *Personality and Individual Differences*, 53, 118-125.

- Peach, H., Lyerly, J. E., & Reeve, C. L. (2014). Replication of the Jensen effect on dysgenic fertility: An analysis using a large sample of American youth. *Personality and Individual Differences, (71)*56-59.
doi:10.1016/j.paid.2014.07.017
- Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology, 90*(1), 175-181. doi:10.1037/0021-9010.90.1.175
- Pianka, E. R. (1970). On r-and K-selection. *The American Naturalist, 104*(940), 592–597.
- Razmyar, S., & Reeve, C. L. (2013). Individual differences in religiosity as a function of cognitive ability and cognitive style. *Intelligence, 41*(5), 667-673.
doi:10.1016/j.intell.2013.09.003
- Reeve, C. L. (2009). Expanding the g-nexus: Further evidence regarding the relations among national IQ, religiosity and national health outcomes. *Intelligence, 37*(5), 495-505. doi:10.1016/j.intell.2009.06.003
- Reeve, C. L., & Bonaccio, S. (2011). *The nature and structure of 'intelligence'*. In T. Chamorro-Premuzic, S. von Stumm, A. Furnham, T. Chamorro-Premuzic, S. von Stumm, A. Furnham (Eds.), *The Wiley-Blackwell handbook of individual differences* (pp. 187-216). Wiley-Blackwell.
- *Reeve, C. L., Lyerly, J. E., & Peach, H. (2013). Adolescent intelligence and socio-economic wealth independently predict adult marital and reproductive behavior. *Intelligence, 41*(5), 358-365. doi:10.1016/j.intell.2013.05.010
- Retherford, R. D., & Sewell, W. H. (1988). Intelligence and family size reconsidered. *Biodemography and Social Biology, 35*(1-2), 1-40.
doi:10.1080/19485565.1988.9988685

- *Retherford, R. D., & Sewell, W. H. (1989). How intelligence affects fertility. *Intelligence*, 13(2), 169-185. doi:10.1016/0160-2896(89)90015-9
- Rindermann, H. (2012). Intellectual classes, technological progress, and economic development: The rise of cognitive capitalism. *Personality & Individual Differences*, 53(2), 108-113. doi:10.1016/j.paid.2011.07.001
- Rindermann, H., Sailer, M., & Thompson, J. (2009). The impact of smart fractions, cognitive ability of politicians and average competence of peoples on social development. *Talent Development and Excellence*, 1, 3–25.
- Rushton, J. P. (1985). Differential K theory: The sociobiology of individual and group differences. *Personality and Individual Differences*, 6, 441–452.
- Rushton, J. P. (1999). Secular gains in IQ not related to the g factor and inbreeding depression — unlike Black–White differences: A reply to Flynn. *Personality and Individual Differences*, 26, 381–389.
- Rushton, J. P. (2004). Placing intelligence into an evolutionary framework or how g fits into the r–K matrix of life-history traits including longevity. *Intelligence*, 32(4), 321–328.
- Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings (3rd ed.)*. Thousand Oaks, CA; London and New Delhi.
- Shackelford, T. K., Schmitt, D. P., & Buss, D. M. (2005). Universal dimensions of human mate preferences. *Personality & Individual Differences*, 39 (2), 447-458. doi:10.1016/j.paid.2005.01.023
- Spearman, C. (1904). “General intelligence,” objectively determined and measured. *American Journal of Psychology*, 15, 201–292.

- Spearman, C. (1927). *The nature of intelligence and the principles of cognition*. 2d ed. Oxford, England: Macmillan.
- Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The End of the Flynn Effect? *Intelligence*, 32(4), 349-362.
- te Nijenhuis, J., & van der Flier, H. (2013). Is the Flynn effect on g?: A meta-analysis. *Intelligence*, 41(6), 802-807. doi:10.1016/j.intell.2013.03.001
- Teasdale, T. W., & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse. *Personality & Individual Differences*, 39(4), 837-843. doi:10.1016/j.paid.2005.01.029
- Teasdale, T. W., & Owen, D. R. (2008). Secular Declines in Cognitive Test Scores: A Reversal of the Flynn Effect. *Intelligence*, 36(2), 121-126.
- Trivers, R. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man, 1871–1971* (pp. 136–179). Chicago: Aldine.
- Trivers, R., & Willard, D. (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science*, 179, 90–92.
- *Vining, D. R. (1982). On the possibility of the reemergence of a dysgenic trend with respect to intelligence in American fertility differentials. *Intelligence*, 6(3), 241-264. doi:10.1016/0160-2896(82)90002-2
- *Vining, D. R. (1995). On the possibility of the reemergence of a dysgenic trend with respect to intelligence in American fertility differentials: An update. *Personality and Individual Differences*, 19(2), 259-263. doi:10.1016/0191-8869(95)00038-8

- *von Stumm, S., Batty, G. D., & Deary, I. J. (2011). Marital status and reproduction: Associations with childhood intelligence and adult social class in the Aberdeen children of the 1950s study. *Intelligence*, *39*(2), 161-167.
- *Wang, M., Fuerst, J., & Ren, J. (2016). Evidence of dysgenic fertility in China. *Intelligence*, *57*, 15-24. doi:10.1016/j.intell.2016.04.001
- Wiederman, M. W., & Allgeier, E. R. (1992). Gender differences in mate selection criteria: Sociobiological or socioeconomic explanation? *Ethology & Sociobiology*, *13*(2), 115-124. doi:10.1016/0162-3095(92)90021-U
- Wiederman, M. W. (1993). Evolved gender differences in mate preferences: Evidence from personal advertisements. *Ethology & Sociobiology*, *14*(5), 331-351. doi:10.1016/0162-3095(93)90003-Z
- Williams, R. L. (2013). Overview of the Flynn Effect. *Intelligence*, *41*(6), 753-764. doi:10.1016/j.intell.2013.04.010
- * Willoughby, R. R. (1928). Fertility and parental intelligence. *The American Journal of Psychology*, *40*, 671-672. doi:10.2307/1414387
- *Willoughby, R. R., & Coogan, M. (1940). The correlation between intelligence and fertility. *Human Biology*, *12*(1), 14-119.
- Woodley, M. A., & Figueredo, A. J. (2013). *Historical variability in heritable general intelligence: Its evolutionary origins and socio-cultural consequences*. Buckingham, UK: The University of Buckingham Press.
- Woodley, M. A., & Meisenberg, G. (2013). In the Netherlands the anti-Flynn effect is a Jensen effect. *Personality & Individual Differences*, *54*(8), 871-876. doi:10.1016/j.paid.2012.12.022

*Woodley of Menie, M. A., Figueredo, A. J., Dunkel, C. S., & Madison, G. (2015).

Estimating the strength of genetic selection against heritable g in a sample of 3520 Americans, sourced from MIDUS II. *Personality and Individual Differences*, 86, 266-270. doi:10.1016/j.paid.2015.05.032

*Woodley of Menie, M. A., Schwartz, J. A., & Beaver, K. M. (2016). How cognitive

genetic factors influence fertility outcomes: A mediational SEM analysis. *Twin Research and Human Genetics*, 19(6), 628-637. doi:10.1017/thg.2016.82

Table 1: Study Characteristics

Author	Study N	Region	Study Name and Description	Test	g -score	Parent Only Sample?	Full Fertility ^a	Full Fertility ^b
Bajema (1968) Chen, et al. (2013)	437 73	US Asia	Kalamazoo Fertility Study. Sample of Taiwan adults collected by author.	Terman Group Intelligence Test WAIS-III (Taiwan Version)	No No	No No	Yes Yes	Indeterminate Yes
Conrad & Jones (1932) Higgins, Reed & Reed (1962)	230 2,032	US US	Rural New England sample collected by author. Study of two cohorts. Minnesota State School and Hospital Study by the Dight Institute for Human Genetics. Age data not clearly reported.	Army Alpha Results were obtained from a variety of cognitive ability and aptitude tests based on availability	No No	Yes Yes	Both Unknown	No Unknown
Kamazawa (2014)	4,973	Europe	National Child Development Study (NCDS). DV was dichotomized (Have Children/No Children).	A range of tests administered in waves. Factor rotated to extract general cognitive ability (<i>g</i>)	Yes	No	Yes	Indeterminate
Lynn & Van Court (2004) Messenberg, et al (2006)	6,244 570	US Dominica	General Social Survey (GSS). Study consists of multiple cohorts. Dominica samples collected by author. Study of two cohorts.	General Social Survey (GSS) Vocabulary Scores Average of Raven's and 30-Item Vocabulary Test	No No	No No	Both Both	Both Both
Meisenberg & Kaul (2010)	6,370	US	National Longitudinal Study of Youth 1979 (NLSY79)	Armed Services Vocational Aptitude Battery (ASVAB). Raw scores from 10 test batteries, <i>g</i> -extracted and age adjusted.	Yes	No	No	Indeterminate
Reeve, Lyerly & Peach (2013)	317,755	US	Project Talent Database	Project Talent Ability Battery	Yes	No	No	No
Retherford & Sewell (1989)	9,098	US	The Wisconsin Longitudinal Study of Social and Psychological Factors in Educational and Occupational Aspirations and Achievements (WLS)	Hennon-Nelson Test of Mental Ability (Revised 1954)	No	No	No	Indeterminate

Date of Study moderators were scaled to the year in which the sample's age (mean, or if not reported, median of a reported range) is 30 years old. Full Fertility (a) reflects the sample's age (mean, or if not reported, median of a reported range) is at least 45 years old. Full Fertility (b) reflects the sample's age (mean, or if not reported, median of a reported range) is at least 50 years old, with the indeterminate age range (35-49) removed.

Table 1 (Continued)

Author	Study N	Region	Study Name and Description	Test	g -score	Parent Only Sample?	Full Fertility ^a	Full Fertility ^b
Vining (1982/1995)	4,150	US	National Longitudinal Study of Labor Market Experience. A 1995 update was published for women only using the original sample that consisted of both genders, first published in 1982. The effects for women in the 1982 study were excluded in favor of the 1995 results.	Results were obtained from a variety of cognitive ability and aptitude tests based on availability	No	No	No	No
von Strumm, Batty & Deary (2011)	5,647	Europe	The Aberdeen children's Study of 1950's. Effects were calculated based on the Cochrane Collaboration Effect Size Calculator based on the reported mean, standard deviation and N for two groups: "have children" and "no children".	Moray House Verbal Reasoning Tests I and II, and an arithmetic and English test.	No	No	Yes	Indeterminate
Wang, Fuerst & Ren (2016)	4,770	Asia	China Family Panel Studies (CFPS) Dataset	Immediate Word Recall, Delayed Word Recall and Number Series Tests	No	No	Yes	Yes
Willoughby (1928)	195	US	Sample of California families collected by author	A series of combined cognitive ability tests	No	Yes	No	Indeterminate
Willoughby & Coogan (1940)	373	US	Providence RI high school records combined with marriage and birth records from the State Bureau of Vital Statistics	Unknown	No	No	No	No
Woodley of Menie, et al (2015)	3,520	US	Midlife in the United States: A National Longitudinal Study of Health and Well-Being (MIDUS II)	Brief Test of Adult Cognition (BTACT)	Yes	No	Yes	Yes
Woodley of Menie, Schwartz & Beaver (2016)	1,886	US	Add Health	Picture Vocabulary Test (PVT), a modified version of the Peabody Picture Test - Revised (PPVT-R).	No	No	No	No

Date of Study moderators were scaled to the year in which the sample's age (mean, or if not reported, median of a reported range) is 30 years old. Full Fertility (a) reflects the sample's age (mean, or if not reported, median of a reported range) is at least 45 years old. Full Fertility (b) reflects the sample's age (mean, or if not reported, median of a reported range) is at least 50 years old, with the indeterminate age range (35-49) removed.

Table 2: Overall point estimate and moderator analyses for methodological artifacts

Analysis	<i>k</i>	Weighted <i>r</i>	95% CI for <i>r</i>		<i>Q_b</i>	<i>Q_w</i>	<i>I</i> ²
Overall Point-Estimate	17	-0.11***	-0.13	-0.08		292.17***	94.52
Methodological Artifacts							
Raw Score vs. <i>g</i>-Score	17				0.01		
Raw Score	13	-0.10***	-0.14	-0.07		55.66***	78.44
<i>g</i> -Score	4	-0.11***	-0.16	-0.06		110.64***	97.29
Parents Only Samples	17				0.34		
Parents Only	3	-0.08 [†]	-0.16	0.00		0.67	0.00
Incl. Zero Children	14	-0.11***	-0.14	-0.08		281.03***	95.37
Incomplete vs. Full Fertility^a	19				6.25*		
Incomplete Fertility	10	-0.14***	-0.17	-0.10		115.52***	92.21
Full Fertility	9	-0.07***	-0.11	-0.04		37.70***	78.78
Female Only:	16				1.21		
Incomplete Fertility	8	-0.14***	-0.17	-0.11		25.85***	72.92
Full Fertility	8	-0.12***	-0.15	-0.09		18.51**	62.18
Male Only:	17						
Incomplete Fertility	8	-0.10**	-0.16	-0.03	2.33	159.28***	95.61
Full Fertility	9	-0.03	-0.09	0.03		27.18***	70.57
Incomplete vs. Full Fertility^b	12				26.49***		
Incomplete Fertility	7	-0.16***	-0.17	-0.16		5.63	0.00
Full Fertility	5	-0.11***	-0.13	-0.09		4.73	15.35
Female Only:	9				0.90		
Incomplete Fertility	4	-0.17***	-0.20	-0.13		5.23	42.66
Full Fertility	5	-0.14***	-0.18	-0.11		4.62	13.36
Male Only:	9				6.78**		
Incomplete Fertility	4	-0.15***	-0.18	-0.11		11.53**	73.98
Full Fertility	5	-0.07**	-0.11	-0.03		2.89	0.00

Note. *** $p < .001$, ** $p < .01$, * $p < .05$, [†] $p < .10$

Full Fertility (a) reflects the sample's age (mean, or if not reported, median of a reported range) is at least 45 years old. Full Fertility (b) reflects the sample's age (mean, or if not reported, median of a reported range) is at least 50 years old, with the indeterminate age range (35-49) removed.

Table 3: Results for substantive moderators

Analysis	<i>k</i>	Weighted <i>r</i>	95% CI for <i>r</i>		<i>Q_b</i>	<i>Q_w</i>	<i>I</i> ²
Substantive Moderators							
Gender	29				5.96*		
Female	14	-0.13***	-0.17	-0.09		52.70***	75.33
Male	15	-0.06**	-0.10	-0.02		362.50***	96.14
Race	16				21.87***		
Asian	2	-0.11***	-0.18	-0.05		2.21	54.65
Black	3	-0.17***	-0.22	-0.12		4.84 [†]	58.67
Hispanic	1	-0.23***	-0.31	-0.14		0.00	0.00
White	10	-0.07***	-0.09	-0.04		35.63***	74.74
Race (Female Only)	15				56.87***		
Asian	2	-0.14***	-0.19	-0.10		1.83	45.45
Black	3	-0.25***	-0.29	-0.21		0.53	0.00
Hispanic	1	-0.27***	-0.35	-0.19		0.00	0.00
White	9	-0.09***	-0.11	-0.08		10.03	20.22
Race (Male Only)	15				4.42		
Asian	2	-0.08	-0.17	0.02		0.37	0.00
Black	2	-0.01	-0.10	0.08		0.09	0.00
Hispanic	1	-0.17*	-0.30	-0.04		0.00	0.00
White	10	-0.05*	-0.08	-0.01		44.10***	79.59
Geographic Region	17				3.43		
Asia	2	-0.12**	-0.21	-0.04		2.21	54.65
Dominica	1	-0.12*	-0.24	0.00		0.76	0.00
Europe	2	-0.05	-0.11	0.02		0.00	0.00
U.S.	12	-0.11***	-0.14	-0.08		151.66***	92.75
Date of Study^a	18				0.04		
Post WWII	13	-0.11***	-0.14	-0.07		279.23***	95.70
Pre WWII	5	-0.10**	-0.16	-0.03		2.96	0.00
Date of Study^b	13				6.90**		
Post Roe vs. Wade	6	-0.14***	-0.17	-0.11		50.23***	90.05
Pre Roe vs. Wade	7	-0.08***	-0.11	-0.04		11.64 [†]	48.44
Female Only:	10				19.52***		
Post Roe vs. Wade	5	-0.15***	-0.15	-0.15		7.44	46.23
Pre Roe vs. Wade	5	-0.10***	-0.12	-0.07		0.26	0.00
Male Only:	11				1.51		
Post Roe vs. Wade	5	-0.10***	-0.17	-0.04		108.77***	96.32
Pre Roe vs. Wade	6	-0.04	-0.11	0.03		7.12	0.21

Note. *** $p < .001$, ** $p < .01$, * $p < .05$, [†] $p < .10$

Date of Study moderators were scaled to the year in which the sample's age (mean, or if not reported, median of a reported range) is 30 years old. Roe vs. Wade analysis was limited to U.S. studies only.

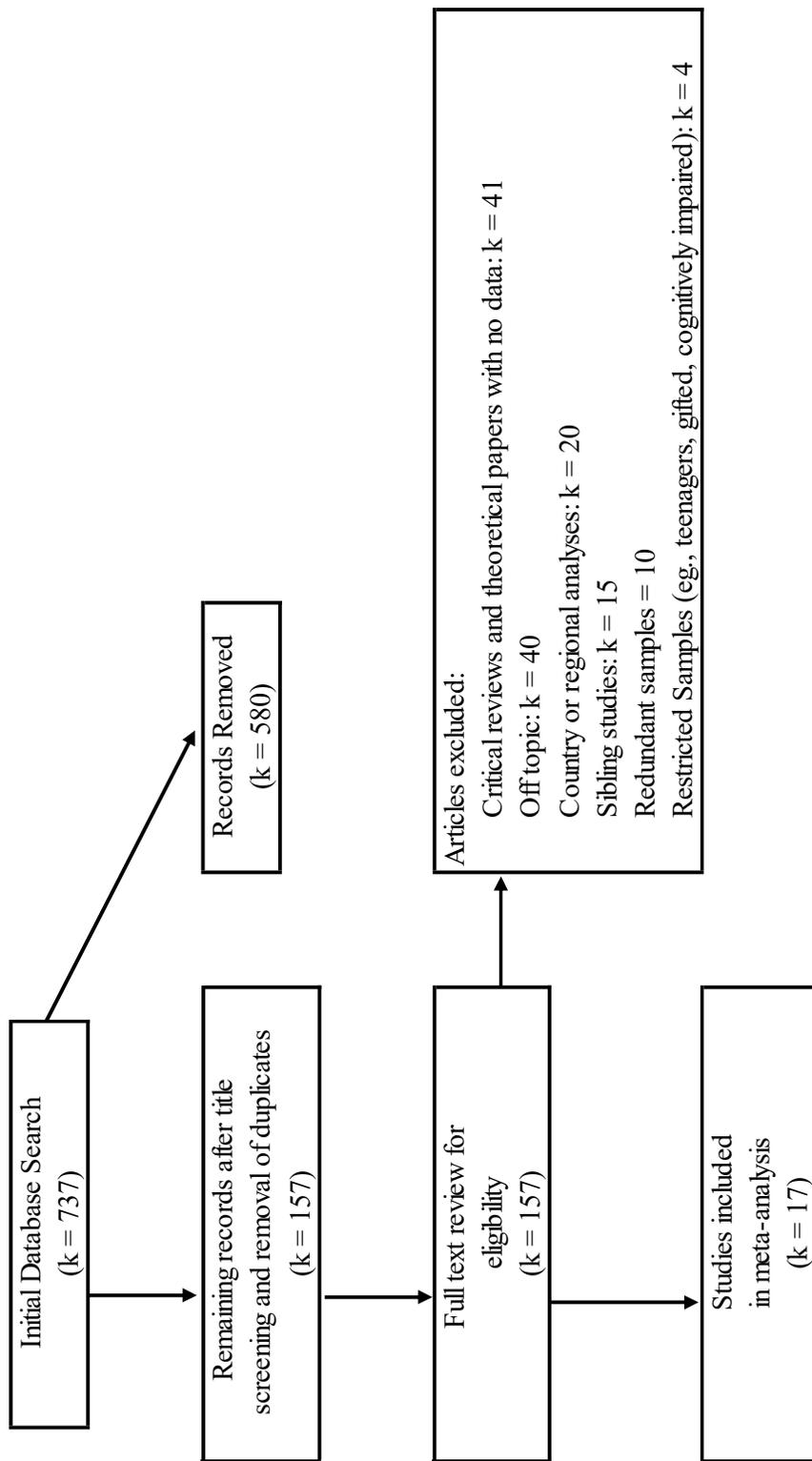


Figure 1: Flowchart for literature selection procedure

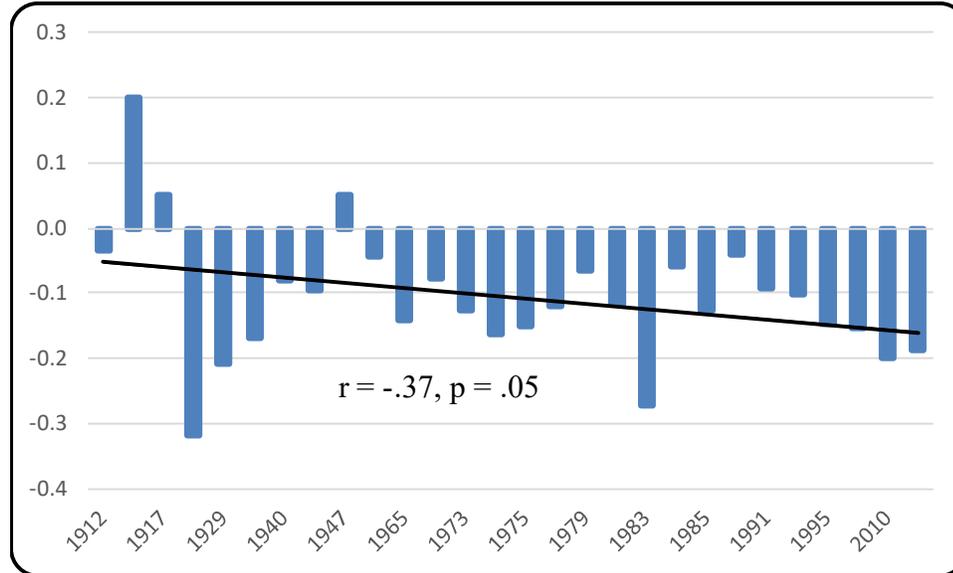


Figure 2: Distribution of effects scaled to year when parent is age 30

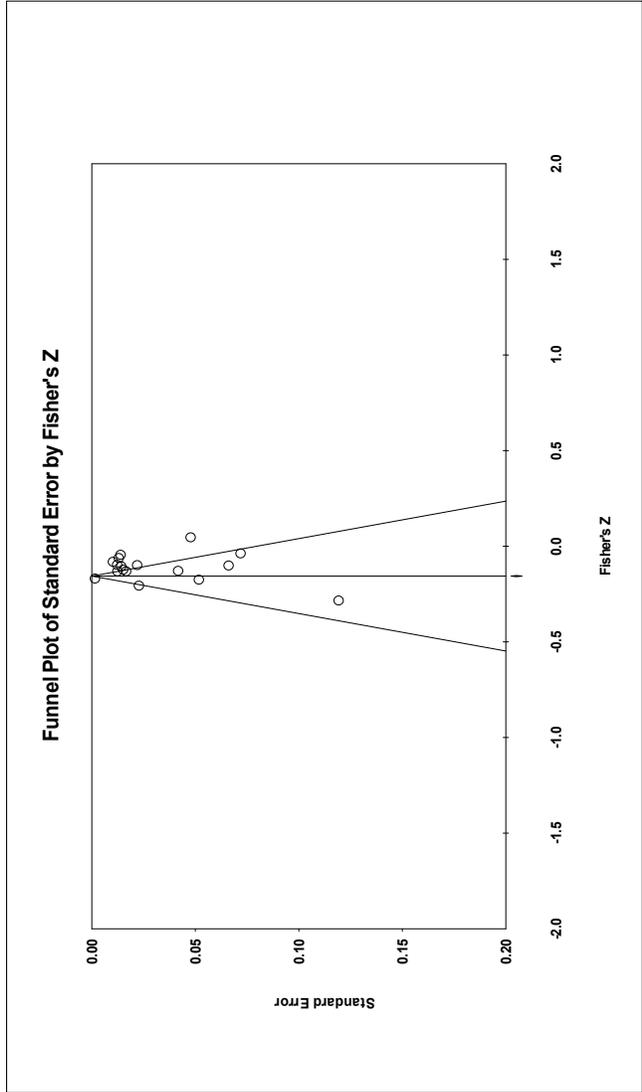


Figure 3: Funnel plot of standard error by Fisher's Z for study level effects between cognitive ability and fertility