Summer 6-9-2014

The Impact of Student Performance on Large-Scale Assessments: A View of Long-Term Health, Career, and Societal Outcomes

Roman Usatin
roman.usatin@student.shu.edu

The Impact of Student Performance on Large-Scale Assessments:

A View of Long-Term Health, Career, and Societal Outcomes

Roman Usatin

Dissertation Committee

Elaine Walker, Ph.D., Mentor
Daniel Gutmore, Ph.D.
Bruce L. Dennis, Ed.D.

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Education

Department of Educational Administration and Supervision

Seton Hall University
May 2014

SETON HALL UNIVERSITY
COLLEGE OF EDUCATION AND HUMAN SERVICES
OFFICE OF GRADUATE STUDIES

APPROVAL FOR SUCCESSFUL DEFENSE

Doctoral Candidate, **Roman Usatin**, has successfully defended and made the required modifications to the text of the doctoral dissertation for the **Ed.D.** during this **Summer Semester 2014**.

DISSERTATION COMMITTEE

Mentor:
Dr. Elaine Walker — Elaine M. Walker — 6/9/14

Mentor:
Dr. Daniel Gutmore — Daniel Gutmore — 6/9/14

Mentor:
Dr. Bruce Dennis — Bruce Dennis — 6/9/14

The mentor and any other committee members who wish to review revisions will sign and date this document only when revisions have been completed. Please return this form to the Office of Graduate Studies, where it will be placed in the candidate's file and submit a copy with your final dissertation to be bound as page number two.

ABSTRACT

This study examined the predictive power of student growth for large-scale assessments on meaningful life outcomes, focusing on the three categories of health, career, and societal involvement. Analysis was conducted using the NELS:88/00 dataset–a longitudinal study that followed a nationally-representative sample of over 12,000 eighth grade students from 1988 to 2000, until the students were 26 years old and entered into the work force. The large-scale assessment variables included math and reading performance in the 1988 cognitive batteries administered by NELS. To gauge growth levels, I generated Student Growth Percentiles (SGP) from tests administered by NELS from 1988 to 1992. Measurable outcomes related to health included binge drinking and cigarette use. Career outcomes included yearly income and job satisfaction. Outcomes related to societal involvement included voting habits, social integration, and the frequency of obtaining information from the outside world.

This quantitative study revealed that student growth on large-scale assessments is meaningfully predictive for three of seven outcome variables: binge drinking, cigarette smoking, and social involvement. Interestingly, I found that students' performance growth on large-scale exams did not yield more desirable outcomes linearly. For occurrences of binge drinking at age 26, only low reading growth increased the likelihood of binge drinking. Typical and high growths in reading performance were statistically identical in reducing binge-drinking occurrences. The use of cigarettes at age 26 saw similar results for both reading and math growth: only low growth on the large-scale assessments increased the

likelihood of the respondent smoking as a young adult. Finally, only respondents who had exhibited typical growth in math performance were more likely to be highly socially involved as young adults.

From the methods and conclusions of this study, I support two major recommendations. First, I recommend that policymakers and school leaders make a habit of collecting longitudinal data along with large-scale assessment results in order to allow researchers and school personnel to investigate long-term program effectiveness. Second, I recommend that a philosophical shift occur among educational researchers in the interest of tracking long-term outcomes that benefit the adult lives of students and society instead of short-lived gains in performance scores and signals.

ACKNOWLEDGMENTS

My first note of thanks goes to my teacher and adviser, Dr. Elaine Walker. Her calm approach and guidance helped this dissertation become lucid, statistically accurate, and meaningful. Dr. Walker encouraged my voice and research interests to come out in this manuscript. At the same time, Dr. Walker provided me with constant feedback that made this dissertation a clear and cohesive text. Truly, Dr. Walker inspires the old adage: Say what you mean and mean what you say.

I would like to thank Dr. Daniel Gutmore, my first professor at Seton Hall University. His wisdom and thoughtfulness is contagious. Dr. Gutmore held debates, asked me to develop my scholarly writing, and synthesized different educational theories. Through his courses and feedback, I have expanded my repertoire to give me the tools to complete this work.

To Dr. Bruce Dennis, I cannot overstate the impact that you have made on my life. I have known Dr. Dennis for seven years, and he is my biggest supporter, my mentor, and my friend. He hired me for my first administrative position; and through watching him work, I have seen all of the leadership theory that I learned come to life. His leadership style is inclusive, his attention is undivided, and he is able to make the complex and intimidating seem like, of course, we can do it! I am forever indebted for his unwavering belief in me throughout all these years.

To my family and friends, I cannot thank you enough for your support through all these years. Specifically, I would like to thank my wife, Cynthia. She is my life partner and my intellectual partner. Through listening to countless hours of my goings on about validity, regressions, and confounding variables, she has helped

me become more coherent and fluid in my words and thoughts. She lifted my spirits when the going was difficult and cheered me on when the going was strong. If this dissertation process was like a tree growing, then Cynthia was like the earth. I can never thank her enough for all of her support.

DEDICATION

I dedicate this dissertation to my parents, Fred and Anna Usatin. Thank you for bringing my brother Vladimir and me into this world and raising us. We crossed the ocean to escape oppression in the USSR, and you were there every day to make everywhere feel like home. I appreciate all of your guidance, support, and love. My heart is bursting with gratitude for all that you have done for me. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Imagine a common American schooling experience. Sitting four feet apart, eighth-grade students are focused with their eyes squarely on a test booklet, bubbling in answers with a number two pencil. After two hours, the proctor collects the exams and sends them to be graded en masse. Three months later, the tests are graded, and students try to make meaning of their performance results. This is where the study takes hold. This study sought to answer the most important question: What is the value of large-scale assessments? Can we use these tests, even at an early age, to determine whether or not they have any effect on health, career, and social integration?

To place this study in its proper context, it is important to understand how large-scale assessments are being used today in the United States. This is discussed in detail in Chapter I, along with the various benefits and the unintended costs of implementing large-scale assessments in a high-stakes manner. In the past decade, high-stakes large-scale assessments have become a major tool for student assessment and accountability in the United States.

The benefits of implementing large-scale assessments are clear and are mandated through policies throughout the United States. Large-scale assessments, by nature, are intended to objectively compare sizable numbers of students from varying backgrounds. This allows policymakers to implement real reform to help those in most need of a high-quality education. These benefits, however, can be outweighed by the drawbacks when large-scale assessments are not implemented carefully. A quick unwinding of the history of this practice reveals several fundamental flaws in the process and in the policies associated with high-stakes large-scale assessments. These limitations show a need for further scrutiny and analysis into the validity of large-scale assessments and the long-term effects on students.

The second chapter reviews the factors that influence student performance on large-scale assessments and what the long-term benefits are of positive student performance on the test. These two chapters set the contextual stage for this study, which sought to analyze large-scale assessments in order to better understand its role in the United States today. This study attempted to provide much needed scrutiny and analysis by examining the predictive power of large-scale assessments on the long-term welfare of test takers in order to retroactively judge the usefulness of the tests administered throughout the country.

History of Large-Scale Assessments

The historical context of large-scale assessments is best summarized through two overarching themes in educational history: (1) the development of the philosophy of education by theorists such as Horace Mann and John Dewey and (2) the history of regulatory policies associated with accountability in education, including the No Child Left Behind Act (NCLB) of 2001. The following discussion critically reviews the history of large-scale assessments through the abovementioned themes in order to elucidate fundamental flaws that appear in the evolution and proliferation of testing for accountability, ultimately presenting the need for further investigation into meaningful outcomes that are associated with a student's performance on these tests.

In the United States, schooling is one of the few common experiences for all citizens. Throughout the tumultuous history of the United States, the educational system evolved from its early stages as a privilege for only those who can afford it to a right of every child in the country (Goodman, 1964). Early theorists, such as Horace Mann (1855), advocated vociferously for universal education on the grounds that a good education can serve as the ultimate equalizer. Those who receive a good education have opportunities presented to them in every aspect of life (Mann, 1855).

Taking this idea further, John Dewey (1903a, 1938) suggested that social reform is the ultimate purpose of education. As an early constructivist thinker, Dewey viewed schools as an incubator for young minds. According to Dewey, the role of the school is to provide authentic learning experiences through real world problems. By having an environment where students tackle meaningful issues, students can indirectly develop all the skills they need to be lifelong learners and productive citizens. In this manner, Dewey advocated for interdisciplinary curriculum and for schools to customize the learning experiences of the child.

Today, the tension between equal access and the customized curriculum is heightened by the contemporary mandates of NCLB (see **Error! Reference source not found.**). The theory of equality behind accountability, in fact, is evident through the very name of the No Child Left Behind Act. The series of accountability reforms imply that the act of measuring students through large-scale assessments can meet Mann's philosophy of equal access to education. At the same time, Dewey (1938) advocated for the constructivist model of teaching and learning whereby teachers facilitate student learning by offering rich learning experiences to help them understand the world. Thus, the education of a person is a lifelong pursuit of knowledge rather than a series of isolated learning activities. As such, this study sought to investigate this tension by examining the degree to which the large-scale assessments predict long-term outcomes. Thus, the usefulness of a test can be measured by how well it relates to long-term life outcomes rather than short-term assessment of skills.

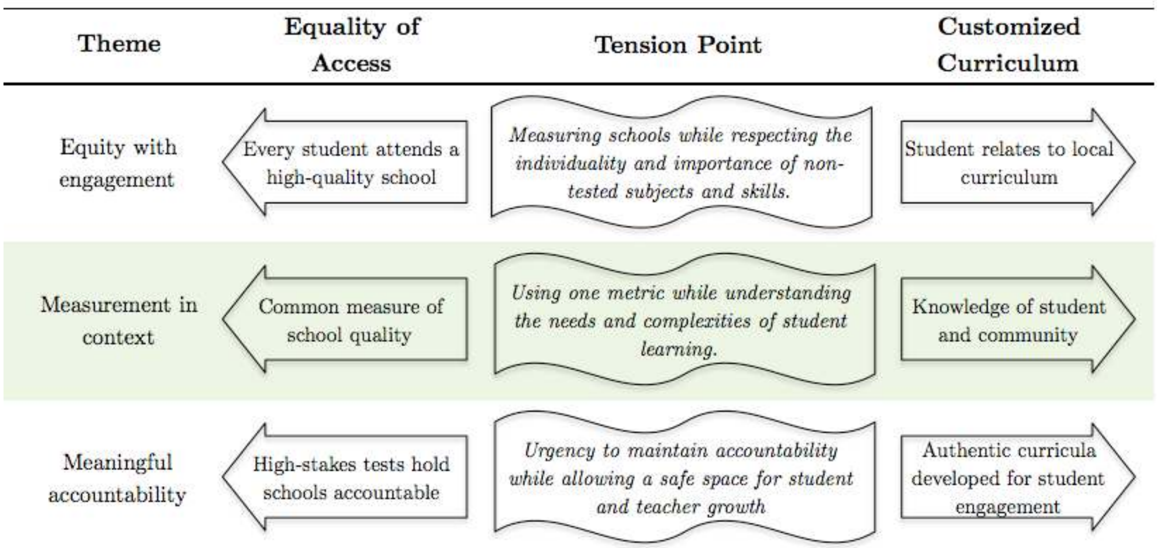


Figure 1. Educational philosophy and tensions created through testing.

The federal government is prohibited from mandating curriculum since the passage of the Elementary and Secondary Education Act (ESEA) of 1965. Nevertheless, two major federal inquiries, identified below, eventually resulted in the passage of a major policy that indirectly influences curriculum nationally via the No Child Left Behind Act (NCLB) of 2001. NCLB mandates accountability to the federal government through high-stakes large-scale assessments. As expected, this mandate resulted in curriculum changes in many schools, districts, and states throughout the United States, with a multiplicity of unintended curriculum effects. The following section discusses the two major federal initiatives and their links to present-day accountability policies as they relate to high-stakes large-scale assessments.

In the history of the United States, two emergent political movements influenced the federal government to intervene in the name of equal access to quality education: the Civil Rights Movement of the 1960s and the standards movement of the 1980s. The policies developed to support the standards movement eventually shifted from suggested standards to large-scale assessments. Consequently, present-day federal policies measure school quality by achievement on large-scale assessments (Ravitch, 2011) without acknowledging the intricacies of what makes a quality school. As such, it is clear that today's federal policies are heavily influencing the education system (Hanushek, 2006; Ravitch, 2011) to move away from Dewey's customized, child-centered, interdisciplinary, and authentic curriculum that is so critical to helping students get the most out of their years in the education system.

The first major federal intervention in education came with the passage of the Civil Rights Act of 1964, which obliged President Lyndon Johnson to commission a study that measured equality of educational opportunity in the United States. At that time, the Coleman Report (Coleman et al., 1966) was the largest study ever commissioned. This first major attempt to survey the educational system found that student outcomes are not solely related to school factors such as funding, but rather that student outcomes are largely associated with demographic and social characteristics of the student. Various subsequent policies based on the report's findings were not effective in raising student outcomes. The report's controversial findings are still a point of debate surrounding school quality (Hanushek, 1998). In the years since the Coleman Report, researchers found that the effects of financial legislation are not likely to have a lasting impact on student outcomes (Hanushek, 2006), causing legislators to seek alternative means to educational equality.

The second series of federal educational reforms can be summarized as the standards movement. This movement started with a commissioned study by President Reagan. This study was controversial at the time, as critics claimed the strong existence of sample bias through, for one, oversampling of the prison population (Carson, Huelskamp, & Woodall, 1993; Tanner, 1998) and that the interpretations of said data were, furthermore, hyperbolic and misleading (Koretz, 2009, p. 86). Acknowledging the poor evidence presented, Ravitch (Ravitch, 2011, pp. 22-31) still makes the case that this report, titled *A Nation at Risk* (National Commission on Excellence in Education, 1983), correctly identifies a need for higher, more equitable standards around the country. This report was the impetus behind the second major shift in educational policy. Furthermore, Ravitch implies that, with the passage of the NCLB Act, the standards movement fell apart and high-stakes testing became a means unto itself—thinly veiled by the guise of raising standards (Ravitch, 2011, p. 30).

The regulatory background of high-stakes testing indicates the federal government's policy-related involvement in education, primarily through NCLB, as a hasty and unsubstantiated response to the urgent demands for equity at the cost of reliable feedback on school quality and of customized education.

The result of NCLB is that the U.S. federal government is increasing its involvement in the local educational sector by favoring schools based on results on large-scale assessments. Since large-scale assessments, without appropriate context, are not accurate enough to be used in a high-stakes manner (Koretz 2002; Koretz 2009), the results sought by the NCLB legislation do not necessarily reflect school quality. The background of accountability through high-stakes testing can be linked to the heart of Horace Mann's educational philosophy, which holds that the vital ambition of schooling is to provide equality of access to quality education. Unfortunately, the large-scale assessment policies that have resulted from NCLB, in their current implementation, do not accurately measure either equality of access or school quality.

Attempts at providing equality of access to quality education for all students are noble quests. However, research has not shown that the avenue of high-stakes large-scale assessments leads to equality of access to quality education; in fact, contemporary federal policies seem to favor the pursuit of equality while neglecting the ideals of the customized education, which, by its nature, cannot be standardized to a paper and pencil exam. Consequently, this study sought to add to the literature what, specifically, large-scale assessments mean to an individual student in the long run.

Background of the Problem

Large-scale assessments have a long history in the United States. Despite their proliferation, the mechanisms of large-scale assessments are intricate and have technical limitations that need to be taken into account before drawing inferences from the data produced. Large-scale assessments, at its roots, attempt to measure school quality in order to ensure that all students have the highest quality education possible. Therefore, large-scale assessments, in and of themselves, are a necessary tool, and can help policymakers identify and measure educational quality on a large scale.

The value of an education goes well beyond the specific competencies measured on a test. *Human and Social Capital Theory* (Becker, 1964; Coleman, 1988) presents the view that the value of an education goes beyond what is learned in school. Becker and Coleman, in separate works, made the argument that there are significant public and private returns to education. In line with human and social capital theory, I measured the most meaningful long-term outcomes to the individual and society using private and public benefits of education.

Using large-scale assessments to measure student achievement growth, I measured the early predictive power of large-scale assessments using three categories: health, societal integration, and economic benefits. *Education production function* (Welch, 1970) is a theory that lends itself to this kind of predictive measurement. The theory states that individuals are able to benefit by going through years of schooling and gaining capital, skills, and knowledge as measured by the three aforementioned categories. These benefits can be *private*, for the individual, or *public*, for society at large. The health of a person is clearly important privately but also allows others to benefit, such as dependents and society at large. The good health of a person lowers the societal cost of healthcare and allows more years of productivity at work. Societal integration is related to such positive outcomes as better health (Christakis & Fowler, 2007), lower crime rates (Owens, 2004), and a closer, fairer community (Helliwell, 2007; Wolfe, 2002). The economic benefits of education are also important to the individual and society at large. For example, if an individual is employed, that individual is more likely to have health insurance and is less likely to rely on public welfare. Indeed, for all of the aforementioned examples and more, it is evident that the private and public returns of education are intrinsically intertwined (Behrman, 1997). However, the limitation of large-scale assessments is that their predictive power applies to the subject tested and not to the wider goals of education, which are to provide long-term benefits to the individual and to the society at large (Dewey, 1903a).

Statement of the Problem

When large-scale assessments are implemented without the understanding of the technical and societal limitations, substantial drawbacks accompanying high-stakes large-scale assessments abound (Amrein-Beardsley et al., 2010; Jacob, 2005; Koretz, 1991, 2009; Tienken, 2011). If these tests can be predictive of long-term outcomes, then this predictability may outweigh some of the inherent technical drawbacks that are associated with large-scale assessments. The problem is that the foundational purposes of education are not aligned with the outcomes sought by large-scale assessments. To truly measure the success of a complex system of education, the benefits to the individual and society must be considered.

Measuring the predictive power of large-scale assessments can allow the major focus of student attention both in and out of school to also be aligned with some of the most important outcomes. Instead of participating in authentic learning opportunities through sports, music, or family bonding, students spend additional time preparing for their exams. If these exams truly measure important skills, then it would mean that these hours preparing for exams are justified to a larger extent than the simple act of improving test scores. Students will typically spend hours studying content and exam-taking skills, both of which are largely unrelated to real-world experiences involving the application of the subject they are learning. However, if the skills measured in these tests were aligned with the larger outcomes of education, it would be visible in the predictive power of the test.

Will the students who demonstrate high achievement on standardized exams fare better in life in comparison to those who demonstrate low achievement? Ironically, there is remarkably little breadth of research into the long-term outcomes of students' lives that would substantiate the robustness of the practice of using large-scale assessments as a tool for measuring educational outcomes.

An exception to the scarcity of research into long-term outcomes of students' lives, are studies linking large-scale assessments to income levels, with the argument that society benefits from more productive citizens who generate higher incomes. It is critical to measure education's benefits through a wider array of outcomes than economics. For example, Dewey (1903b) makes the case that an education should also allow a scholar to engage with society as a productive citizen. In contemporary society, such outcomes encompass voting habits, volunteering habits, excessive drinking or smoking habits, and other indices of health and involvement with society.

Research into the long-term effects of large-scale assessments ought to investigate these measures of worth as well as income earned.

The political climate in the United States now advocates for increased educational accountability through large-scale assessments. There are heavy consequences associated with high-stakes testing, yet the validity of using these tests as a predictive tool was not explored widely in research. Indeed, the most important measures of education, which are directly related to the individual and society, are ignored altogether. The skills taught in a classroom must have value to both the individual student and society at large, and large-scale assessments should also be analyzed for these important outcomes. This study attempted to shed light on the predictive power of large-scale assessments on these most important factors.

Research Questions

The research questions below are concerned with student performance on large-scale assessments and the resulting outcomes to young adults at age 25.

Questions Surrounding Career Outcomes

1. Do students with higher test growth tend to earn more money?
2. Are students with higher test growth more satisfied with their jobs?

Questions Surrounding Health Outcomes

1. Do students with higher test growth tend to smoke less?
2. Are students with higher test growth less likely to be excessive in their use of alcohol?

Questions Surrounding Societal Involvement Outcomes

1. Are students with higher test growth more likely to vote?

2. Are students with higher test growth more likely to spend time socializing with others?
3. Are students with higher test growth more likely to spend time getting information about the outside world?

Study Limitations and Delimitations

This study focused on the long-term predictive value of large-scale assessments. The complex realities of an individual and the society at large place a significant limit on the inherent predictability of large-scale assessments administered in high school. Several factors affect long-term outcomes such as career, health, and societal integration as elucidated in Chapter II of this study. However, it should be noted that the problem of over-reliance on large-scale assessments is inherent not only in the primary and secondary levels of education, but are also evident in higher education and beyond. Furthermore, since an individual leads a life that is far more complex than a single test score indicates, this limitation is further compounded when considering how much of the individual's health is out of his or her own control and not at all related to health-related behaviors.

Subject Delimitation

This study analyzed only English and mathematics large-scale assessment scores and their impact on long-term outcomes. The delimitation is consistent with current trends in the United States that measure student performance only in English and mathematics. Reference to English and mathematics scores only in judging student performance is written into the statute, NCLB, and is the basis of major college admissions tests throughout the country.

Population Delimitation

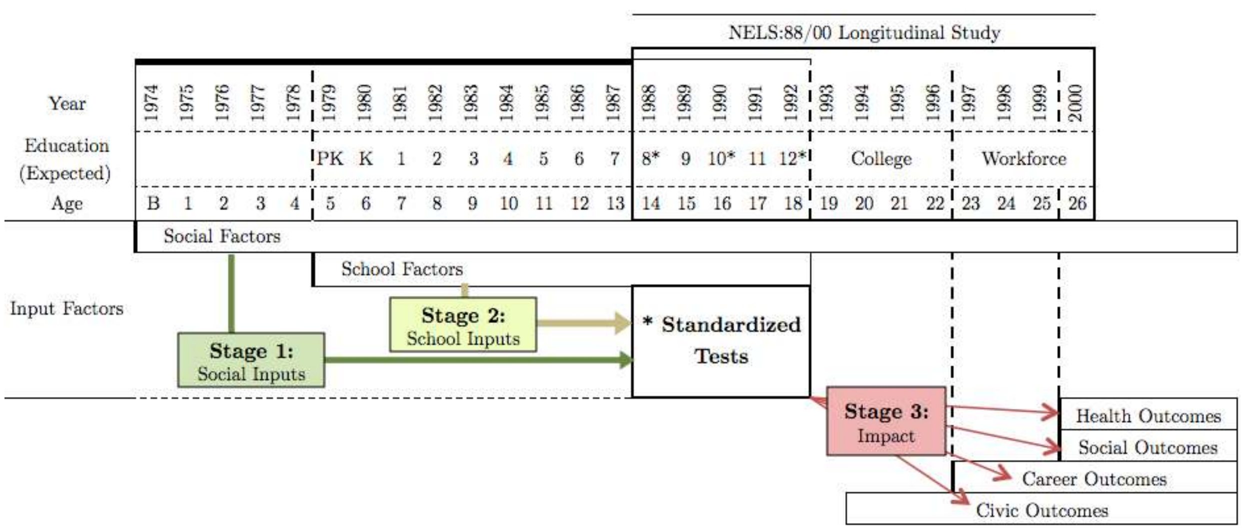
Student performance measures in this study involved the use of Student Growth Percentiles (SGP), as well as status measurements of proficiency. In order to use SGP with sound validity, however, multiple years of test scores per student must be incorporated into calculation of the variable, as each year of scores increases the reliability of the measure. Filtering the participant pool of the National Longitudinal Study of 1988/2000 report to extract students who have multiple years of scores, and therefore a more reliable SGP score, leaves only 7,781 students available for analysis. These students completed all parts of the cognitive batteries during the base year, first follow-up, and second follow-up. Their scores allow growth calculations to be measured consistently using 8th and 10th grade test scores to predict 12th grade growth percentiles. A delimitation of this kind is standard in studies involving SGP since students who have only one year of test scores may be non-randomly different from those who have multiple years of test scores (Rock et al., 1994).

CHAPTER II

REVIEW OF THE LITERATURE

Theoretical Framework

The review of relevant literature, as presented herein, took a three-stage approach. For the first and second stages of the review, I analyzed literature relating to student background and school variables in order to isolate the factors that influence student performance on standardized exams. As Figure 2 shows, there exists a distinction between factors that relate to student performance on standardized exams and the relationship of that performance to health, societal, and career outcomes at age 26. Concurrently, the third and final stage of review utilized the framework of educational function to evaluate the impact of student performance on long-term outcomes.



Note. The figure above juxtaposes the conceptual framework of this literature review with the timeline of the eighth grade cohort of the NELS:88 participants. The lifetime of the participants prior to the completion of the study, and until the age of 26, is included and represented in the upper portion of the chart. The upper portion of the chart also references the expected age and grade of the students. The green and yellow colored boxes, to the right and below the timeline, represent the relationship between the first two input stages, which relate to achievement in large-scale assessments. The red box, nearer to the right, represents the impact stage and the four categorical outcomes of the impact stage as outlined in the lower right-hand corner.

Figure 1. Conceptual framework using (NELS) cohort.

Educational functioning (Hanushek, 1995), as used in this study, tests the predictive power of student performance after controlling for known associated factors. Based on this framework, the benefits of going to school can be viewed via two theories. The first theory holds that school allows a student to gain abilities by learning new skills and making social connections within the school. Thus, a student can build *human and social capital* through schooling. The second theory views enrollment in advanced courses or obtainment of diplomas as a *signal* of qualifications. In other words, under this perspective, obtaining a degree is an external flag that denotes a student's intrinsic ability level.

Human and social capital theory and signaling theory map the measures of student performance to the measures of societal, health, and career outcomes. When human and social capital theory (Becker, 1993; Coleman, 1988) is applied to education, a student learns academic and non-academic skills in school. Non-academic skills include critical thinking, ability to acquire new information, and ability to apply learned information meaningfully to other parts of life through synthesis and transference (Tanner & Tanner, 2006). As such, expected long-term benefits are considered to relate to the greater capacity obtained through schooling. Consequently, students are equipped with more information that will allow them to make healthier choices, regarding the use of cigarettes or alcohol for example. Students will also have acquired skills that aid them in the job market, which will manifest itself through opportunity for work that further allows for higher job satisfaction. Income levels would rise in accordance with the skills that the student obtained through his or her schooling.

In addition to human and social capital theory, signaling theory (Chatterji, Seaman, & Singell, 2003; Hungerford & Solon, 1987; Weiss, 1995) implies that students enter school with predisposed abilities to obtain signals of educational attainment such as completion of a difficult course or a diploma. In this view, students do not directly obtain skills through challenging coursework or increasing levels of education; rather, the fact that they completed those courses indicates that they had the ability to complete the schoolwork in the first place and that the skills and dispositions required to complete the course, such as motivation and resilience, are not taught in school. Suppose that calculus is a course that, on average, is more highly associated with long-term benefits compared to another type of course. Students who complete calculus in high school could potentially benefit from the course in three ways: (1) through skills gained during the course, (2) through inherent traits that led the student to enroll in and complete the course, or (3) through a complex mixture of both avenues.

The benefits of education may be due to a web of environmental, innate, and unobservable factors, as well as observable measures. Although observable measures are the focus of this review of related literature, the possibility of interference from unobservable measures must always remain in the background. For practical purposes, this chapter uses measurable inputs from a student's social and school life to draw upon two major theories in educational functioning: human and social capital theory and signaling theory. Human and social capital theory correlates with evidence that shows that education builds capacity through acquisition of skills and social connections. Signaling theory goes beyond the skills and abilities gained in school and shows evidence of additional returns to school inputs (Bowles & Gintis, 2000).

Social Inputs Related to Student Performance

External measures of a student's background are well known to be the most powerful predictor of student performance, as measured by large-scale assessments. The Coleman Report (Coleman et al., 1966) found that student achievement was intensively related to background factors, such as peers and family, and only loosely related to school factors. These findings are confirmed in a large body of literature (Crum, Ensminger, Ro, & McCord, 1998; Hanushek, 1998; Rivkin, Hanushek, & Kain, 2005). When the studies, cited in this review of literature, controlled for student background, most school-level factors became statistically insignificant and often ensnared with unobserved variables. In fact, studies on school-related inputs found many more variables that are not related to performance than predictive variables.

Unobserved variables frequently influence the measurable societal inputs and outputs that researchers rely on to make inferences. Often, unobserved variables confound the ability of a researcher to identify associations. This is due to undetected connections that alter variable interactions. Even when measures as simple as family income and verbal abilities are used, fundamental exceptions to the trends almost always exist. For example, a study by Chaplain et al. (Caplan, 1992) found that a community of immigrants was able to succeed at high levels, even though the community, on average, did not display typical external variables associated with high performance. Conversely, students in households that promote intellectual environments need to adopt life-skills that are positively correlated with achievement on large-scale assessments, yet the underlying intellectual environment of the student is undetectable through such proxies. It is evident that the link between inputs and outputs is shaped not only by external variables, but also, and profoundly, by underlying characteristics, traits, and dispositions of individuals and their environments.

Not only is it problematic when unobserved factors interact with observed factors but more so when the factors interact with each other. Observable characteristic factors, such as parental income, family structure, neighborhood selection, and school choice, may all interact with one another and possibly with some unobserved factors as well. For example, a person who experiences poverty in one city may have access to additional resources that poverty in another city would not offer (Behrman, 1997, p. 20).

Family Environment Variables

Given the generally positive correlation between test performance and family background, separating the relationship between these two variables can become extremely difficult. Unobserved factors relating to family background can lead to overestimates or underestimates of students' test performance. For example, if students who have favorable family backgrounds are able to afford private tutors to boost their test performance, then the impact of family background on their performance is likely to be overestimated. However, if students from a disadvantaged family background are provided with free services that attempt to improve their test performance, then the impact of family background on their performance may be underestimated. To obtain estimates that are as accurate as possible, it is important to be aware of the method by which the data were collected.

The language that students engage in within their community was found to be related to performance on standardized exams (Gee, 1999). Those students who are surrounded by adults and peers who have advanced language capacity tend to have more intellectually challenging conversations that require sustained attention, recall, ability to synthesize, and switch genres of conversation. This finding suggests that student performance in school and exams is equivalent to learning a new school-oriented language. Gee found that all students are extremely capable of learning languages. Furthermore, the type of language that is necessary to succeed in school is learned socially. As a result, students from environments that sustain intellectually stimulating conversations tend to show enhanced performance in school and on exams.

Caplan et al. (1991) conducted a series of studies that linked student achievement with the home environment. In these studies, the researchers found that a low-income group of immigrant refugees succeeded at high levels in terms of student performance because of their home environment. More specifically, Caplan et al. found that the external measurements of the participants countered many notions of what are generally regarded as disadvantages. For example, the participants lived in low-income neighborhoods, had large families, did not speak much English, and lived near poverty levels. These factors are known to be major hurdles for student performance, yet the families were able to neutralize and perhaps turn the perceived disadvantages into advantages (Caplan, 1992, p. 399).

The main advantage to a home environment that values education is that knowledge is frequently transferred from school to home, which makes schooling more valuable and relevant for the students (Caplan, 1992). In their qualitative study, Caplan et al. found that the immigrant families supported academics in the home through various activities, which extended their schools' impact in meaningful ways. Parents assumed most chores and other household responsibilities so that students were able to spend more time on their schoolwork at home. Additionally, students in these households were encouraged to help their siblings within the home. Furthermore, almost half of the families who participated in the study read aloud to their children, which was associated with a gain of 5% for student GPA. Reading aloud and homework support in those families blurred the line between school and home, allowing the children to transfer skills to multiple domains.

Egalitarianism and role sharing are also meaningful drivers of student performance. Homes that agreed with the statement "A wife should always do as her husband wishes" performed 16% worse in school than those who disagreed with the statement (Caplan, 1992, p. 400). Furthermore, Caplan et al. found that when fathers help with chores around the house, student performance is, on average, 13% higher. This however is not limited to parents alone. When the family views gender roles as equivalent and have similar expectations for all their children, there is a 10% increase in performance. Although this study is extremely limited in its applicability to other populations, it demonstrates that observed variables of social characteristics do not impact all students in the same way.

On the other hand, there is evidence that the performance gap between those who have advantageous home lives and those who do not begins at the age of 7-9 months, growing substantially in the first few years of life (Hart & Risley, 2003). In a major study, Hart and Risley interviewed 42 families for one-hour conversations during the duration of two and a half years. Their study found that the work status of the parents was a major factor in the child's language acquisition. Children of unemployed parents hear an average of about 600 unique words per hour, while children of working-class parents hear an average of about 1,200 words per hour. Children of higher-income families, in sharp contrast, hear an average of about 2,100 words in an hour. Upon extrapolating these numbers, the authors found that differences in words heard multiplied into the millions by the age of four. Furthermore, the higher-income families were found to use terms of encouragement when speaking to their children 32 times in an hour, while the working-class families used terms of encouragement within the same context 12 times in an hour; and finally, the unemployed families used terms of encouragement only 2 times in an hour. This is a major difference that was found over thousands of observations. Indeed, this study shows that family environments can offer marked advantages or disadvantages to student performance over the long run. Although the case of each individual family varies greatly, on average, language acquisition has been found to be substantially linked to family background.

The size of a family is also a critical variable in student performance. The number of children in a family is inversely related to the amount of time a family can devote to each child. Hill and Stafford (1974) advanced the concept of dividing parental time into budgetary terms, which include *public time* and *private time*. Public time indicates general activities from which many children benefit, such as a trip, an intellectual atmosphere, or using advanced language at home. Private time is nontransferable to other children, such as individual help with assignments and private counseling. Larger families tend to have less private time, which is generally considered more valuable than public time. As such, larger families tend to have lower test scores. Being first born is advantageous only in the sense that you are more likely to be in a small family (Hanushek, 1992).

Parental Resource Variables

Variables regarding parental characteristics are commonly controlled for when measuring student outcomes (Haveman & Wolfe, 1995). However, external measures of income and education may not necessarily represent linear relationships. Haveman and Wolfe found that large positive changes in family income levels increase student attainment very minimally. Specifically, in their literature review, Haveman and Wolfe found evidence from a synthesis of multiple studies that adding 10% to a family's income raises educational attainment by only 1%. A family's income is a crude measure of resources available to the child. The reason is that although a family's income is directly measureable, the amount of time and resources transferred to the child were not captured by any of the studies that Haveman and Wolfe reviewed.

In the same respect, the time parents spend on homework with their child may not necessarily indicate a positive trend for their child's performance. For example, parents who spend prolonged time with their child may be doing so because their child is especially resistant to homework or has experienced prolonged absence or suspension. In this case, the extra time spent with children may not influence overall performance gains significantly.

When it comes to family income, the level of earnings in the early years of a child's life considerably impacts performance in school throughout the rest of that child's life. Duncan et al. (1998) found that for low-income families in the first five years of their children's life, an increase of an average of $10,000 yearly nearly triples the odds of their children completing high school and adds an additional year of their children's overall schooling. In the same study, Duncan et al. (1998) found that the benefit of increased parental income during early childhood is nonlinear. At the upper margins of family income, additional earnings do not impact student performance. However, the benefits of raising the income level of a family in poverty can substantially increase student performance. This evidence supports a twofold implication: (1) there is a curvilinear relationship for parental income and performance and (2) the child's early experiences in life have substantial impacts on their later achievement and performance in school.

Issues of entanglement between income and family factors such as parental education, marital status, and neighborhood characteristics limit a majority of the literature around family income and educational attainment (Brooks-Gunn & Duncan, 1997). For example, the mother's educational attainment is important in relation to the student's performance to a point of diminishing returns (Haveman & Wolfe, 1995). The returns for children are relatively small after a mother attains some post-secondary education. Nevertheless, the level of education that the mother achieves is entangled with family structure, subsequent income levels, and choice of neighborhood (Jaffe, Eisenbach, Neumark, & Manor, 2006), which do, in turn, influence the child's educational performance. For example, if the family structure is such that both parents are employed, one of the parents may find the opportunity to invest in furthering his or her education. If a single parent works full time to support the household, however, that parent may be less inclined to devote time to attaining a college degree due to the threat of lost wages in the immediate term. At a result, there are confounding positive associations between the variables of family structure, education, and income. This web of associations limits independent interpretation of those variables.

Family-Related Adversity and Long-Term Outcomes

Advantageous family backgrounds are not necessarily related to substantial long-term outcomes. The current literature in psychology states that those who are most successful in their field do not always come from advantageous family backgrounds (Simonton, 2012). Instead, those who make substantial impacts in their domain are subject to a 10-year rule. The rule suggests that regardless of individual characteristics, true impact in any domain comes after 10 years of deliberate practice. Furthermore, Simonton (1993) found that backgrounds that are excessively advantaged produced students who perform better in the short run, but these performance advantages may not lead to substantial impacts on a field of study or work in the long term.

Events that are commonly perceived as disadvantageous are counter-intuitively common amongst the most successful individuals in their fields. Indeed, there is evidence that the experience of adversity is regularly overcome, regardless of background, to allow the person to succeed and impact the world at the highest levels. In fact, there is mounting evidence that prominent individuals have experienced degrees of adversity much higher than the rest of the population. In the book *David and Goliath*, Gladwell (2013) argues that what is commonly perceived as a major disadvantage, the loss of a parent, is also a common factor amongst the leaders of the United States and England. Iremonger (1970) studied the family backgrounds of English prime ministers and found that a full two-thirds of the prime ministers lost a parent before the age of 16, which is far greater than the average population for that time. In the same vein, Gladwell (2013) found a similar pattern where 27% of U.S. presidents lost a parent at a young age.

In another study involving 699 prominent individuals, Eisenstadt (1978, p. 214) found that 25% of them had lost a parent before the age of 10 and 45% had lost a parent before the age of 20. Although the estimations of expected death rates of a parent vary greatly by century, location and other factors, Eisenstadt estimates that the disproportionate number of prominent individuals who had experienced parental loss is universally significant to the average population of their respective countries. These findings indicate that perceived adversity could be neutralized via individual motivation to succeed in conjunction with abilities and opportunities.

Poverty and Student Performance

Poverty is the most direct predictive variable for low student achievement (Chapman, Laird, & KewalRamani, 2013; Ravitch, 2011). Indeed, when controlling for every other factor linked to low student performance, children who live in poverty were found consistently to show lower performance on large-scale assessments (Smith, 1997). Consequently, poverty in the early years of a child's life may create a gap in test performance that can relate to misplacement in special education classes due to test performance, but not to other factors. Thus, membership in these classes may mask the true ability levels of poverty-stricken students by the cover of limited feedback from large-scale assessments scores. While some students rise above such adversity, many do not (Ravitch, 2013).

On average, students living in poverty tend to face a disproportionate myriad of burdens that may affect long-term achievement on large-scale assessments. Students living in poverty are at higher risk of not seeing a dentist, of not being read to by their parents, of not having access to three nutritious meals a day, and of often experiencing burdensome circumstances that are generally associated with the basic needs of students (Ravitch, 2013). Children of poor families tend to be absent more frequently from school, which is a major indicator of dropout risk and low performance on tests (Rumberger, 2011). There is further evidence that parents who live in poverty work longer hours, thus sacrificing potentially valuable time with their children without gaining much relative earnings in exchange for their sacrificed time (Ryan, Fauth, & Brooks-Gunn, 2006).

There are indications that children living in poverty may have less access to print materials and age-appropriate games (Bradley, Corwyn, McAdoo, & García Coll, 2001). This is particularly unfortunate since students who are exposed to reading and to intellectually stimulating games early in their development gain tremendous performance benefits that include early reading and advanced interactions with adults and peers (Bradley & Corwyn, 2002; Duncan et al., 1998).

Long-term poverty is especially associated with lower student achievement. After controlling for various important family variables, a study evaluating the National Longitudinal Survey of Youth (NLSY) found that the number of years students experience poverty drastically lowers test performance for mathematics and reading (Korenman, Miller, & Sjaastad, 1995). That is, long-term poverty impacts the student even when controlling for family structure, mother's education, and academic ability. This implies that the effects of poverty may be cumulative and that the obstacles a student faces may compound the effects of absenteeism and other factors associated with poverty, which, in turn, create obstacles to student achievement.

Peer-Related Variables

The influence of peers is interwoven with many other unobserved variables. Since the neighborhood in which a student lives is not random, peer influence is associated with parental choice. For example, a family may choose to live in a neighborhood based on income, job location, or school choice. Once in a neighborhood, parents may choose to send their children to public school, private school, or alternative schools. Once enrolled, parents may advocate for a particular teacher or influence their children's course selection. Thus, the interwoven fabric of neighborhood and choice must be understood in order to articulate a relationship between a student and the influence of his or her peers.

Other unobserved variables that intermingle with peer related influences include the non-random placement decisions made by educators, where students are commonly sorted into classes based on behavior, performance, or interests. These factors further compound the question of whether peers influence one another in school or whether their perceived influence is the result of related unobserved variables so much so that the influence of peers may not be measurable directly at all.

Effects of Peers in the Classroom

In the classroom, peers can enhance or impede performance in various ways. The characteristics of the social network of a student can affect that student's performance on tests. Hoxby (2000) found that student performance is positively associated with peer exam performance. In other words, when students perform better on an exam, their peers can be expected to follow with a similarly raised performance score. Interestingly, regardless of peer performance, students, both male and female, perform better on reading and mathematics tests when they are in a class where the majority of students are female (Hoxby, 2000). Although decoupling peer effects from other choice variables is challenging, there seems to be evidence of positive overall effects in having high-performing peers.

For teenagers, being assigned to a high-ability class is also associated with increases in performance (Vardardottir, 2013). There is evidence that high-performing peers are associated with assisting the learning process through motivation (Eisenkopf, 2010). However, research also shows that students who perform at high levels perform worse when grouped heterogeneously with lower-performing students (Feldhusen & Moon, 1992; Shields, 1996). Also, when lower-performing peers are in a classroom, higher-performing peers tend to lose enthusiasm and have a negative outlook towards school.

Confounding the problem, the negative impact of homogeneous grouping is felt significantly more for lower-performing students. The social impacts of grouping low-performing students together are severe. Shwartz (1981) found that students in low-ability tracks have more classroom disruptions, off-task behavior, and a negative association with the group label. Also, social skills such as asking for help or collaborating with others is lessened if the class is tracked for low-performers (Wilczenski, Bontrager, Ventrone, & Correia, 2001). The act of tracking essentially labels the student internally and externally, making it challenging for the student to move to a higher track. The benefits of tracking and grouping based on performance are prominently outweighed by social disadvantages in the classrooms tracked for low-performing students.

Neighborhood and Community Variables

Extracting the influence of a student's neighborhood from the student's family background and other social factors is challenging, given the interconnected nature of parental income, education, and social capital. In a review of the literature surrounding neighborhood impacts upon student attainment, Haveman and Wolf (1995) found that all the studies they reviewed were unable to successfully separate family and neighborhood characteristics. For example, those who live in higher-income neighborhoods tend to be able to put forth political pressures for better school resources and may already have a higher probability of settling near better schools.

However, if the political pressures are concerned with equity, those who live in lower-income neighborhoods may have a higher probability of going to a high-quality school (Behrman, 1997, p. 34). Furthermore, it is well known that the income of a family is a significant factor in the choice of neighborhood. As such, the research shows that neighborhood variables are insignificant when controlling for family background (Haveman & Wolfe, 1995). However, there is also evidence that when a neighborhood's qualities are far below average, the student is negatively impacted (Haveman & Wolfe, 1995).

In addition to home and community factors, the classroom itself is often the place where a student learns language. Gee (1999) found that the social learning of English in school is critical to successful complex language acquisition, which facilitates high-level performance on standardized exams. All children are born with excellent language-acquisition skills, and students who do not perform well on large-scale assessments tend to lack the specific verbal skills tied to school practices and school-based knowledge (Gee, 1999). This finding suggests that to perform well on standardized exams, students need a specific social language that is taught in schools and expressed by adults and peers with advanced language capabilities.

School Inputs Related to Student Performance

The landmark study referred to as the Coleman Report (Coleman et al., 1966) found that school resources do not account for much variance in student performance. Instead, the background variables of a student were found to be of critical importance to student performance as measured by test performance and graduation rates. This finding, however, should not be taken to mean that schools do not matter. There is a large body of research, much of which is presented in this section, suggesting that the measures of school quality are consistently confounded by the difficulty of measuring social sorting (Lankford, Loeb, & Wyckoff, 2002), student support (Ashton, 1985), and teacher quality. There is no doubt that a good teacher can change a student's life and that the student may never know it because the effects of the teacher may not be felt until years after the student completed his or her course (Lazear, 2003).

School Leadership

There is evidence that high-quality school leaders, particularly principals, can increase student performance on large scales primarily through hiring quality teachers. Branch et al. (2013) found that, particularly in high-poverty schools, principals are a critical factor of attracting and retaining high-quality teachers. They found that a top-performing principal could increase average performance of the entire school by 0.11 SD (standard deviations), while a principal at the bottom-performing quartile can decrease the average performance of the school by 0.15 SD. These findings, like any attempt at associating student performance with the skills of a professional, are confounded by various measures. For example, Branch et al. (2013) found that for the most disadvantaged schools, the top tenth percentile of principals are associated with only 0.07 SD gain, while the lowest performing principals lose an average of 0.2 SD. Therefore, the measures of effectiveness of principals by association with student performance are incomplete and highly influenced by the characteristics of the population that the school serves.

Confounding the measurement of a principal's effectiveness are two main issues. The first is the confounding issue of school and student characteristics as reflected in the errors associated with aggregated measures of student performance, as discussed in the large-scale assessments section of this literature review. The second issue involves the related effect of the characteristics of the student population. The ceiling effect may cap gains that principals see if they enter an already high-performing school. When controlling for these factors, Branch et al. (2013) found a strong impact of principal quality on student performance. For one SD of improvement in principal quality, student performance also increases by 0.05 SD. Even though studies relating to teacher quality have shown a relatively larger impact on student performance of about 0.1 SD of student performance improvement per standard deviation of teacher quality (Hanushek, Rivkin, Figlio, & Jacob, 2010), principal quality reached a broader range of impact. Principal quality affects every student in the school, while teacher quality affects selected classes of students. Studies analyzing the effects of school leadership generally show positive associations with effective leadership and student performance. At the same time, however, these studies were unable to separate the effects of principal leadership from teacher quality and student population adequately.

Class Size

A large body of research speculating upon class size exists in the United States. Hanushek (1997) summarized 277 study estimates of the impact of class size and found that class size does not significantly affect student performance. Krueger (2003) reanalyzed Hanushek's findings by weighing the study estimates differently. Krueger (2003) found that Hanushek gave some studies significantly more weight by virtue of the quantity of estimates. Further, two studies, compiled by the same authors with 24 estimates accounting for 17% of Hanushek's estimates, had methodological issues that should have excluded those studies from review in the first place. Upon recalculation, Krueger found that the studies with fewer than seven estimates showed class size to be related positively with student performance. In a follow-up paper, Hanushek (2003) criticized Krueger's use of weights for two reasons. First, many of the studies with just one estimate are considered to be of lower quality, yet they ended up having the most weight in Krueger's study. Second, even when using the alternative weights, Krueger's estimates of the benefits were too small to be significant. The end result is that the argument over class size and student performance continues to be an ongoing debate in the United States and all over the world, with compelling evidence for lowering class size, while, at the same time, taking into account the economics and individual situation of a school or school district.

Literature in favor of smaller class sizes argues that smaller classes allow teachers to devote more time to each student, which allows teachers to make more of an individual impact (Achilles, 1999). Although there is no literature in favor of larger class sizes per se, the stance against prioritizing small class sizes at all costs is compelling. Hanushek (1999) argues that, in certain specific cases, smaller class sizes can make a positive difference, but legislating smaller class sizes without weighing the benefits of other economic investments of the schools, such as raising salaries or promoting teacher quality, is an inefficient use of funds allocated to the schools. Hanushek's arguments are limited, however, in practice since there is no evidence that the constraints he presents to lowering class size exist in schools widely.

For disadvantaged subgroups, however, there is significant evidence that small class sizes early in academic life can greatly influence later college aspirations. In a study of the K-3 STAR program in Tennessee, Krueger and Whitemore (2001) found that students in small classes were twice as likely to take a college entrance exam when compared to their subgroup peers across the country. Furthermore, in another study, Krueger (1999) found that small class sizes out-rightly increases student performance for grades K-3. This finding is consistent with similar studies of the Tennessee STAR program (Finn & Achilles, 1990). In Israel, Angrist and Lavy (1999) used Miamonides' rule of no more than 40 students per grade to conduct a natural discontinuity study. They found that, like Finn and Achillies (1990) and Kreuger (2001), smaller class sizes improve student performance, specifically for those who are members of disadvantaged subgroups.

Contradictory studies, on the other hand, found that class sizes make no significant difference between disadvantaged subgroups of students and the rest of the population (Hanushek, 2003; Rivkin et al., 2005). Further, although class size is related to achievement, a study found that the relationship fades and reverses as students pass through the middle school grades of six and above (Rivkin et al., 2005). These contradictory findings indicate that the issue of class size does not have a one-size-fits-all solution. Further, the discussion of class size, when measured by test scores and graduation rates, significantly discounts the pedagogical and social arguments for class size. Instead, Hanushek and Rivkin basically argue that a high-quality teacher should have as many students as possible. Although there are merits to this view, it lacks the pragmatic issues that schools face every day, such as the variability of teacher and student quality. Indeed, there are specific cases of schools where economics favors smaller class size, while there are other cases of schools that cannot accommodate the reduction of class size at all. Economics and student performance are not the only issues that determine class size. Size is also an issue of the physical space within the school and the pool of available high-quality teachers for extended subject areas. That said, research on effective class size is an ongoing and important field of study that will help school leaders and policymakers understand the complexities of class size.

Social Capital

Schools provide a critical venue for gaining social capital. Involving parents in the school community is associated with increased student performance. Putnam (1995) found that there is a strong link between parental involvement in schools and academic performance of the students in that school. In fact, Putnam finds that social networks, created by parents or the students themselves, are a major predictor of academic achievement. This finding is not surprising and supports the previous literature presented about the benefits of supportive home environments. Since family involvement is critical in extending the benefits of school through transference of knowledge (Caplan, 1992), having a parent involved in the school can help blur the line of home and school, forming a larger academic sphere.

For students, teachers are a chance to make social connections to adults in safe and healthy ways. A study using the data from NELS:88 found evidence that teachers are one of the main sources of social capital for students (Croninger & Lee, 2001). In his study, Croninger et al. found that teachers can significantly reduce the risk factors associated with leaving school and can improve academic performance, particularly for students from disadvantaged backgrounds. Indeed, teachers can be one of the most trusted adults and role models for students. The teacher can be an adviser to the student, help the student through difficult times, and dramatically influence the feeling of belonging within that student's school environment—all of which help students perform at their best.

Teacher-Related Variables

Research shows that, of all school-level variables, teacher quality matters most to student performance (Darling-Hammond, 1999; Hanushek, 1971, 2009a; Lazear, 2003; Rivkin et al., 2005; Rockoff, 2004; Sanders, Wright, & Horn, 1997). Isolating teacher influence from confounding factors is challenging (Rockoff, 2004) however. In a review of the literature of the most frequently measured school inputs, a majority of studies found no positive association with school inputs and student performance. Specifically, over 90% of the studies found no positive relationship to the school inputs of teacher/student ratio, teacher education, administrative variables, and facilities. Furthermore, 78% of studies showed no positive effects from teacher experience, 84% found no positive effects from teacher salary, and 80% found no positive effects from expenditures per student. These results are further discussed in the review below.

Teacher expectations.

Student performance can be positively influenced when teachers have high expectations for their performance. Students are astute observers of teacher expectations and tend to fulfill the overt or hidden expectations that a teacher holds for them (Weinstein, 1989). Indeed, student performance is highly related to the teacher's perception of their abilities (Rist, 1970; Roeser, Eccles, & Sameroff, 1998). Specifically, teachers who hold low expectations for selected students are more likely to offer less autonomy and more tedious work to those students than to students for whom they have high expectations (Roeser et al., 1998).

Carol Dweck, an expert in motivational theory, wrote extensively about the motivational theory of students meeting teacher expectations, a concept known as the self-fulfilling prophecy (Dweck, 1986, 2007, 2008). Students listen to and watch teachers carefully, even when they do not appear to be doing so. These students are influenced by the overt and subtle messages that teachers give daily in their classroom. The self-fulfilling prophecy, however, works both ways. Teachers with low opinions of student ability will have students who meet those expectations, while teachers who have high opinions of student ability will have students who perform at high levels. The teachers who believe that they can affect all of their students are more likely to reach out to their most challenging students and to dramatically improve the performance of those students (Ashton, 1985). Simultaneously, there is significant evidence of negative impact from low teacher expectations on subsequent student performance (Brophy & Good, 1974; Jussim, Eccles, & Madon, 1996; Parsons, Kaczala, & Meece, 1982).

Wilczenski (2001) argues that the focus of an effective teacher should be mastery of skills, rather than student perceptions. That is, the body of research implies that teachers who are most effective in raising student achievement hold their students to high expectations, relate their students to the class material and zeros their focus upon subject mastery, while supporting the students' socio-emotional needs as those needs relate to performance in the classroom.

Measures of teacher quality.

There is a large body of research that finds that teacher quality can improve student performance dramatically. However, the evidence that the studies in this section present are severely limited by two significant factors. First, there are consistent confounding effects of unobserved background and sorting variables associated with the measures of teacher quality (Hanushek, 1971; Rivkin et al., 2005; Sanders et al., 1997). Second, there is significant evidence warning that value-added measures are highly error-prone (Hanushek, 1971; Koretz, McCaffrey, & Hamilton, 2001; Newton, Darling-Hammond, Haertel, & Thomas, 2010; Olson, 2001; Schafer et al., 2012; Schochet & Chiang, 2010; Scott, Rock, Pollack, & Ingels, 1995) and only measure teachers of tested subjects like English and mathematics. Consequently, the following paragraphs hinge on the limited notion that high quality teachers raise student performance.

The perception that teacher quality matters is well grounded in the reality of students. The school experience for a student involves sitting in the classroom of teachers for an entire school day. In each class, the experience is necessarily different depending on the environment that the teacher helps create for the students. Clearly, teacher quality does matter. However, the quality cannot be unobservable by external measures without visiting the classroom (Sanders et al., 1997).

The classic measures of teacher quality such as experience, certification, and educational attainment increase the confounding effects of unobserved sorting and selection mechanisms. In a study of elementary schools in North Carolina, Clotfelter et al. (2006) found strong evidence of teacher sorting between schools and within schools based on the three aforementioned measures of teacher quality. That is, teachers with more experience and education are more likely to work in higher-performing schools. Furthermore, sorting is evident within schools. Teachers that have signals of quality tend to resist being assigned low-performing students. Higher-performing students are assigned to teachers with these flags with relative consistency throughout all of the studies on teacher quality. Student sorting could be due to reputation, perceived qualifications, or advocating for particular student or course assignments. When controlling for this non-random student assignment, the impact of experience and education are not significant. Overall, the literature centering on student performance indicates that teachers with more experience are able to get better students assigned to them, which makes the confounding factors of sorting difficult to separate from the measures of teacher quality.

On the other hand, the body of research that attempts to link teacher quality to student performance is substantial. This body of literature tends to overstate the degree of teacher impact when considering the significant limits of value-added measures in the studies cited above and in the later section of this study concerning issues relating to large-scale assessments. In studies that attempt to control for background and sorting mechanisms, teacher quality was found to be positively associated with student performance. In a more recent study, Hanushek et al. (2010) found that 1 SD in teacher quality can increase student performance on math by 0.13

SD and reading by 0.17 SD. These findings agree with other research that shows significant student performance gains related to teacher quality. Rockoff found even larger gains for two elementary schools in New Jersey. For each standard deviation of teacher quality, a student potentially gains 0.2 SD on reading and 0.24 SD on mathematics (Rockoff, 2004). In another experiment, elementary school teachers in Los Angles were randomly assigned groups of students. Kane et al. (2008) found a similar result: 0.22 SD gains in math and 0.18 SD gains in reading as estimations of teacher quality. Although the measures of teacher quality are not consistently measured between the studies, after normalizing the impact of teacher quality, the findings are very similar across these and other recent studies of teacher impact on student performance. In his study on teacher quality in Texas, Hanushek (2005, p. 280) addresses some entanglement issues of teacher selection. Acknowledging the limitations surrounding the linkage of student performance and teacher quality, Hanushek found about a 0.15 SD increase in student performance for each one SD of teacher quality (Hanushek, 2005). In another study involving students in Grades 2 through 6 in Gary, Indiana, Hanushek (1992, p. 107) found that the difference between having an effective teacher versus having an ineffective teacher could be as much as one grade level in student performance. Consequently, when students have a series of teachers of high or low quality, the compounded effect can be very large in terms of student performance. With all of the teacher quality research in mind, caution should be exercised when attempting to interpret this body of research. Although mathematically attractive to quantify effective teachers, in practice, teacher quality is not measured in standard deviations. Therefore, the implications of this body of research is interesting to note, but not applicable to schools or policy.

Furthermore, although the evidence presented about teacher quality is fairly compelling, the significant limitations of measurement are not explicitly addressed in the conclusions of the aforementioned studies. As such, the majority of the literature does place the hard numbers they find in the proper context. This is most clearly seen in a theory called *teacher deselection*. Hanushek (2009a) uses this compounding effect of teacher quality to theorize that, by permanently replacing the lowest-performing teachers in the year 2010, the education system would lead to returns of $200 billion for the U.S. GDP by the year 2030 and that students would see dramatic gains in performance. He indicates that the lowest 10% of teachers are damaging students irreparably and makes the case that low-quality teachers should be removed with haste and in a consistent manner. There is an issue with this conclusion by Hanushek for two reasons. Student performance is a poor proxy for teacher quality (Baker et al., 2010) and, according to Fullan (2008), an expert on change in large organizations, it is unwise to adopt a management style similar to Jack Welch's vitality curve. The vitality curve is the famous management system that Jack Welch used during his time as the CEO of General Electric, where 10% of his managers were fired annually. Many companies started adopting this policy in what is called the halo effect, which is a management delusion whereby a company copies the external traits of an organization after it is successful (Fullan, 2008). For example, if a CEO of Company A has a controversial personality which leads the company to gain more popularity, it may not necessarily be wise for another CEO to adopt this trait, imagining that this is the cause of the success Company A gained. As such, the halo effect and the dubious connection between student performance and teacher quality should inspire policymakers to proceed with caution when raveling student performance measures with measures of teacher quality.

Characteristics of teacher quality.

Effective instruction, classroom management, and social skills of a teacher are challenging to evaluate externally through aggregate. Sometimes, a teacher connects instruction for certain students and not others, while classroom management can be an issue in some classes though it may not be an issue in other classes for that same teacher (Glazerman, Mayer, & Decker, 2006, p. 92). Indeed, teacher quality does matter and can potentially account for tremendous learning growth in students. However, to find out what makes a teacher successful, visiting the classroom is profoundly important; one cannot observe teacher quality through external measures alone. Indeed, in a study controlling for external measures of teacher effectiveness, 97% of the variance in student performance is attributed to unobservable teacher characteristics (Goldhaber & Anthony, 2007, p. 36). In fact, a study using the NELS:88 data found that much of the quality of a teacher cannot be observed until years after the student completes the course (Lazear, 2003).

Teacher experience.

Many researchers have attempted to examine whether or not teacher experience matters. One of the most-cited reviews of the literature found no evidence of teacher experience positively associated with student performance (Hanushek, 1986). Specifically, in a later study of school expenditures, Hanushek (1989) found that of 140 studies that used teacher experience, less than a third found positive associations with student performance. This finding indicates that the connections between teacher quality and experience are generally not related to student performance.

There is evidence of a sharp learning curve for those entering the teaching profession. In a study of Grades 4-7 in Texas, Rivkin (2005) found that beyond the first two years, teaching experience is not associated with an increase in student performance. In the first two years in the profession, however, teachers tend to go through an adjustment period where they learn the curriculum and culture of a school and may also find that the profession of teaching may not be a match for their skills and disposition. As such, there is little evidence in Rivkin's study of teacher performance gains after three years.

Teaching experience was found to be significant, but not necessarily linear (Murnane & Phillips, 1981). Boyd et al. (2006) found that teacher experience does not matter after five years, regardless of the path of certification. Rockoff (2004) found that ten years is the cutoff past which the returns of experience, particularly scores in reading, become marginal. Rivkin et al. (2005) found the first year to be critical for teacher effectiveness. This gain in quality concludes after three years, when not much evidence of improvement is found. In another study, the peak years of performance were found to be between 13 and 26 years of experience (Clotfelter et al., 2006). These findings indicate that the experience a teacher gains in his or her first few years is important for effectiveness, but experience after the initial first few years does not matter and can even show declines in performance.

Teacher education and certification.

Subject matter knowledge as measured by teacher performance on tests is not related to student performance (Darling-Hammond, 1999). General intelligence tests, such as IQ tests, also show no relationship with student performance. Indeed, the literature indicates that there is no direct relationship between general knowledge and student performance. However, there is evidence that specialized knowledge in the content area may improve student performance to a point. A study conducted on a large, national dataset found that teacher coursework in his or her subject matter improved student performance (Monk, 1994). However, Monk (1994) found this relationship to be curvilinear, where too much coursework in the subject area was found to bring diminishing returns.

Verbal abilities, however, were found to be impactful on student performance. In a reanalysis of the data from the Coleman Report (Coleman et al., 1966), Ehrenberg et al. (Ehrenberg & Brewer, 1995) found that verbal abilities influenced test score gains for their students. The ability to explain concepts lucidly and in a manner that the students can understand was hypothesized to increase student performance (Murnane, 1985).

Advanced degrees are not associated with an increase in test scores but are associated with higher compensation (Hanushek, Kain, O'Brien, & Rivkin, 2005; Rivkin et al., 2005). Hanushek found that all 113 of the studies he reviewed found no significant results for the relationship between degree earned by a teacher and student performance (Hanushek, 1994). However, there is evidence that having advanced degrees or significant coursework in mathematics does, in fact, impact student math performance (Goldhaber & Brewer, 1996, 2000; Monk, 1994). These findings were limited since the positive impact was only visible with teachers of advanced mathematics courses (Monk & Rice, 1994). Teacher education, in and of itself, does not relate to student performance, except in specialized circumstances.

Teacher certification was also found to be conditionally effective in raising student performance. The type of certification is not as important as a credential itself. For example, certified 12th grade mathematics and science teachers were found to be more effective that non-certified teachers, yet the type of certification obtained by the teacher did not make any significant difference in student performance (Goldhaber & Brewer, 2000). In addition, a study of the New York teacher licensing pathways conducted by Boyd et al. (2006) showed that the traditional pathways of teaching yielded approximately the same results as the alternatively certified teachers. In fact, in certain aspects, teachers who entered through alternative pathways saw more achievement gains in mathematics and English than traditionally certified teachers. In another study of state-level certification, teacher certification in the appropriate discipline was found to be more impactful on student learning than class size (Darling-Hammond, 1999). However, on the individual level in New York, Kane et al. (2008) found no performance difference between certified, uncertified, and alternatively certified teachers. The literature, by and large, implies that certification and degrees matter, but not in the classic sense. Too much bureaucracy can keep highly qualified teachers away from the profession where they could make a large impact. Consequently, the evidence is clear that certification is important, but the path to certification should be more open to college graduates with the appropriate subject-matter knowledge to teach in their discipline.

Large-Scale Assessments

When used appropriately, large-scale assessments can offer a tremendous boon to students and educators. When tests are viewed in context and are valid and reliable, the data produced can empower decision makers, such as teachers, administrators, and policymakers, to help ensure high-quality education for all students. Individual results from these tests can help students judge their mastery of a subject; offer parents insight into their child's performance; help educators isolate students in need of further support; and, finally, help give the educational system, as a whole, feedback on how to best serve its population (Ravitch, 2011, p. 150). The catch, however, is that few large-scale assessments are reliable and valid due to the tremendous margins of error associated with the test itself and with the limits of the interpretation of test data.

Limitations Intrinsic to Large-Scale Assessments

Fundamentally, large-scale assessments are not a complete measure of a student's content mastery due to the nature of knowledge and learning. On a large-scale assessment, students are asked, with severe time constraints, questions that are representative of a larger body of knowledge. Thus, large-scale assessment questions sample knowledge in much the same way political polls sample opinions—small samples that become representative of very large inferences.

To compound the sampling problem, the representativeness of a test can be undermined immensely through subsequent education practices such as teaching to the test. Furthermore, the variation in student performance is largely attributable to background factors, as opposed to school factors (Coleman et al., 1966; Hanushek, Rivkin, & Taylor, 1996). Thus, actual measurement of content mastery needs to be interpreted with caution. Often, oversimplification of test results misleads policymakers to infer school quality from test scores without appropriately investigating the actual signs of educational effectiveness. Understanding this limitation of large-scale assessments should raise flags of caution for policymakers and educators when making decisions using data.

Limitations Intrinsic to the Interpretation of Aggregate Data

The averaging of scores can hide and misrepresent the real story behind the data. Although a single numerical value summarizing the educational competency of a student can be attractive for the purpose of presentations, conversations, and arguments, average values are a misleading statistic since there is no average student and there is no average school. In aggregate, we crudely combine the student who lives in a house with no desk to complete homework at night with another child who receives daily private tutoring. While the two students may have similar experiences in school, their average score gives no relevant information about their actual learning.

Shifts in population can also show misleading trends in aggregation. Suppose that a town recently opened a large university. Viewing only aggregate educational data, it may seem that the university caused a rise in overall test scores. However, this inference would be misleading. It is more likely that the development of a new university drew in academically inclined families who have students that are preemptively prone to performing better on tests. As such, the aggregate data, when not contextually investigated, can be misleading when used as a measure of educational quality.

Threats to Validity

Consider a hypothetical thermometer whose mechanism is broken. Reliability issues that coincide with the measurement level of the device will suggest that the measurement is not consistent. For example, one may sample his or her body temperature three times in a row to find that each reading is wildly different than the preceding. Validity is a quality that, when achieved, allows inferences to be drawn from the data. The broken thermometer in this example is not valid and no inferences should be made upon that limited data. However, the law of large numbers implies that if one takes temperature measurements all day with the broken thermometer, the eventual average of the numbers will show the real temperature, assuming that the inconsistency is random. In fact, each individual measurement ought to have been interpreted cautiously and with reliability in mind, since inferences drawn upon the inaccurate and incomplete data could lead to undesirable regulations consequences.

Moreover, threats to validity are not based solely on reliability. Measurement error, score inflation, bias, and misinterpretation of the results are frequently cited as reasons why test scores ought to be interpreted with consideration. In fact, even when using extremely reliable tests, Koretz (2009, p. 158) found that up to 14% of students who fail an exam would otherwise pass if absolutely no reliability issues had arisen. Thus, threats to validity are important to take into account when considering the test itself and aggregate data.

Test Bias

Bias in tests can be detected when there is evidence of a systematic distortion of test results that are unrelated to the content knowledge being tested. This distortion can occur for several reasons and is generally difficult to detect by evaluating the question content.

Bias, however, does not necessarily exist on the sole predicate of subgroups performing differently on large-scale assessments. Suppose that students who live in higher-income neighborhoods are able to attend better schools than their lower-income counterparts. Theoretically, better schools should improve student performance. Therefore, in this example, which is fairly authentic, the ranges in student performance by subgroup on standardized exams may not be due to bias, since the test may, indeed, measure the underlying content without bias. As such, when subgroups show widely different performances, but not due to bias, it is called adverse impact (Koretz, 2009, p. 265).

Imbalance between the performances of subgroups can occur without adverse impact or test bias. For example, consider two groups of high school students where one group is interested in computer science and the other in English. The students who are interested in computer science are likely to take more mathematics courses, and students who are interested in English are likely to take more English courses. Consequently, overarching differences in performance by the two groups in the mathematics versus the English sections of a standardized exam can be anticipated without the suspicion of bias.

When adverse impact is exacerbated by test bias, however, fundamental problems with test validity will arise. Suppose that a district, through indirect policies, excludes females from enrolling in higher-level mathematics courses. Consider further that any exam given to the general student body of this district may potentially contain one of many well-researched biases, such as the use of complex language with second-language test takers; incorporation of phrasing that is centered around experiences that all students do not share; and existence of particular biases that are commonly associated with lower performing subgroups such as *stereotype threat* (Steele & Aronson, 1995). Combining the possibility of test bias against certain subgroups with the potentiality that the subgroups may have also been systematically disadvantaged can exacerbate the misinterpretation of a student's competency.

Unfortunately, the compounding effects of adverse impact and test bias are not always easily seen due to *Simpson's Paradox* (Simpson, 1951). This paradox states that the aggregated effects of subgroup scores may be hidden when the scores are aggregated into the scores of the larger population. As such, the aforementioned test bias and adverse impact may be lost to interpreters unless the aggregate data are specifically subdivided into subgroup averages.

In addition to test bias that affects individual student performance, test bias can also affect aggregate scores. For example, consider the latest growth of college attendance in the United States. Today college is more accessible to students and therefore larger ranges of students are likely to take college entrance exams. When college education was not attainable to most, a larger portion of students who performed poorly academically simply chose not to take admissions examinations. The intrinsic filtration of average and high performing test-takers from low performing students artificially raised test scores by systematically keeping out students who choose not to go to college. As more students make the choice to go to college, the pool of test-takers will change in proportion to the underlying population of high, average, and low performers. Inferences generated from long-term declines in admissions test scores are inevitably misleading since this trend is not related to shifts in the content knowledge being tested.

Error Rates

When evaluating the year-to-year measurement of a student test score, it is important to understand the complexity behind a test score. Student test scores can be attributed to a large number of factors including, but not limited to, curriculum and instruction; family and community support; student ability and health; peer culture and performance; current and prior teacher performance; past performance on exams; and specific type of test used for the exam (Hinchey, 2010). Furthermore, highly idiosyncratic variables also largely contribute to a student's performance on an exam day, such as noise pollution from a construction site near the school (Kane & Staiger, 2002). In fact, over 50% of yearly changes in test scores were due to factors other than long-term learning (Olson, 2001). Thus, the margin of error associated with performance gains makes it challenging to make inferences on performance gains.

In fact, a large body of research regarding error rates refutes the validity of performance gains. Kane and Staiger (2002) completed a study that found that 74% of the changes in student scores were temporary, and further, 90% of the variance in gains that students achieved can also be attributed to noise or sampling variation.

This finding suggests that when measuring year-to-year growth of student performance, caution must be applied since the impact of unobserved variables is highly volatile. Error rates for interpreting scores on high-stakes exams are not negligible and can have severe negative consequences (Koretz, 2008).

Considering that there are so many issues within a test score, it is surprising that value-added measures became popular in the United States. These measures are found to be unreliable and noisy due to random error (Kane & Staiger, 2002). With larger samples, schools tend to show less growth (Koretz, 2002, 2009, p. 167). On the other hand, when samples are too small, scores fluctuate wildly from year to year. In fact, most of the literature surrounding growth measures warns against using this sort of data for accountability and, rather, encourages the use of it only as impetus for further investigation (Betebenner & Linn, 2010).

Limitations of Norm-Referenced Interpretations

Large-scale assessments are often norm-referenced, which encompasses one of the advantages of large-scale testing. Norm-referenced exams show how a student is performing relative to his or her peers. The advantage of this system is obvious: students are able to find out where they stand compared to their peers. This type of measure makes it relatively simple to sort students by ability level. However, this type of scoring has the disadvantage, by design, of enforcing competition. This competition fosters in the students the misconception that there is limited room at the top and that success can be achieved only in comparison with others. While norm-referenced tests provide insight into student academic standing in relation to peers, actual mastery of a subject cannot be measured through this measure alone.

Limitations of Criterion-Referenced Interpretations

In an attempt to create goals for mastery and levels of achievement, large-scale assessments are often accompanied by measures of proficiency. These measures, however, are multifaceted and must be interpreted with caution. For example, a common misinterpretation is applied to the label proficient for the National Assessment of Education Progress (NAEP) exam. The term *proficient* means that a student displayed competency in difficult subject matter. However, students who fall below proficient may still have the basic ability to be successful in that particular subject. These types of labels are widely misinterpreted by popular media such as the film Waiting for Superman. In this documentary, David Guggenheim incorrectly claims that a majority of students are below grade level when in fact only 24% of students are below grade level. What Guggenheim was referring to, incorrectly, was the number of students not reaching proficiency (Ravitch, 2013, p. 115). Criterion-referenced reporting of large-scale assessments results can inadvertently raise alarms when they are understood without proper context.

Often, the difficulty of a given question or standard is difficult to judge. Proficiency standards are highly inconsistent from year to year; thus, large numbers of students who are deemed proficient one year may lose the label of proficiency due to arbitrary labels and criteria applied by test developers (Koretz, 2009, p. 191). As such, policymakers must carefully examine the data and dissect the meaning of the labels and criteria in order to draw any inference about large populations.

Confounding Effects of High-Stakes Testing

When limitations are considered and interpretation is conducted with caution, large-scale assessments can be very beneficial to the education system. However, when large-scale assessments are used to make high-stakes decisions, the testing becomes an end unto itself. Students and teachers focus on the passing of exams for the sake of receiving external rewards or avoiding consequences. There are a number of major unintended consequences resulting from high-stakes testing that infringe on the effective education of students. As this happens, the positive value of large-scale assessments becomes outweighed by its cumulative negative impact.

After the passage of NCLB, each state was required to administer large-scale assessments to students in Grades 3 through 8 in mathematics and in English. The effects of high-stakes examinations can be seen on students through promotional and graduation requirements. States throughout America are rapidly adopting standardized exams as graduation criteria (Amrein & Berliner, 2002). In fact, 26 states require exit exams for graduation, which affects nearly 70% of the students in the United States (McIntosh, 2012). Moreover, the effects of high-stakes testing can also be seen on teachers through teacher evaluations. Using student performance to measure teacher quality has become more prevalent across the United States, with 29 states requiring student performance to be taken into account in teacher evaluations (Baker, Oluwole, & Green, 2013). These effects on teachers and students trigger a range of adverse practices that encompass score manipulation by school officials and teachers, narrowing of academic curriculum, and unnecessary financial burdens carried by the academic system.

Score Inflation

Score inflation is found across the country and is supported by a large body of research (Amrein & Berliner, 2002; Koretz, Barron, Mitchell, & Stecher, 1996; Koretz, Linn, Dunbar, & Shepard, 1991). This effect can be most clearly seen when students show achievement gains on a high-stakes test; but on a similar low-stakes test, they show no progress or even a decline. When complex social processes are scrutinized, the very measures being analyzed tend to become corrupt. Substantial evidence supports this claim. In fact, Koretz (2009, p. 237) provides several examples from outside of the education field (listed in Table 1) that show how the act of measuring and scrutinizing fundamentally make the measurements invalid.

Table 1

Non-Educational Examples of Inflation that Threaten Measurement Validity

Industry	Measure	Inflation Factor	Threat to Validity
Airline	On-Time Arrival	Extending the flight time to include time on the ground	On-time arrival is now padded by an artificially long flight time.
Post Office	Letter Delivery Time	Devoting special treatment to the sampled letters	Measure reflects fastest possible time as opposed to average or authentic time.
CPU Chips	Benchmark Speed Tests	Designing chips with benchmark tests in mind	The speed of the CPU is misleading in regular usage.
Automotive	Pollution Tests	Cars are designed optimally to pass the specific measurements of known pollution tests	Cars that may pollute alarmingly during authentic scenarios still manage to pass pollution tests.

Performance on high-stakes tests is easily manipulated through methods that have nothing to do with student learning. Practices such as narrowing curriculum, teaching to the test, and discouraging low-performers from taking certain exams can be used in order to artificially increase test scores (Amrein & Berliner, 2002; Neal & Schanzenbach, 2007; Nichols & Berliner, 2007; Phelps, 2005). As evidence of artificial grade inflation associated with high-stakes testing, a study found that when using low-stakes tests as a benchmark of learning, states that show significant gains on high-stakes exams tend to show no growth on the benchmark exams (Amrein & Berliner, 2002). As a matter of fact, some of the benchmark scores declined after high-stakes tests were implemented in the state. This provides further evidence that high-stakes testing artificially inflates scores, even if the cost is potential learning.

Score inflation can also be the result of outright cheating. A significant body of research suggests that, as the stakes for testing become tied with teacher evaluation, the incentive to cheat also becomes higher (Phelps, 2005). A particular study that sampled educators from Arizona found that 50% of educators are aware of colleagues engaging in questionable test-related practices and 25% of educators admitted to being personally engaged in such behaviors themselves (Amrein-Beardsley, Berliner, & Rideau, 2010). This questionable behavior ranges in severity from obtaining information from restricted materials to willingly changing student answers. Most of the questionable grade inflations by teachers, however, are instances of inappropriate behavior in test-settings as opposed to outright violations of integrity. Nevertheless, these findings imply that the political climate in states like Arizona, which ties student performance with teacher evaluation, incentivizes teachers to obtain advantages, even if by breaking test protocols.

In corresponding research, Nichols and Berliner (2007) found that some educators who are against high-stakes assessments choose to corrupt the system in order to undermine the validity of the tests. In fact, Nichols and Berliner compiled evidence to argue that when teacher performance is judged by high-stakes tests, the underlying validity of the exam will always be corrupted. Indeed, narratives are widespread regarding teacher hostility towards the culture of high-stakes testing (Amrein-Beardsley et al., 2010).

Narrowing Curriculum

Perceived irrelevance and low interest in academic curriculum is often associated with truancy, apathy, and disaffection from school (Jackson & Davis, 2000). This can be particularly common to groups that are typically underrepresented (Fine, 1991). Furthermore, NCLB measures only math and reading aptitudes. This is a critical limitation since various integral dimensions of learning such as civics, health, and other disciplines are devalued, as knowledge of other subjects is not incentivized the same way. Accountability measures influence teachers to narrow the curriculum in favor of tested subjects (Koretz et al., 1996; Stecher & Barron, 1999; Stecher & Mitchell, 1995). This is a specific problem since there is literature supporting the value of extra-curricular activities that enhance a child's educational experience (Kronholz, 2012).

There is evidence that school personnel adjust curriculum to align to large-scale assessments. A study found that, in Kentucky, when high-stakes accountability measures were implemented for a subject, teachers were significantly more likely to spend time on that subject for the duration at the cost of the non-tested subjects (Stecher & Barron, 1999). Also, principals are more likely to move their most effective teachers to grades and subjects that are followed by high-stakes examinations to maximize student performance as considered by the state (Koretz et al., 1996). For example, if a high-stakes mathematics exam is administered in Grade 4, then principals are likely to shift their most effective math teachers to teach that grade. Considering that a student's learning is the accumulation of a lifetime of experiences, this type of illogical curricular shift disadvantages students from retaining year-to-year information in the long run. Further, this situation offers little flexibility for teachers to address authentic learning situations that occur naturally or in response to community ties. The result is that tested material takes precedence over authentic coursework.

Certain criterion-based accountability systems may discourage teachers from spending time on students who are already proficient or who are perceived as too far from proficiency (Neal & Schanzenbach, 2007). Teaching is a complex task, and when the performance outputs are incentivized, teachers, like any professional with a multi-faceted job, will tend to prioritize those tasks that increase their performance measures (Holmstrom & Milgrom, 1991). Actions such as teaching to the test and narrowing curriculum to only those topics that are covered on an exam are obvious reactions to performance incentives, which inflate scores but do not increase subject mastery (Koretz, 2002).

There is evidence of direct score inflation associated with large-scale assessments. A study of Chicago's newly implemented high-stakes testing system found that math and reading achievement sharply increased when compared to similar populations that did not share such policies (Jacob, 2005). Furthermore, this achievement gain was not present for the lower grades that took equivalent low-stakes exams. This implies that faculty neglected to teach the underlying skills associated with subject mastery and were merely coaching students specifically towards skills reflected on the high-stakes exam. This implication is supported by Jacob's item-level analysis that found that teachers focused their curriculum on test-specific strategies and strategic retention and placement of students in order to maximize test score gains. Furthermore, in an unpublished study, Jacob (2002) found that subjects not tested by high-stakes standardized exams, such as science and social studies, witnessed a leveling-off effect or even a decline in performance. Furthermore, there was no evidence of increase in the parallel low-stakes exams offered to students during that same time. Jacob's study suggests that the gains students saw in Chicago were not due to mastery but were influenced by the change in curriculum, which focused on gains towards that particular high-stakes test.

Financial Costs

The financial costs associated with federally mandated large-scale assessments are substantial. Developing and administering exams are the tip of the iceberg. Large financial costs stem from supporting exams through curriculum, teacher and administrator professional development, remediation, and curricular assistance. For example, when new testing is required by a district, the schools in that district are compelled to purchase new sets of textbooks aligned to that test and to train teachers and administrators to become more familiar with what is being asked of the students in professional development sessions. These costs are not often included in estimates and require a significant shift of resources. In the years since this estimate, the federal government spent $390 million each year, totaling almost $4 billion in spending on NCLB testing (Levine & Levine, 2013).

Costs of the large-scale assessment reforms throughout the United States were estimated at an increase of $600 per pupil spending in 2009 dollars (Dee, Jacob, & Schwartz, 2013). There is limited research available regarding the investment of finances. However, a preliminary cost-benefit calculation showed that an increase in mathematics performance over the last decade could be attributed to accountability measures (Dee & Jacob, 2011). This small but significant increase in mathematics performance, when using Krueger's (2003) estimates of increased earnings through the rise in elementary school math grades, accounts for the increased spending per student. However, Dee et al. (2013) warned about using this calculation with caution and that financial estimates alone do not take into account other negative consequences of high-stakes testing.

Another recent study (Levine & Levine, 2013) found results similar to Dee (2013). However, in the cost-benefit analysis, Levine and Levine (2013) found that the biggest winners of the testing expenditures are the testing companies who earn $2.8 billion annually. At the same time, instructional and administrative resources are lost each year in the preparation and administration of large-scale assessments (Zellmer, Frontier, & Pheifer, 2006). Thus, it can be viewed that while the testing companies gain resources from the expansion of large-scale assessments, schools shift their resources and gain nothing directly through the expansion of large-scale assessments. Meanwhile, the perceived gains in mathematics scores can be explained by the literature, which implies that students simply get better at tests when they are trained to take them.

Long-Term Outcomes and Student Performance

In addition to earnings benefits, education can improve many factors for the individual's way of life (Behrman & Stacey, 1997). Thus, to complete an analysis of the benefits of student test performance would necessarily include benefits outside of the labor market. An analysis of the full effects of student performance would enhance the body of knowledge about student performance and serve to inform the knowledge base of policymakers.

The major measures of non-market benefits of education include health, family, fertility, child welfare, and the environment. That said, of all of education's benefits, earnings and health are by far the most widely researched fields historically (Behrman & Stacey, 1997). The following section relates to individual long-term outcomes and explores the body of literature concerning the broad categories of health, career, and societal outcomes.

Education is measured through student attainment and performance. The attainment of a degree or years of school completed are both known as educational attainment or education levels in the literature reviewed in this section. Student performance as measured by large-scale assessments is a measure of cognitive abilities that are associated with schooling. In the literature, large-scale assessment performance was found to be a better predictor of outcomes than educational attainment (Adebayo, 2008; Alderman, Behrman, Ross, & Sabot, 1996; Blau & Kahn, 2005; Geary, 2011; Heckman, Stixrud, & Urzua, 2006; Murnane, Willett, Duhaldeborde, & Tyler, 2000). Nevertheless, Behrman (1997, p. 46) states that even though student performance is an important product of schooling, it is not used systematically to predict outcomes in the literature. Consequently, using Behrman's framework (Behrman & Stacey, 1997) of analyzing the social benefits of education, I extended the definition of educational attainment to student performance since it is a measure of educational attainment.

This study and review of literature is limited to the private benefits of education. Nevertheless, the same benefits of education for individuals may also benefit society at large (Becker, 1993; Coleman, 1988; Portes, 1998). For example, if better test performance is related to making better health choices, being more environmentally conscious or enjoying longer lives, then the benefits can translate well beyond the private returns. Better health choices may include purchasing healthier foods, thus influencing local food manufacturers to lean towards healthier food production in general. Living longer and healthier lives can translate into more years of productive work and consumerism, especially in the later stages of life when an individual has a higher probability of accumulating wealth.

Conceptual Framework

The educational production theory is used to describe the relationship between student performance and long-term outcomes. Preferences formed by children, and subsequently as adults, are a mixture of heredity and environment. Schools heavily influence the values and preferences of an individual given the amount of time a child spends in that environment (Arrow, 1997, p. 15). Arrow makes the case that schools inherently have a significant impact on student values by providing students with information on a continual basis. In his discussion of the benefits of education, Arrow points to the incongruity of the research that states that additional schooling adds to an individual's income later in life, yet has little effect on his or her cognition as measured by large-scale assessment performance. That is, the signals obtained from years of schooling and degree completion increase benefits without the evidence of skill acquisition. This finding supports signaling theory and shows a limitation to the effects of human and social capital theory in education.

Causality

It is well known that the amount of educational attainment is claimed to afford the individual a buffer with which to cope with change and take advantage of opportunities which he or she may otherwise not pursue (Hanushek, 1995; Lazear, 2001; Welch, 1970). Education may more easily allow a person to learn and adapt to new technologies and skills required for creating income or making health and life choices based on the most current information available. Nevertheless, causality should not necessarily be assumed for the relationship between education and its benefits (Behrman & Stacey, 1997). It may be reasonable to assume that having more education, as measured by years of school and degrees attained, will allow an individual to obtain higher-level employment and other benefits. At the same time, the reverse can be true as well. Individuals can both benefit from an education, and, at the same time, the individual can benefit solely from their skills and dispositions regardless of education. An individual who is predisposed to traits such as inherent ability, motivation, and social connections can benefit from those traits in and of themselves and would also be more likely to obtain higher levels of educational attainment as a side effect.

Health Outcomes

There is a large body of literature that supports the positive relationship between health and educational attainment (Cutler & Lleras-Muney, 2006, 2010; Kenkel, 1991; Lee, 2000; Ross & Wu, 2007). Even when controlling for other factors such as income and health insurance, there is still a strong positive association between health and education (Grossman & Kaestner, 1997). Furthermore, those who have low levels of education and performance are more likely to experience depression, hostility, and stress (Adler et al., 1994; Sewell &

Hauser, 1975) and have less social supports to help them through tough times (Adler et al., 1994; Christakis & Fowler, 2007). The higher likelihood of suffering illness and mental stress are associated with significant health risks and can reduce the longevity of a person substantially (Elo & Preston, 1996; 2004; Kitagawa & Hauser, 1973).

Those who are more informed about the harmful effects of certain behaviors would be more likely to engage in more healthy behaviors to the best of their knowledge, which would result in improved health overall. This is referred to in the literature as *productive and allocative efficiency* (Rosenzweig & Schultz, 1982). For example, if a person were fully informed and aware of which foods are most nutritious, he or she would be more likely to seek out healthier options within his or her circumstances. According to Behrman (1997), productive and allocative efficiency combine to create a framework upon which to analyze the relationship between education and health.

The *preference of time* adds to the theory of productive and allocative efficiency to help explain the unobserved effects of education on health (Becker & Mulligan, 1997). In this theory, a school helps students value future outcomes by focusing a student's attention on the future. For example, a common practice in a school is running simulations of possible future scenarios. This type of practice makes the future seem less distant to a student, and thus, the preference of time can be linked to harmful health habits such as smoking and excessive drinking (Becker & Mulligan, 1997, p. 774). In fact, there is research that supports the immense impact of role-playing on student motivation and self-concept (Kamins & Dweck, 1999).

Further, there is literature arguing that effective use of role-play can be used to change attitudes, concepts, and dispositions (Van Ments, 1999). As such, insofar as school helps the student practice thinking about future outcomes, unhealthy habits become less attractive options in the individual's daily life.

The consumption of cigarettes and alcohol has external costs along with private costs. The chief external cost to alcohol is associated with drunk driving (Guerra & Verghese, 1992). Further, the primary external price associated with smoking is related to pregnancy (Maynard & McGrath, 1997). Women who smoke during pregnancy tend to have children with lower birth weights and a host of neurological problems. Thus, the cost of smoking for pregnant women can be viewed as the direct impact on the child in addition to the health impacts on the smoker. Further, although the health impacts of second-hand smoke are not as clear cut in the literature as smoking while pregnant, the health impacts of smoking is wider than the sole impact of smoking on the individual.

Upward bias.

The impact of education on health may be understated in the literature. Due to the self-reported nature of sickness, observations of health are challenging to observe. In order for a person to not feel well, it would not necessarily need to be documented. A health problem can be subtle and yet can impact the individual severely. A person in good health, with ample energy and vigor, would more likely learn better and perform more successfully on exams. Behrman (1997) argues that, like health, various unobserved factors such as child, family, and community dispositions and choices can potentially interact and cause upward bias.

Buffering effects.

There is compelling evidence that the benefits associated with education create buffers for potential health risks throughout the life of an individual. Sewell and Hauser (1975) followed a sample group of Wisconsin males for 10 years after graduating from high school. This study was particularly influential because of the robustness of the data. Over 90% of the respondents stayed in the follow-up surveys. The study found that education's impact on health is compelling because the more educated participants tended to have higher incomes and were less likely to experience financial struggles. All of those benefits are associated with buffering the effects of adverse health-related variables (Sewell & Hauser, 1975). Furthermore, education is associated with creating supportive relationships (Ross & Wu, 2007) and creating a larger, beneficial social network that is linked to better health (Berkman, Glass, Brissette, & Seeman, 2000; Christakis & Fowler, 2007).

Social capital and health.

A large study (Christakis & Fowler, 2007) found that activities such as drinking and smoking are highly related to a person's social network. As such, the authors recommend that programs that help people quit smoking and drinking are much more successful if they have a large peer component. Indeed, the authors make the case that health is a shared social element. Another study (Wolfe & Haveman, 2002) found that more educated individuals are less likely to spread diseases and rely more on their well-built social network to increase their overall well-being. This implies that social capital matters, and the influence of educational attainment can be even farther-reaching than the current literature implies.

There is compelling evidence that a social network influences the health habits of an individual, regardless of background. In a twin study (Sacerdote, 2004), Korean adoptees were randomly assigned to parents of differing backgrounds. The study found that smoking and drinking were strongly associated with the behaviors of the adopting parents. Specifically, a mother's drinking increases the probability of the adopted child's drinking by 19%, and smoking by the mother increases the adopted child's smoking probability by 11%. This conclusion is not surprising when considering the findings from Christakis et al. (2007) and brings to question the hereditary nature of smoking and drinking.

Lasting effects of student performance.

Educational attainment and cognition, as measured by large-scale assessment performance, is not related to the health issues that are associated with getting older. Most income is earned in adulthood, and therefore differences in test performance and years of schooling are largely unrelated to onsets of adult health issues (Elo & Preston, 1996). This implies that the benefits obtained early in life through schooling provide long-term health gains that a person integrates into his or her life.

There is evidence that the health benefits of education are more relevant today than in years past. A study (Lynch, 2003) that used data from over 800,000 participants ranging in age from 30 to 90 years old, found that the chance of a subject reporting poor or fair health rises with the age of the person but decreases with the cohort. The impact of education on the health of the participants, however, becomes more pronounced by cohort. This suggests that education's effects on health are becoming more significant for those born later. This finding agrees with another study (Kim, 2008) that also studied that mental health of participants over time. Unfortunately, these studies imply that health-related inequality is growing in the United States.

Research shows that education, and not income levels, is one of the major sorting mechanisms for health. Grossman (1972, p. 382) found that past a certain point, income did not positively affect the health of an individual. Aging often increases health problems while it also increases income. This finding implies that past a certain income level, income does not necessarily relate to long-term health outcomes due to the inherent association of income and age.

Simply having health insurance does not imply that a person has equal access to medicine. Newhouse (1993, p. 47) found that there is no difference in the use of health insurance between healthy and sick people. This study implies that having healthcare is unrelated to a person's overall health due to the inequality of medical care individuals seek and receive. Indeed, people who are more educated are more likely to have access to modern medical treatments (Lleras-Muney & Lichtenberg, 2002). Thus, simply making new medicines available and more freely accessible would, in fact, widen the gap of health disparity for the most disadvantaged (Mechanic, 2002). Mechanic argues that to decrease health disparities, direct intervention policies should be adopted to (1) allow individuals to attain healthcare, access to medical care and a living wage; (2) create a safety net for individuals with disabilities, those who are homeless, those who are sick and have no insurance, and those who are temporarily without financial resources (Mechanic, 2002, p. 57).

The benefits of education seem to peak in middle age and tend to neutralize in old age. There is evidence that education and age are not linearly associated with health (Kim, 2008; Lynch, 2003). Specifically, Lynch finds a curvilinear relationship between age, education, and health. At age 30, the participants with eight years of education were seven times more likely to report fair or poor health when compared to those with 17 years of education. At age 60, this number went to its maximum at eight times the difference, relative to educational levels, in reporting poor health. The reports of poor health steadily decline to no difference for education levels at age 90. These findings indicate that although education seems to matter more as time progresses, the buffering effects gained are most important in the middle point in life; and as a person ages, the benefits of education relative to health disappear entirely.

Cigarette and Alcohol Consumption

Higher levels of student performance and education are associated with a healthier lifestyle. In a study using random dialing as a way to select participants between the ages of 18 to 90, Ross and Wu (2007) found that self-rated physical health is positively associated with years of education. They attribute this finding to the average differences in lifestyle over years of schooling. For example, variables such as life satisfaction and positive relationships are positively associated with education levels, while behaviors such as smoking, drinking heavily, and not visiting the doctor are associated negatively with education levels.

A study found that smoking cigarettes, but not alcohol consumption, was related to years of schooling (Lantz et al., 1998). Specifically, 42% of people without a high school diploma smoked cigarettes, while 33% of those with some high school education consume cigarettes, and only 20% of those with a college degree are current smokers. However, the same study did not find that correlation with abstaining from drinking. In fact, Lantz et al. found that alcohol consumption increases with years of schooling.

Alcohol consumption does not affect health linearly. Ross et al. (2007, p. 740) found that rare drinking and moderate drinking have similar effects on self-ratings of health, and heavy drinking is negatively associated with health ratings. Interestingly, abstaining from drinking is negatively associated with health ratings. However, this could be due to the theory that those who have lower ratings of health may abstain from drinking because of health purposes, and therefore, causality cannot be presumed.

Benefits of information.

Health knowledge, particularly relating to cigarettes and alcohol, is an important factor in reporting reduced use (Kenkel, 1991). However, even when health knowledge is controlled for, education level is strongly associated with reducing cigarette and alcohol use. Health knowledge, in and of itself, is not enough to influence healthy behaviors. It is well known that those who are more educated are less likely to smoke cigarettes (M. Grossman & Kaestner, 1997). Surprisingly, a survey taken by Viscusi (1992) found that both smokers and non-smokers overestimate the negative effects of smoking. Indeed, even teenagers, who normally have less information than adults, perceive exaggerated effects of smoking on health. As such, it can be concluded that those who engage in smoking do so by choice, regardless of the knowledge that smoking can have negative internal and external costs.

There is evidence that an increasing knowledge of the population is not enough to create equal access to high-quality medical care and medicine. Mechanic (2002) argues that knowing that certain lifestyle activities are detrimental to health is not enough to help bridge the health gap that is associated with a lack of educational attainment. Mechanic states that helping the entire population become healthier through policy changes increases the disadvantage for the least educated of the population. Considering that those who are more educated are more likely to take advantage of the latest advances in medicine, the theory presented by Mechanic suggests that the way to help those who are most disadvantaged is not by helping them understand information, but rather by aggressively targeting those at most risk and helping them directly through policies and social programs.

Alcohol consumption is particularly harmful if an individual drives while intoxicated or drinks in excess. Kenkel (1993) estimated alternative policies to reduce drunk driving. He found that the price of alcohol in a given city relates directly to the number of times a person drinks five or more drinks on one occasion, which is known as binge drinking. Binge drinking is also associated with negative long-term brain function side effects (Courtney & Polich, 2009) as well as other risky behaviors (Wechsler, Dowdall, Davenport, & Castillo, 1995). Kenkel (1993) further finds that binge drinking is negatively correlated with schooling, and binge drinking is positively associated with drunk driving. The research implies that, although drinking alcohol is difficult to separate from other negative behaviors, drinking and driving and drinking in excess are known to have harmful health effects.

Information regarding the potential harmful health effects of drunk driving and excessive drinking can buffer the negative health impacts associated with drinking. A study found that when knowledge of drinking effects is included in the regression, the negative association between binge drinking and education is reduced by as much as 20% (Kenkel, 1991, p. 297). Thus, those who are more educated may be able to obtain better information about the consequences of drunk driving and be more aware of the risks they take when drinking. Kenkel (1991) found that those who are more educated smoke less, binge drink fewer times, and exercise more. Furthermore, Kenkel acknowledges the high correlations between education and health knowledge. Thus, the gains in health knowledge need to be substantial, and in many ways unreasonably high, in order to surpass the benefits of education on healthy choices.

There is substantial evidence of education's association with healthy choices regarding cigarette and alcohol use. More educated people smoke fewer cigarettes per day and are less likely to be heavy drinkers (Wolfe & Zuvekas, 1995). Indeed, cigarette smoking is directly related to educational attainment (King, Dube, & Tynan, 2012). Each year of school is associated with a reduction of daily cigarette use by 1.6 for men and 1.1 for women with each additional year of schooling (Wolfe & Zuvekas, 1995). The researchers were, by and large, able to control for other factors and still found education levels to be related to healthy behaviors later on in life. This evidence overwhelmingly implies that the attitudes, disposition, and experiences gained in school impact the long-term cigarette and alcohol use of individuals.

Long-term health benefits.

Schooling is associated with longevity (Wolfe & Zuvekas, 1995). In an analysis of census data, a study found that a year of education increased the adult life expectancy in 1960 by 1.7 years (Lleras-Muney, 2005). This result was significant after controlling for income levels. Indeed, mortality rates are related to educational attainment (Backlund, Sorlie, & Johnson, 1999; Singh & Siahpush, 2002). A famous study completed by Kitagawa and Hauser (1973, p. 17) found that at age 25, educational levels could account for an additional life expectancy of four years for males and six years for females. In an earlier study investigating the causes of death related to education level, Kitagawa and Hauser (1968) found an inverse relationship with education levels and mortality for about 83% of the fatality categories in their study for males and females. There were just three major categories of mortality that were unrelated to education levels and they were only related to adult men aged 25 and over. For younger men and women, all major categories of mortality were negatively correlated with education. This study lends further evidence of the benefit achieved during the years a person is obtaining an education.

Health outcomes by gender and ethnicity.

Above and beyond the well-known links between gender, race, and longevity (Felder, 2006), the data show the education gap for health outcomes is increasing (Horton et al., 2010). There is evidence that gender and socioeconomic status (SES) are related to cigarette use later in life. Adler et al. (1994, p. 18) found that, of people who have less than 15 years of education, women tend to smoke cigarettes less frequently then their male counterparts with the same level of education. The linear correlation of daily smoking habits and education holds for both men and women. The same pattern does not hold for consumption of alcohol. High levels of education are associated with moderate drinking outcomes for both men and women (Horton et al., 2010). As discussed previously, the risks of drinking do not come with moderate use; in fact, moderate use of alcohol is related to positive outcomes that include lower levels of stress and heart disease.

Horton et al. (2010) found that the confounding effects of SES, and not health-related behaviors, explain the mortality differences amongst race and ethnicity. When using education as a control, Horton et al. found that the negative effects of health related to ethnicity were buffered by the positive impact of education. A major study of high school seniors in the late 1980s (Bachman et al., 1991) found that, of the major race and ethnic categories, White and Native American seniors had the highest propensity for heavy use of cigarettes, while lack and Mexican American seniors were the least likely to admit to smoking. In that same study, alcohol use was generally higher for males, regardless of ethnicity. Further, somewhat consistent with the cigarette-smoking results, White, Mexican American, and Native American high school seniors reported the highest rates of alcohol use consistent with binge drinking. The authors of the study cautioned, however, that these results are not conclusive, since even heavy use does not necessarily imply life-long addiction. In fact, the authors point to large amounts of evidence that two worlds exist when talking about the use of cigarettes and alcohol. The first world is that many individuals are limited in their use because of the lack of income. For adults with regular access to income, the use of alcohol and cigarettes are some of the largest risks of preventable death amongst all racial groups.

Societal Outcomes

Owens (2004) found that increased education of the population has significant positive effects on society at large. The benefits include better public health, government, lower crime rates, environmental improvements, and positive involvement with the community overall. Education is also the most important predictor for social capital (Helliwell & Putnam, 2007). Specifically, education is generally positively associated with increases in trust and social engagement. This research indicates that education relates positively to political and social engagement (Helliwell & Putnam, 2007; Wolfe & Haveman, 2002). In fact, in a large meta-analysis of education's returns of social capital, Huang et al. (2009) found that one SD of schooling accounts for about a 15% increase in social participation. As such, education, after controlling for many other variables, accounts for a large part of social capital that exists today.

Voting activity.

Educated individuals are more likely to make informed decisions while voting and participate more in their communities (Wolfe & Haveman, 2002). Yet there is evidence that the education differentials in the United States may be due to the voter registration policies that inequitably discourage specific subgroups from voting. Indeed, a study attempting to link education with voting habits in the United States and the United Kingdom found that there is a strong effect on voting in the United States, but not in the United Kingdom (Milligan, Moretti, & Oreopoulos, 2004). The differing laws and policies regarding voting the United States and the United Kingdom are likely to be a reason for education being so highly related to voting in the United States. In the United Kingdom, voters are automatically registered by government agencies; thus, any citizen can show up and vote on election days. In the United States, voters must register to vote in advance of the election. The policies vary by state and local laws; but overall, the registration laws essentially raise the cost to vote in the United States.

The voter registration process in the United States creates barriers that may be too challenging for less-educated individuals to overcome. Wolfinger (1980, p. 62) makes the case that voting turnout will always be lower when there are obstacles in the way. Those who are more educated, according to Wolfinger, are able to better cope with the bureaucratic hurdles since the nature of the issues with voting are similar to that of school completion; that is, in order to obtain a diploma and earn credits, there will be bureaucratic actions that a person will need to take before he or she is able to advance to the next level or simply find out what is needed to continue his or her education. As such, Wolfinger finds that education level is the best predictor of voting in the United States.

The educational gap for voting in the United States may be due to relative education and not necessarily an absolute measure of education. A theory presented by Feddersen et al. (1996) is that voting may not necessarily be directly related to education, but rather to one's relative knowledge of the issues and policies. Wolfinger (1980) finds that education is a strong predictor of voting, and Feddersen et al. (1996) find that individuals would abstain from voting if they perceive themselves to be relatively uninformed of the issues, regardless of their education level.

Knowledge of current events may drive people to vote due to a perceived need to change social issues. Using the data from a large national dataset, Dee (2004) adds further evidence that educational attainment is positively associated with voting and knowledge of current events. Specifically, Dee estimates that college entrance increased the probability of being registered to vote or voting by approximately 30%. Overall, Dee finds that college entrance increases voter participation by about 25%. In the same study, but using a different dataset, Dee (2004) found that an additional year of schooling increases the probability of voting by 5%. Furthermore, an additional year of schooling increases newspaper readership by 3%, a relatively weak influence. Newspaper reading is associated with understanding current events and being more politically aware. This research supports the link between education and cognition with voting activities. The evidence implies that those who are aware of whom and what they are voting for are more inclined to vote.

Volunteering and other social involvement.

Volunteering is positively associated with years of schooling (Freeman, 1997; Musick & Wilson, 2007). Dee (2004) found that college entrance increased the probability of volunteering by 20%. Furthermore, Dee found that an additional year of schooling increased the number of groups such as clubs, unions, and other social groups to which a person belongs by 12%. Adults with more years of education tend to volunteer more of their time (Hayghe, 1991). About 40% of college graduates volunteered their time when compared to 10% of people who did not complete high school. Further, another study found that college graduates donated more of their income to causes than high school graduates (Owens, 2004, p. 13)

Selection bias may indicate an overstatement of the effects of volunteering. Although the highest achievers in high school tend to volunteer more of their time, the relationship may be confounded with the higher likelihood of participating in extracurricular activities in general, which include volunteering (Musick & Wilson, 2007, p. 120). Furthermore, in a recent study using data from 85 sets of adult identical same-sex twins in New Zealand, Gibson (2001) found a source of potential bias with other volunteer studies. Gibson (2001) found a reduction in the probability of volunteering by 12% for each year of schooling. This conclusion counters the findings in other studies and lends credence to the opportunity cost that comes with schooling and also supports the idea of selection bias for volunteers; that is, although volunteers may be more educated on average, having more schooling may not increase the likelihood of volunteering.

The sense of obligation to the community, which is a critical factor for volunteering, is positively associated with schooling. A recent study analyzed a large, representative dataset of adults whose ages ranged from 25 to 74 (Son & Wilson, 2012). Their conclusions were consistent with other studies that found that education was highly associated with volunteering. The study suggests that this is due to an increased sense of two forms of obligation: *altruistic obligations* are a sense of ethics that move people to volunteer and *civic obligations* are a sense of responsibility to get involved and improve the community. Son and Wilson found that, even when controlling for religion, years of education is a far stronger predictor of both types of obligations; and they imply that the disposition to these obligations lead to life-long volunteering habits.

The school is a place where a student learns about other situations than their own. In the book *Volunteers: A Social Profile*, Musick et al. (2007) explain that although education is the universal variable that transcends all other characteristic variables, education may be what intertwines all the other variables that lead to the choice of volunteering. For example, a person who may do well in school would also be likely to learn about the plight of children in need. Once a person learns of the struggles of others, he or she may be more inclined to volunteer his or her time than if the person had never heard of this opportunity to help others.

The domain of volunteering is vast, and may not draw equally in relation to levels of education. Musick et al. (2007, p. 35) find that there are 12 domains of volunteer opportunities, and that some domains are likely to draw individuals with higher educational attainment. For example, volunteering on the board of trustees of a non-profit school program would more likely attract those individuals with higher educational attainment than volunteering at a soup kitchen.

People who are more educated tend to have more cosmopolitan attitudes, larger social networks, and may be asked to volunteer more frequently. Signaling theory may play into this since educated individuals are flagged as those with more capability to do volunteer work (Musick & Wilson, 2007). Indeed, the social pressures of volunteering strongly impact the individual’s behavior. In fact, young adults, usually with smaller social networks than older adults, volunteer their time least often, regardless of levels of education. The actual payoff of education happens between the ages of 40 and 59 (Musick & Wilson, 2007, p. 125). This finding may be due to the building of social capital to make use of those signals of capability associated with volunteering.

Cognitive abilities, as measured by student performance, are also linked to volunteering. Verbal abilities, and not math ability, as measured by large-scale assessment performance, are associated with volunteer hours (Nie & Hillygus, 2001). Specifically, students with the lowest SAT verbal scores performed one hour of service on average while the highest scoring donated more than 12 hours. Majoring in humanities and social sciences is related to more community service time. In college, service increased from about two hours to nine hours as students took more social science classes (Nie & Hillygus, 2001, p. 48). Interestingly, they found a strong link between civic volunteering and social science courses, but no relationship with humanities courses. Finally, after evaluating five disciplines, they found that those majoring in business and science tend to volunteer about 10 hours fewer than those majoring in humanities and social sciences; and those majoring in education were in between the majors. The results of this study show a potential link between education level and cognitive abilities in relation to societal involvement. In a school, students learn about situations other than their own and can be moved to act on behalf of others. Since those who are more educated tend to have stronger social capital, being aware of the plight of others can move communities to act on behalf of those in most need.

Career Outcomes

The application of human capital theory on curriculum rests on the assumption that students can gain skills through coursework in their school that are applicable to their productivity in the job market. The types of skills learned in a particular course or series of courses can be related directly or indirectly to a particular job, thus making the student more productive. Skills such as time management, logic, and reasoning can help increase productivity (see Gamoran, 1994, for a review of curriculum paths to productivity).

In addition, when social capital theory (Coleman, 1988) is applied, those students who take more advanced or specialized courses would be actively creating a social network to support their opportunities later in life to find employment through various means. The adults who teach these more advanced courses can potentially provide guidance for future employment. The students may keep in contact with their like-minded peers who, in turn, can help the students find their way to a higher paying job in the future.

In contrast, Spence (1973) suggests that the courses students choose do not increase productivity. Rather than obtaining skills through the coursework, signaling theory implies that the students who have already obtained the necessary skills are more likely to choose more advanced courses. Further, the students who take more advanced courses in high school signal their ability to colleges, and then prospective employers. It should be noted, however, that the link between high school courses and grades and employment are not clear since evidence presented by Bishop (1989b) suggests that employers do not check high school transcripts, nor do they value high school achievement.

There is evidence that the impact of education on salaries is getting larger.

Allen (2001) found that industries heavy in research and development (R&D), such as electrical equipment manufacturers and chemical producers, provide much higher returns to education than those industries with relatively few scientists and engineers, such as retail sales and apparel. Allen found an almost 3% higher return per year of education in those higher R&D industries. Allen ran his model for data from the 1980s and found that, as the needs of many industries go towards R&D and adopting new technologies, the difference between worker earnings by way of years of education increases greatly. Technology usage amongst all industries is rapidly expanding; thus, the influence of education on earnings will increase along with this trend. The impact of R&D and technology increases earnings of college graduates much faster than high school graduates.

Characteristic level.

Family background variables are only modestly related to earnings later in life. Haveman and Wolfe (1995) found that parents' education status is not related to the income levels of their child later in life. Rather, they find that the income level of the parents is closely related to the income level of the child later in life. This implies that there are unobserved variables in higher-income families that allow more pathways to increased earnings for the student. Haveman and Wolfe (1995) further found that once a student's ability and educational level are controlled for, the influence of parental variables on income is significantly reduced. Lazear (2003) found that salaries are raised by 4.5% per year if the parents completed college.

Teacher quality.

There is significant research on the positive association with teachers and later income levels. Chetty et al. (2011) finds that a one SD increase in teacher quality is associated with a 1% gain in salary over the lifetime of the student. When considering Hanushek's theory of teacher deselection (2009a) estimates that if 5% to 10% of the bottom-performing teachers left the profession permanently and consistently, overall student performance would improve by 0.5 SD. Chetty et al. estimate the benefits of low-performing teacher deselection to be about $200,000 per student over his or her lifetime. However, as stated in the review of teacher quality research, these conclusions are purely theoretical and are problematic in their implementation.

Student performance.

Earnings later in life are interwoven with test performance and noncognitive factors like resilience and determination. Heckman et al. (2006) conducted a study of a large, national dataset and found that cognitive abilities play an important role in predicting earnings. Furthermore, they found that noncognitive abilities are equally impactful but are challenging to measure directly. Specifically, Heckman et al. found motivation, persistence, and positive self-esteem to be predictors as strong as cognitive abilities, as measured by student performance on tests, to be impactful on earnings.

Student performance plays a significant role in the United States in explaining wage differentials. Even after controlling for education, experience, and age, Blau et al. (2005) found that people in the United States have higher income inequality related to test scores than other comparable countries; that is, those performing at the bottom earn 17% less than those in other countries, in the median earn about the same, and at the top earn about 14% more than those in other countries in their study. Specifically, one SD difference in test scores accounts for 16% and 12% increase in earnings for men and women, respectively; when compared to non-U.S. countries, the figures are starkly lower, 9% and 8% for the respective genders. These findings indicate that cognitive measures matter in labor-market outcomes in the United States. Blau et al. (2005) conclude that the differences between the United States and the other countries may be due to the sheer size of the United States when compared to the other OECD countries. The United States has an abundance of low-scoring individuals in the labor market. As such, income inequality in the United States is positively correlated with education and cognitive abilities as shown through large-scale assessment scores. The findings of this study add to a previous study that showed the positive relationship between student performance and earnings for the countries included in the study, regardless of whether the country is developed or developing (Hanushek, 2009b). It is important to bear in mind that comparisons between other countries and the United States are challenging because of the size and nature of the U.S. economy and population.

When studies control for education, student performance still stands out as a significant factor in explaining earnings differentials. In a 1980 longitudinal study of salary of those students who were in high school in 1972, Bishop (1989a) found that a one SD difference on test scores pays off even when the student is not working. Income difference, regardless of education level, rises from is the age of 19, when the student makes about 3% more per SD; and the earnings gap becomes even wider at 25, when the earnings difference averages a 10% premium per one SD rise in test performance. There is further evidence that student performance is positively related to later earnings. For example, a one SD gain in testing performance increases income 15% for males, and 10% for females (Murnane et al., 2000). Using a large national dataset, Lazear (2003) estimates the future earnings of a 1988 class of eighth graders in 2000, based on their combined scores on a composite score of large-scale assessments. Lazear finds 12% income growth for each SD in test score gains. These results imply that test performance influences earnings in unobserved ways beyond what schooling offers. The link between test performance and noncognitive factors may cloud the true effects of cognitive measures that are being measured by the large-scale assessments.

Educational attainment is positively associated with both cognitive and noncognitive factors. In fact, Cawley et al. (2001) found that the impact of schooling on cognitive and noncognitive abilities makes the separation of test performance and educational attainment very difficult (Cawley, Heckman, & Vytlacil, 2001). Thus, when attempting to untangle the effects of test performance on earnings, only a modest impact was found. These findings suggest that the analysis becomes complicated when measuring cognitive abilities due to the nature of performing well on large-scale assessments. Those who have high performance on tests tend also to continue their education further, and those who continue in education rarely perform poorly on standardized exams.

There is evidence that student performance, even early on, has a relationship to future earnings. Krueger (2003) found that an increase in one SD for math scores in elementary school is linked with increased earnings in adulthood of approximately

8%. In another study using two large national datasets, Murnane et al. (2000) found that for high school graduates, test scores and family background account for only 25% of the variation in test scores. The dataset had severe limitations that did not allow the researchers to investigate the true impact of test scores on salary. As such, Murnane et al. (2000) concluded that the primary advantage of higher test scores is due to its relationship with educational attainment. However, the study did find some interesting points of evidence: one SD in mathematics score is associated with a 6% increase in salary for females and 15% increase for males after controlling for education. Mulligan (1999) found that a one SD increase in a test performance is associated with an 11% increase in earnings after controlling for years of schooling and other background factors. This association is relatively large when compared to additional years of schooling, which are associated with approximately 8% increases in earnings (Hanushek, 2009b).

Evidence from international literature suggests similar returns to student test performance. Although the metrics are different from the studies above, a study from the United Kingdom found that test performance, particularly on mathematics, is associated with being employed (McIntosh & Vignoles, 2001). Further, they find that going past the lowest levels in test performance associated positively with earnings. Specifically, McIntosh et al. (2001) found that those scoring at the lowest levels in literacy earn about 17% less than their peers, and those scoring at the lowest level in mathematics earn about 6% less than their peers. In Canada, a study found that literacy scores are positively associated with income, while mathematics scores are not significantly associated (Finnie & Meng, 2002). Overall, international literature provides further evidence for the positive association of student performance and earnings later in life.

Curricular and school-related benefits.

In a review of existing literature, Card (1999) found that years of school increase earnings above all other observable variables. There is a large body of research about the impact of education on earnings (Becker, 1993; Hansen, 1963; Schultz, 1961). Indeed, in 1998, the median earnings of college graduates was 74% higher than those with a high school degree (Wolfe & Haveman, 2002). Furthermore, there is evidence that school attainment accounts for a about an 8% increase in wages (Hanushek & Raymond, 2006).

There is further evidence that the signaling effects of schooling are more impactful than the gained abilities in the courses taken. Altonji (1995) conducted a study investigating the link between courses students take in high school and labor market outcomes. He found that the returns on individual courses are relatively small when compared to the returns on a year of school. When course credits are controlled for, social studies and English showed negative returns, while returns on mathematics, science, and foreign languages showed a small return of 3% (Altonji, 1995). Without controlling for academic subjects, the overall return on an additional year of coursework in all subjects was 0.3%. The added labor market value for coursework is markedly small when compared to the 7% return of an additional year of schooling (Altonji, 1995; Rose & Betts, 2004). The additional courses taken by a student add only marginal value when compared to the signal of a year of schooling. On the other hand, human and social capital theory predicts that a student would become more productive and thus earn a higher income as a result of taking more courses. Therefore, a student who takes advanced coursework would benefit intrinsically from sitting in those courses and learning new skills. However, the evidence presented implies that the coursework is of negligible benefit to the student for the most part. Rather, the research results suggest those years of schooling act as a weed-out process for those students who are unable to continue in school.

Although the research shows that coursework when taken overall only marginally influences labor market outcomes, the types of courses the student takes matter. Out of all test scores, math is the most influential on future earnings (Murnane, Willett, & Levy, 1995). Using a large, national dataset, two studies found that course content, and not necessarily course credits, are predictors for key outcomes. Algebra II is identified as a signal course for two significant outcomes. The course is a key predictor for students who will earn above $40,000 annually (Carnevale & Desrochers, 2001) and is also an important factor when predicting college completion (Adelman, 2006). These findings further support the signaling value of gatekeeper courses and provide evidence against coursework helping students learn skills that ultimately result in higher earnings.

Rose and Betts (2004) support the large body of research suggesting that math matters. Specifically, the study estimates that credits in algebra and geometry can account for an 8%-9% gain in income. Even after controlling for background, attainment, and occupation, the type of math courses the student takes in high school are still found to be strongly related with earnings at age 25 (Rose & Betts, 2004).

The research suggests that although not all courses are created equal, taking advanced mathematics courses is a key predictor of future labor market outcomes. The policy implications from the research are not clear since causality cannot be assumed from the studies. According to signaling theory and evidence presented from Altonji (1995), it is likely that students who take advanced mathematics courses are not necessarily learning new skills within that course; rather, they enter the course with the necessary skills and dispositions to be successful in that course and therefore be more likely to earn a higher income due to intrinsic motivation, skills, and productivity obtained outside the courses. The research implies that the courses are more likely to be a signal of ability rather than an environment where students obtain skills and dispositions for the labor market.

Education beyond high school has a major impact on the types of benefits received by the employee (Smeeding, 1983). A study of the returns of education in relationship to working conditions, found a downward bias for estimates of wage returns for schooling (Lucas, 1977). That is, Lucas (1977) found that jobs that are repetitive, physically demanding, or need high levels of vocational preparation tend to pay more. Therefore, people who are more educated tend to take jobs that are more pleasant and may not need as much vocational preparation. Those who are more educated may spend their time obtaining non-vocational degrees that would, in turn, create a downward bias when estimating salary returns to education. This research suggests that there are skills that lead to financial returns that are not captured in school per se.

Higher education returns.

The benefits of higher education are generally well known. There are conflicting approaches to understanding the value of a college education. For example, in a contested study, Brewer et al. (1999) found evidence of increased earnings for elite-level colleges when compared to other colleges. In a related study, but using a different approach, Dale and Krueger (2002) found little earnings differences for elite-level colleges. Due to their assumption that elite universities tend to draw individuals who have higher earning capacities due to their family background and that elite colleges tend to take in students with higher college entrance exam scores, there is evidence of higher unobserved ability bias when analyzing earnings. Indeed, student background may entangle the earnings differentials for different colleges. Research supports the view that once controlled for background variables, the return on a year of college is the same for two-year colleges as for four-year colleges, regardless of the selectivity of the institution (Kane & Rouse, 1995, p. 605).

In a study completed for wages in the 1980s, Bishop (1989a) found evidence of a strong and growing link between four years of college and the average subsequent earnings at age 26. The study found that there is also a large difference between the earnings of those students who major in humanities and those who major in business or the sciences. Specifically, physical science, engineering, and business majors earned twice as much as humanities majors. Indeed, for every control employed by Bishop, including years of education and years of experience, majors in humanities earned significantly less than those who major in fields like computer science. These results remained significant throughout the study for both men and women.

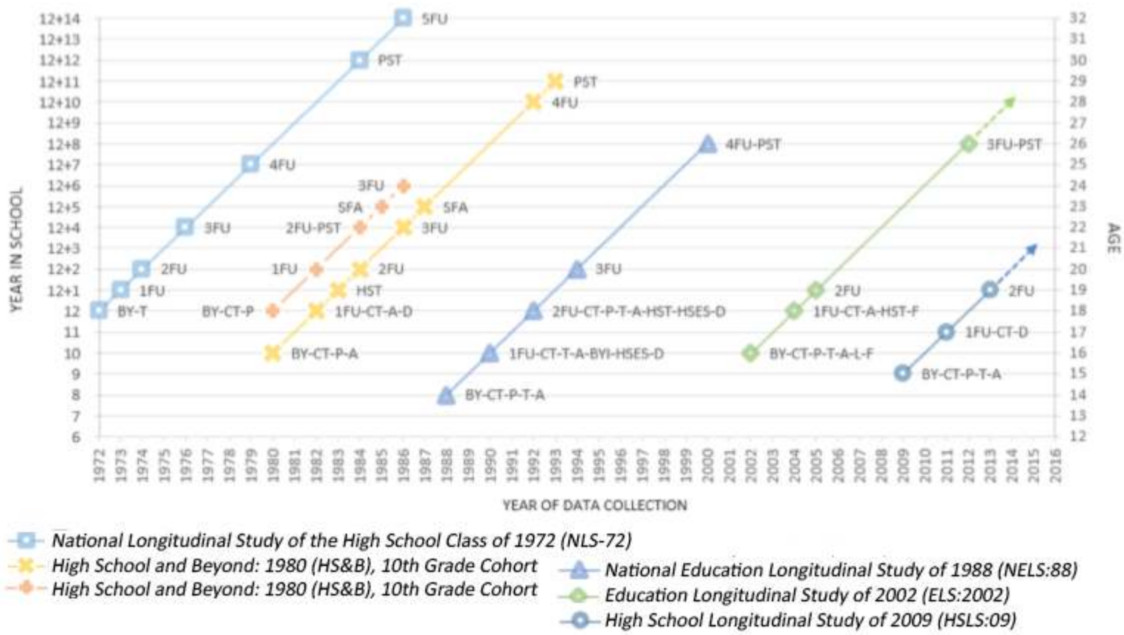
CHAPTER III

METHODOLOGY

Data Source

I selected the National Education Longitudinal Study of 1988 to 2000 (NELS:88/00) for analysis and response to the research questions of this study. The NELS:88 was a 12-year longitudinal study that followed students from eighth grade into adulthood and their entrance into the workforce. Data collected in this study were designed to investigate the role of schools in promoting positive life outcomes (Curtin, 2002, p. 3). Consequently, the dataset contains several technical and design advantages to accompany the analysis of long-term outcomes. I drew upon the measures of student performance as captured by specially designed large-scale standardized exams and tested those measures for predictive power towards long-term outcomes as indicated by the final follow-up in the year 2000.

The NELS:88 study is just one of several longitudinal studies that track student performance over a significant period of time for the examination of long-term outcomes. However, two specific advantages of using this study, as opposed to other National Center for Education Statistics (NCES) high-school cohorts, include the following. First, the NELS:88 administered three rounds of standardized exams where each additional round increased the accuracy of growth tracking (Betebenner, 2007). Second, the NELS:88 was the most contemporary of the available longitudinal studies that measure long-term outcomes. Note that the High School Longitudinal Study (HSLS) had not released its data for the 2012 cohort at the time of this study. Figure 3 highlights the advantages of the NELS:88 dataset in relation to the other NCES studies. As shown in the graph, the NELS:88 is the most contemporary, indicated by its position to the right of the graph, of the longest running studies, as indicated by its height within the graph, with the most frequent round of examinations, as indicated by the corresponding note on the graph.



Key

HST=High School Transcript	3FU=3rd follow-up data collection
PST=Post-Secondary Transcript	4FU=4th follow-up data collection
SFA=Student Financial Aid	5FU=5th follow-up data collection
BYI=Base Year Ineligible	CT=Cognitive test P=Parent
HSES=HS Effectiveness Study	survey T=Teacher survey
D=Dropout Survey	A=Administrator survey
BY=Base Year data	L=Library/media center survey
1FU=1st follow-up data collection	F=Facilities checklist
2FU=2nd follow-up data collection	

Note. This figure was updated for 2013 data and adapted from *Base-Year to Fourth Follow-up Data File User's Manual* (Curtin, 2002, p. 3).

Figure 2. Longitudinal design for the NCES high school cohorts.

Data Collection

Data for the NELS:88 study were gathered in the form of questionnaires and exams administered to groups of approximately 23 students from each school sampled. During each session, students were asked to respond to a questionnaire and then allotted one and a half hours to complete the corresponding standardized exam (Ingels, 1994). Each exam included sections of varying lengths: (a) the reading section contained 21 questions to be completed in 21 minutes; (b) the mathematics section contained 40 questions to be completed in 30 minutes; (c) the science section contained 25 questions to be completed in 20 minutes; and (d) the history, citizenship, and geography section contained 30 questions to be completed in 14 minutes (Rock et al., 1995). Tests administrators then reviewed the completed exams for missing answers or stray marks before students left the classroom.

Participant Pool

The NELS study gathered its data from a representative sample of over 25,000 participants in the eighth grade from 1,000 different schools. Random selection of schools and participating students allowed this sample to represent approximately three million eighth graders and 40,000 public and private schools all over the country for 1988 (Curtin et al., 2002, p. 11). This large and nationally representative participant pool increases the applicability of the corresponding data to a relatively wide population.

Cognitive Batteries

Exams administered as part of the NELS:88 study were technically robust. For example, in order to minimize floor and ceiling effects when measuring growth of student performance, a multilevel design was employed that involved tests of varying difficulty. Also, through application of item-response theory (IRT), the administered tests were norm- and criterion-referenced (Curtin et al., 2002). These tests were further modeled upon the exams created by the National Assessment of Educational Progress (NAEP). As a result, the NELS:88 study did not sample schools that were part of an NAEP study in order to avoid artificial inflation of scores.

Application of universal weights renders the NELS:88/00 dataset robust to subsampling (Curtin et al., 2002, p. 91). Subsampling was indispensable to my study, as students who had not completed multiple exams could not be included in calculations of Student Growth Percentiles (SGP), which was a major mode of analysis in this study of student performance in relation to long-term outcomes.

Sequential cognitive batteries were scaled for each follow-up. As a result, this dataset had the distinct advantage of allowing year-to-year comparisons of student performance. A student who improved beyond the previous scaled score showed evidence of growth above and beyond prior ability levels. Furthermore, it is of note that, to the extent possible, cognitive batteries were administered in an equitable manner. Most students were able to complete the exam in the given time and there was no evidence of test bias related to gender or ethnicity (Rock et al., 1995).

Nonresponse Limitations

Evidence indicates that participants who had dropped out of school were also less likely to complete the course of study by partaking in its full battery of tests (Ingels, 1994, p. 115). Approximately one quarter of dropout respondents completed an abbreviated student questionnaire and were not administered further cognitive test batteries. Dropping out of school is a known red flag for adverse long-term outcomes; lack of long-term measurements for students who have done this may, therefore, trigger an upward bias in overall reported long-term outcomes. Moreover, for all test subjects, lower SES was related to higher nonresponse, causing the overall view of outcomes to be further distorted (Rock et al., 1995).

The nature of this longitudinal study required that students participate in five survey rounds and that they complete four full exams. Selection of participants who were willing to abide by this requirement inherently forced a gap between the selected population and the general population. Error bias results when the population of a study differs from the general population in terms of skills, knowledge, and follow-through abilities. If key attributes of the selected participants were, in fact, different from the general population, then statistical conclusions might be limited in the generalization to the total population of eighth graders in 1988.

Nonresponse limitations and associated upward bias that came from the subsampling of students who completed all cognitive batteries is worth further research that is outside the scope of this study. Since students who dropped out were less likely to complete all exams (Rock et al., 1995), the examined population can be thought to perform at a higher level than the general population. This should be kept in mind when applying the conclusions of this study.

Research Design

Sample Weights

Weights used in the NELS:88 study compensate for the complexity of the selection process. This study tried to take into account the inferred population of U.S. eighth graders in 1988, creating weights for school types such as private or public, gender and ethnicity balance, SES, and other characteristics. Weights are used to count students in proportion to the overall population. For example, a specific student may represent 50 people, while another student represents 5,000.

According to the user manual for NELS:88/00, the weight F4PNLWT must be used when a researcher seeks to apply the fourth follow-up data for those who also participated in each prior survey round starting from eighth grade in 1988 (Curtin, 2002). As such, the inclusion criterion is to hold a weight of more than zero.

The NELS:88/00 longitudinal research dataset initially attempted to generalize the sample population to the entire 1988 student body. Since the purpose of the study was to measure outcomes longitudinally and generalize the findings to a meaningful population, the weight F4PNLWT was chosen. This weight applies to respondents who participated in all five survey rounds and was appropriate for describing longitudinal outcomes projected for 1988 eighth graders (Curtin et al., 2002, p. 84). The result of choosing this weight excluded 1,317 out of 12,144 participants, or 11% of the sampled population.

Selection of Factors

Using the Statistical Package for Social Sciences (SPSS) software, Version 21, I conducted partial correlations between all variables listed in Appendix A and created a sorted list for each outcome variable based on absolute value of correlations. The top correlations valued in each category of independent variables were included in a logistic regression model. I conducted collinearity tests and eliminated those variables with values higher then .8, keeping the variable that had a higher initial absolute value with respect to the outcome variable. Mason and Perreault (1991) suggest using cut off points of .8 for partial correlation matrices. Further, a variance inflation factor (VIF) of greater than 10 is considered above the acceptable norm for collinearity (Marquardt, 1970).

Mason and Perreault (1991) state that the effects of collinearity are often exaggerated in the literature. Collinearity must be viewed in combination with the overall accuracy of the model along with the sample size. Usually, collinearity does not occur in high levels and the effects on accuracy and coping techniques for lower levels are not well understood (Mason & Perreault, 1991). Further, Mason and Perreault (1991) found that the effect of collinearity is low when the sample size is large enough ($n > 250$).

Collinearity is measured by the degree to which two predictor variables can be measured through a linear relationship. All predictors have some measure of collinearity, ranging from no relationship to a perfect relationship. The literature suggests that all predictors should have the lowest possible collinearity. This cannot always be avoided since the trend can be representative of the underlying data of the population (Mason & Perreault, 1991).

Dropping one of the variables with high collinearity is one of the simplest ways to deal with any problematic relationships among the predictors (Mason & Perreault, 1991). This method comes with issues that include the selection method of which variable to drop and that the model may be biased since the relationship between the dropped predictor and dependent variable is not zero.

Mason and Perreault (1991) discuss another technique to deal with the problem of collinearity. In methods of this style, a composite variable of the two or more related predictors is used to generate a referential matrix (Farebrother, 1974; Massy, 1965). Although this approach deals well with collinearity, intrinsically it confounds the set of variables. For this study, therefore, the composition properties of creating indices make this technique unsuitable for straightforward analysis of the results.

Preparing Data for Student Growth Percentiles

In order to use the latest techniques to measure growth, I used Student

Growth Percentiles (SGP) developed by Betebenner (2011) instead of simple differences. Using simple differences to detect growth would provide misleading results for those who perform at the top and bottom levels. Therefore, to more effectively understand student performance growth, SGP is used to measure students against their peers. Specifically, student growth is compared only to those students who scored within the same percentile. Using the three sequential exams students completed in eighth, tenth, and twelfth grade, I created a new variable for each test subject. The technique for applying the SGP function to the NELS:88/00 data is described in Appendix A.

Data Analysis

Using the factors identified earlier in this section through a correlation table, I built logistic regression models for each of the desired outcome variables described below. The models are disaggregated based on gender and ethnicity. The purpose of this granular level of analysis was to find out whether large-scale assessments are more predictive for individuals of a given gender or ethnicity. In aggregate, these findings may be hidden due to Simpson's paradox (Simpson, 1951).

Dependent Variables

Variables Measuring Health Outcomes

Measuring excessive alcohol use.

As discussed in the literature review, a large body of research shows that those occasions where a person consumes five alcoholic drinks or more (binge drinking) is associated with undesirable health outcomes. This question was asked in 2000, when most respondents were either 25 or 26 years old. Self-reports of alcohol use are known to be reliable (Del Boca & Darkes, 2003). Specifically, a question that asks the respondent to remember the specific number of drinks helps increase reliability (Embree & Whitehead, 1993; Lintonen, Ahlström, & Metso, 2004). Otherwise, if there is a presence of social stigma in answering the question, results may be reported with a downward bias (Embree & Whitehead, 1993).

Reliability of responses may be high, but unstable drinking patterns may show misleading results (Gruenewald & Johnson, 2006). Considering that the interviews of the 26-year-old participants were conducted either over the phone or in person, there is no clear systematic reason to suspect bias of alcohol reporting.

The researchers for NELS queried how many times a person had consumed five or more alcoholic beverages in the two weeks prior to the interview. For consistent analysis, this question was recoded dichotomously. The resulting question was thus modified to ask whether or not the respondent binge drank in the last two weeks.

Use of cigarettes.

The technique for coding cigarette use is identical to binge drinking. Respondents were asked in 2000 to state how many cigarettes they use in a day. I recoded the response to a binary identification of the respondents into two groups, those individuals who smoke, and those individuals who do not smoke any cigarettes.

The efficacy of self-reports regarding smoking cigarettes is mixed. Patrick et al. (1994) conducted a meta-analysis of the literature surrounding self-reports of smoking. They were followed up by various independent measures to verify the claims that included observations and chemical tests. They found that, like self-reports of alcohol consumption, the presence of social stigma is related to an underreporting of cigarette use. Stigma is associated with an expectation of the cessation of smoking habits and general negative perceptions of unhealthiness. For this dataset, there is no evidence of bias for measuring cigarette use.

Variables Measuring Societal Outcomes

Respondents were asked whether or not they voted in the two years prior to the interview in 2000. This variable is a proxy measure for the degree of civic engagement of the participant. Since this interview took place prior to the presidential election in 2000, the likelihood of voting in national elections is not captured in this study.

To measure social interactions, I created a dichotomous composite variable, which separates respondents into two groups: highly socially involved respondents, and respondents who have low social involvement. For this study, four variables were combined to create this composite: religious activities, attending sports and concerts, playing organized sports, and volunteering. These variables were combined in a consistent manner to measure days in which a respondent participated in social activities. For details on the computation of social involvement, see Appendix A.

The informational involvement variable was created to measure information gathered by the respondent that was not related to work or school. This composite variable was created identically to the social involvement variable. As such, the respondents were divided into two groups: individuals who frequently obtain information from the outside world, and individuals who acquire information less frequently. The composite was generated using a consistent measure and included five variables: reading newspapers and magazines, reading books, using Internet to obtain information, watching news on the television, and going to the public library. For the associated formulas and details of this computation, see Appendix A.

Variables Measuring Career Outcomes

The variable measuring income levels was determined by the interview in 2000. This self-reported measure was tracked numerically. For the purposes of analysis, I created two subgroups, using the median salary as a binary split. As such, two groups were created, respondents with high incomes, and respondents with incomes less than the median yearly salary in 1999.

The second career outcome was a self-rating of job satisfaction based on eight categories. Job satisfaction consisted of binary answers to whether the respondent was satisfied with the following: fringe benefits, training opportunities, job security, pay, promotion opportunities, use of past training, work importance, and overall job satisfaction. Using these variables, I created a binary composite variable for the purpose of analysis that indicated either high or low job satisfaction. The computation of job satisfaction is detailed in Appendix A.

Independent Variables

Using the methodology described previously, I identified nine covariates in three categories: background factors, school factors, and college enrollment status. Figure 4 below summarizes the covariates used in the logistic regression models in this text. The technical aspects of the inclusion of each variable can be found in Appendix A.

	Binge Drinking in 2000	Cigarette Use in 2000	Yearly Earnings in 1999	Job Satisfaction in 2000	Voting 1998-99	Social Involvement in 2000	Informational Involvement in 2000
Gender	✓		✓				
Socioeconomic Status (1988)	✓		✓		✓		
Age					✓		
% Free Lunch (1988)	✓						
Repeated a Grade (1988)		✓					
Ever Named MVP (1992)						✓	
Ever Enrolled in AP Program (1992)							✓
In Academic School (2000)			✓	✓			✓
College Graduate		✓	✓	✓			✓

Figure 3. Usage of covariates in logistic regression models.

The three background factors were dummy coded for the purpose of analysis in the logistic regression. Gender was coded 1 = male, 2 = female. Socioeconomic status and age were coded into quarters. School factors included two types of variables. Whether the respondent repeated a grade, was in an AP program, and named MVP was coded into 1 = yes and 0 = no. Percent free lunch in the school was coded into subgroups by NELS. College enrollment status was coded into a binary 1 = yes and 0 = no.

It should be noted that not all variables were used for each model due to the inconsistence of impact on outcomes. For example, no school factors were associated with career outcomes and, at the same time, the college graduation status of the respondent had associations with outcomes in all three categories.

CHAPTER IV

DATA ANALYSIS

Introduction

The broad objective of this study was to determine whether large-scale assessment performance growth in the subject areas of math and reading are predictive of meaningful long-term outcomes in a student’s life. Meaningful long-term outcomes are identified and categorized into three domains: health, career, and societal outcomes. The use of logistic regression models, while controlling for background factors, calculates the degree to which test growth can predict meaningful outcomes at age 26 for the chosen respondent population.

Descriptive Statistics for Key Variables

Outcome Variables

Occurrences of binge drinking.

The generated dichotomous variable coded all values greater than or equal to one, as one. The 4,561 *none* responses and the 3,250 *legitimate skip* responses were coded as zero. None responses refer to individuals that claimed to have drunk alcohol in the previous month but did not binge drink in the past two weeks. The legitimate skip responses refer to individuals who claimed to have abstained from alcohol the entire previous month.

Table 2

Frequency of Coded Variable for Binge Drinking in 2000 (N = 10,604)

Response	*n*	%
No	7,811	72.1
Yes	2,793	25.8

The first output variable indicates binge drinking over the last two weeks of the interview. This variable was dichotomously coded to indicate at least one instance of binge drinking, which is consuming five or more alcoholic beverages in a row. This question was asked during the fourth follow up and coded as an ordinal value in F4IBINGE. Respondents who indicated that they did not drink alcohol on a prior question, F4IDRINK, were not asked this question.

Table 3

Frequency of Binge Drinking in Last Two Weeks in 2000 (N = 7,354)

Number of Occasions	*n*	%
None	4,561	62.0
1	1,340	18.2
2	733	10.0
3	262	3.6
4	189	2.6
5	106	1.4
6	46	0.6
7	24	0.3
8	24	0.3
9	3	0.0
10	66	0.9

Cigarette use.

The outcome variable for cigarette use was collected in 2000, when respondents were age 26. The variable question asked how many cigarettes the respondents smoked on a typical day.

Table 4 organizes this data to represent the frequency of cigarette use for the respondents. To answer the respective research question, I recoded the variable into two categories consisting of smokers and of non-smokers.

Table 4

Frequency of Cigarette Use in 2000 (N = 10,614)

Number of Occasions	*n*	%
I don't smoke cigarettes	8,060	75.9
Less than one cigarette a day	138	1.3
1-5 cigarettes a day	670	6.3
About half pack a day (10 cigarettes)	728	6.9
More than half and less than 2 packs	949	8.9
2 or more packs a day (40+ cigarettes)	69	0.7

Table 5

Frequency of Coded Variable for Cigarette Use in 2000 (N = 10,614)

Daily Cigarette Use	*n*	%
Does not smoke	8,060	75.9
Smokes daily	2,554	24.1

Yearly earnings.

In the 2000 questionnaire, respondents were asked to estimate their income for that year. This income included all wages, salaries, and commissions earned in 1999 before taxes and other deductions. This variable was recoded dichotomously, using median earnings for modeling and using logistic regression. The median earnings for the NELS:88/00 cohort was $24,000.

Table 6

Frequencies of Yearly Earnings in 1999 (N = 9,971)

Earned Income Level	*n*	%
Less than $9,999	1,598	16
$10,000 - $19,999	2,228	22
$20,000 - $29,999	2,677	27
$30,000 - $39,999	1,952	20
$40,000 - $49,999	835	8
More than $50,000	681	7

Job satisfaction.

To model job satisfaction, I compiled an index by combining eight separate binary questions of job satisfaction, using summation. The questions are detailed in the table below. The outcome of note, job satisfaction, was coded high job satisfaction if the respondent was satisfied in at least seven of the eight job satisfaction categories. These categories involved the respondent's satisfaction with (1) pay, (2) fringe benefits, (3) importance and challenge of the job, (4) opportunities for promotion and advancement, (5) opportunities to use prior training and education, (6) job security, (7) opportunities for further training and education, and (8) overall job satisfaction as a whole. Consistent with prior coding, this variable was recoded dichotomously, using the median job satisfaction composite.

Table 7

Frequency for Job Satisfaction Responses in 2000

Satisfaction Category	*n*	% Satisfied
Job Satisfaction Composite (% high satisfaction)	10,827	57
Fringe benefits	10,214	76
Further training	10,392	77
Job security	10,468	87
Overall job satisfaction	10,516	85
Pay	10,521	72
Promotion opportunity	10,262	70
Use of past training	10,421	79
Work importance	10,486	83

Voting habits.

Respondents were asked in 2000 about their voting habits in the prior 24 months. One critical point of analysis is that there were significant differences in voting indicators based on the type of voting habit being tracked. As shown in the table below, the majority of respondents did not vote in the two years prior to their interview. This is consistent with general voting habits for local elections in comparison to national elections.

Table 8

Frequencies of Voting Habit Indicator

Voting Category	*n*	% Yes
Voted in elections 1997 - 1999	11,897	41

Social integration.

The social integration variable is measured in times per month that the respondents participated in organized religious activities, playing groups, team sports, recreational activities, or concert or museum visitations. Volunteer hours were tracked in average hours per week over the last 12 months of the interview, which took place in 2000. For the purposes of using a regression model, a composite variable was created that used the median as the cutoff point for low and high social involvement.

Table 9

Descriptive Statistics of Social Involvement Indicators in 2000

Social Involvement Category	*M*	95% *CI*	*SD*	*Min*	*Max*
Social involvement composite	5.8	[5.6, 5.9]	7.1	0	64
Volunteer (hours/week)	1.2	[1.2, 1.3]	3.3	0	57
Organized religion (days/month)	2.5	[2.4, 2.6]	4.0	0	30
Attend plays, concerts (days/month)	1.2	[1.2, 1.3]	2.0	0	30
Participate in group sports (days/month)	2.6	[2.5, 2.7]	4.9	0	30

Note. Social involvement composite was computed through summation, using days per month attending organized religion, plays, concerts, group sports, and volunteering hours divided by four.

Table 10

Descriptive Statistics of Social Involvement Outcome Variable in 2000

Social Involvement Composite Category	*n*	%
Low social involvement	6,523	54
High social involvement	5,621	46

Information integration.

The informational involvement composite was a measure of the respondent's engagement in obtaining information from the outside world. Included in this measure were visits to the library, accessing the Internet for information, reading books, reading articles, reading newspapers or magazines, and watching news on television. As consistent with prior coding, this variable was dichotomously coded, using the median of the composite variable.

Table 11

Descriptive Statistics of Information Involvement Indicators in 2000

Information Involvement Category	*M*	95% *CI*	*SD*	*Min*	*Max*
Informational involvement composite	14.7	[14.6, 14.8]	5.8	0	33.5
Go to public library (days per month)	1.6	[1.5, 1.6]	3.2	0	30
Access Internet for information (week)	2.7	[2.6, 2.7]	2.5	0	7
Read books (week)	2.9	[2.8, 2.9]	2.5	0	7
Read papers or magazines (week)	4.1	[4.0, 4.1]	2.3	0	7
Watch TV news (week)	4.7	[4.7, 4.7]	2.3	0	7

Note. Information involvement composite is computed through summation of all sub-categories in this table.

Input Variables

To identify appropriate control variables for the regression models for each dependent variable, I ran a bivariate correlation for each dependent variable and sorted by the absolute correlation. I selected control variables from those variables with the highest correlations with the outcome variable while eliminating variables that were generated from one another when necessary. For example, family income in 1987 is a variable that is used in the calculation for SES. Therefore, in each case where family income was highly correlated with the outcome variable, so was SES. In every case, I selected the variable that had a higher correlation and excluded the other variable from the model. Further, for consistency, I selected variables that were highly correlated for multiple models. Therefore, the variable family income in 1987 was excluded from the analysis below due to the consistently high correlations of socioeconomic status in 1988 for three of the models.

As shown in the tables below, background information from the respondents was collected in several waves of surveys starting in 1988, when the respondent was in 8th grade. During that first wave, respondents were asked questions regarding their ethnicity, gender, and age.

Table 12

Descriptive Statistics of the Personal Characteristics of NELS:88/00 Respondents

Characteristic	*n*	%
Ethnicity		
American Indian or Alaska Native Respondents	94	1
Asian or Pacific Islander Respondents	558	5
Black, not Hispanic Respondents	895	9
White, not Hispanic Respondents	7,334	69
Hispanic or Latino Respondents	1,405	13
Respondents who are more than one race	295	3
Gender		
Male Respondents	5,056	47
Female Respondents	5,771	53
Age as of September 1988 (years)		
13.5 and younger	48	<1
13.6 - 14	1,399	13
14.1 - 14.5	4,326	41
14.6 - 15	3,239	31
15.1 - 15.5	909	9
15.6 - 16	464	4
16.1 and older	200	2

School factors were generated from several surveys and sources. The percentage of students receiving free lunch was derived from the school representative questionnaire in 1988. In this questionnaire, the students were also asked if they had repeated a grade prior to eighth grade. In 1992, the students were asked whether they had completed an Advanced Placement (AP) course or if they had been named the most valuable player (MVP) for a school sports team in 1992.

Table 13

Descriptive Statistics of the School Factors of NELS:88/00 Respondents

Factor	*n*	%
% Students Receiving Free Lunch in School (1988)		
None	1,621	15
1% – 5%	1,472	14
6% - 10%	1,155	11
11% - 20%	1,844	17
21% - 30%	1,521	14
31% - 50%	1,702	16
51% - 75%	921	9
76% - 100%	404	4
Repeated Grade		
No	8,762	86
Yes, repeated	1,438	14
Completed Advanced Placement (AP) Courses		
No	5,756	60
Yes, was in AP course	3,795	40
Named Most Valuable Player (MVP) on Sport Team		
No	8,441	89
Yes, was MVP	1,061	11

The tables below acknowledge critical covariates, gathered from the final round of surveys for NELS:88/00, that were used in the regression models for this study. The final round of responses included questions regarding the respondents' highest diplomas achieved and the activities in which they were taking part in 2000. For consistency of analysis, the variable representing the highest diploma earned was recoded into 1 = college graduate and 0 = did not complete college.

Table 14

Descriptive Statistics of Highest Earned Diploma in 2000

Education: Highest Diploma Achieved	*n*	%
No Diploma	501	5
GED	525	5

High school diploma	4,370	40
Associates/Certificate	1,629	15
Bachelors	3,355	31
Masters or above	447	4

Table 15

Descriptive Statistics of Respondent Activities in 2000

Activity	*n*	Yes %
Employed full time (*N* = 10,826)	8,291	77
Employed part time (*N* = 10,825)	1,875	17
In school (*N* = 10,825)	1,979	18
In vocational/tech school (*N* = 12,140)	843	7
Work experiences* (*N* = 12,141)	730	6
Keeping house (*N* = 1,673)	987	59
On leave/awaiting job (*N* = 2,866)	348	12
Job held within 12 months of interview (*N* = 10,667)	10,033	94
Average Hours Worked per Week (*N* = 10,827)		
Less than 20	1,716	16
21 - 40	5,255	49
41 - 60	3,300	30
61 - 80	556	5

* Work experiences include internships, job training, and apprenticeships

Testing Variables

The tables below show performance indicators for math and reading. Respondents completed three rounds of tests in math and reading periodically from 1988 to 1992. The student growth percentile (SGP) was generated from all three exams taken during the three questionnaire rounds; see Chapter III for more information on this technique. By design, SGP values are not correlated with prior student performance; therefore, they are ideal for use in regression models. SGP values were interpreted as the growth from the 1988 test to the 1992 test in relation to peer performance in 1988. That is, if a student scored 40 points on the original math test administered in 1988, he or she was compared, in 1992, only to those students who initially scored approximately 40 points in 1988. This measure is similar to using growth charts for heights of children. Since children of different initial heights tend to grow at different rates, they are grouped together with peers in a similar range in order to adequately measure growth while accounting for the initial height. This also minimizes floor and ceiling effects since a student who scored very low or high on an exam in 1988 was only compared to other students with similar test performance.

Table 16

Descriptive Statistics of Math and Reading Student Growth Percentiles for 1988 to 1992

Testing Variable	*n*	*M*	*SD*	Min	Max
Reading Student Growth Percentile (1988-92)	8,245	50	29	1	99
Math Student Growth Percentile (1988-92)	8,240	51	29	1	99

Data Analysis for Health Outcomes

Research Question 1: Are students with higher test performance growth less likely to be excessive in their use of alcohol?

The table below shows the results of the logistic regression models for whether the respondent drinks in excess. The table is disaggregated for ethnicity and then by gender. The results show that overall, math growth is not predictive of whether the respondent binge drinks. Conversely, there is evidence that growth in reading performance lowers the odds-ratio (*OR*) for the binge drinking. Further, upon analysis of the disaggregation, Simpson's paradox is evident. Even though reading growth is significant for the total population, when the data are broken down by ethnicity, only White, non-Hispanic respondents show a significant influence from reading growth. The technical appendix contains further details and SPSS outputs associated with the table below.

Table 17

Logistic Regression Models for Binge Drinking for Gender and Ethnicity

	Math		Reading	
	OR	Wald	*OR*	Wald
Overall				
Low Growth*		1.39		13.72
Typical Growth	*NS*	0.30	0.86	4.88
High Growth	*NS*	1.38	0.79	13.37
Ethnicity				
Asian or Pacific Islander				
Low Growth*		3.61		0.03
Typical Growth	*NS*	3.60	*NS*	0.00
High Growth	*NS*	0.82	*NS*	0.02
Black, not Hispanic				
Low Growth*		2.26		0.64
Typical Growth	*NS*	0.18	*NS*	0.64
High Growth	*NS*	1.19	*NS*	0.06
White, not Hispanic				
Low Growth*		0.31		12.03
Typical Growth	*NS*	0.05	*NS*	2.45
High Growth	*NS*	0.30	0.78	12.03
Hispanic or Latino				
Low Growth*		2.02		2.26
Typical Growth	*NS*	0.10	*NS*	2.26
High Growth	*NS*	1.11	*NS*	0.49
Gender				
Male				
Low Growth*		0.85		11.90
Typical Growth	*NS*	0.85	*0.81*	6.41
High Growth	*NS*	0.21	*0.77*	10.41
Female				
Low Growth*		6.63		3.79
Typical Growth	*NS*	3.79	*NS*	0.13
High Growth	*0.79*	5.49	*NS*	3.39

* Low growth is reference category for each model
NS = Odds ratio (OR) is not significant

Research Question 2: Do students with higher test performance growth tend to smoke less?

As consistent with prior outputs, I ran a series of logistic regression models to examine the impact that a respondent's growth on large-scale assessments has on cigarette use later in life. Interestingly, the table below shows evidence of significant impact of large-scale assessment growth on smoking habits later in life. However, there is also evidence that when the data are disaggregated, the significant impact disappears entirely for Asian and Black respondents. Furthermore, the impact of performance growth only held for female respondents when I broke the data down by gender.

Table 18

Logistic Regression Models for Cigarette Use for Gender and Ethnicity

	Math		Reading	
	OR	Wald	*OR*	Wald
Overall				
Low Growth*		15.48		7.14
Typical Growth	0.81	8.81	0.85	5.38
High Growth	0.79	13.10	0.86	4.99
Ethnicity				
Asian or Pacific Islander				
Low Growth*		1.26		0.51
Typical Growth	*NS*	1.21	*NS*	0.42
High Growth	*NS*	0.14	*NS*	0.36
Black, not Hispanic				
Low Growth*		0.93		1.98
Typical Growth	*NS*	0.93	*NS*	1.97
High Growth	*NS*	0.30	*NS*	0.31
White, not Hispanic				
Low Growth*		14.56		4.21
Typical Growth	0.81	6.74	*NS*	3.06
High Growth	0.75	13.25	*NS*	3.09
Hispanic or Latino				
Low Growth*		1.67		7.74
Typical Growth	*NS*	0.94	*0.66*	3.91
High Growth	*NS*	1.44	*0.58*	6.67
Gender				
Male				
Low Growth*		3.30		2.82

Typical Growth	*NS*	1.88	*NS*	1.51
High Growth	*NS*	2.94	*NS*	2.46
Female				
Low Growth*		18.10		3.92
Typical Growth	*0.74*	9.15	*NS*	3.29
High Growth	*0.69*	15.39	*NS*	2.41

* Low growth is reference category for each model
NS = Odds ratio (*OR*) is not significant

Data Analysis for Career Outcomes

Research Question 1: Do students with higher test performance growth tend to earn more money?

Recall that the earnings variable was collected in the year 2000. The question asked the respondent to state all before-tax earnings in 1999, including wages, salaries, and commissions that the respondent earned from employment. This allowed the respondent to select a value from $0 to $500,000. The table below displays information about the employment status of the participant pool in 2000.

Of the employed respondents in 2000, 794 had both part-time and full-time jobs. Of all part-time employees, 92% were either also employed full-time or also getting professional or academic experiences. Of the respondents who were employed full-time, 22% were engaged in academic school or other vocational pursuits.

Table 19

Current Activity and 2000 Work Status of Respondents

	Work Status		
Current Activity	Unemployed *n* (%)	Full Time *n* (%)	Part Time *n* (%)
Full-time job	-	-	794 (42%)
Part-time job	-	794 (10%)	-

Academic school	217 (34%)	1,038 (13%)	603 (32%)
Vocational/tech school	51 (8%)	495 (6%)	178 (9%)
Work experiences*	53 (8%)	375 (5%)	174 (9%)
Keeping house	405 (64%)	-	-
On leave/awaiting job	30 (5%)	-	96 (5%)
Number of respondents	633	8,291	1,875

Note. Percentages do not add up to 100% due to participants falling into multiple categories (i.e., respondents who work both full- and part-time jobs)
* Work experiences include internships, job training, and apprenticeships

Respondents who were employed full time earned approximately $9,000 more than individuals who were employed part-time. Recall that there was an overlap of 794 respondents that were both part- and full-time employed. There were clear interaction effects due to the nature of this conflation of work status.

Table 20

Income of respondent in 1999 by Work Status

Work Status	*M*	*SD*	*Mdn*	*n*
Unemployed	$932	$4,040	$0	634
Part-time job	$20,184	$13,835	$18,000	1875
Full-time job	$29,372	$19,694	$27,000	8291

Female respondents comprised 76% of the unemployed population. Female respondents were equally employed in full-time capacities as their male counterparts, and made up a higher proportion of the part-time employed respondent pool. In this dataset, male respondents represented a larger portion of job-specific work experiences such as internships, job training, and apprenticeships, and were less representative of the respondents who were in an academic school.

Table 21

Current Activity in 2000 by Gender

Current Activity	Male *n* (%)	Female *n* (%)
Unemployed	155 (24%)	479 (76%)

Part-time job	764 (41%)	1111 (59%)
Full-time job	4248 (51%)	4043 (49%)
Academic school	900 (45%)	1079 (55%)
Vocational/tech school	367 (50%)	372 (50%)
Work experiences	355 (57%)	266 (43%)
Keeping house	91 (11%)	758 (89%)
On leave/awaiting job	133 (43%)	173 (57%)

Asian or Pacific Islander respondents attended school at a rate of 31% of the population category. This was more than 10% higher than any other ethnic subgroup. At the same time, Asian or Pacific Islander respondents had the highest unemployment rate of any ethnic category. The employment descriptive statistics of Black, White, and Hispanic respondents were similar to one another. American Indian or Alaska Native respondents and respondents of more than one race were not included in the regression analysis splits due to the low sample sizes of 94 and 295, respectively.

Table 22

Employment Status in 2000 by Ethnicity

		% Population in Category			
Ethnicity	*n*	Unemployed	Part Time	Full Time	Academic School
American Indian or Alaska Native	94	11%	16%	72%	16%
Asian or Pacific Islander	558	11%	16%	70%	31%
White, not Hispanic	7,334	5%	17%	78%	17%
Black, not Hispanic	895	5%	17%	76%	18%
Hispanic or Latino	1,405	7%	17%	74%	21%
More than one race	295	7%	19%	68%	21%

The table below displays earnings means by gender and ethnicity. The table shows that female respondents earned $10,000 less than their male respondent counterparts. Given that the employment statistics are similar for male and female respondents, this table shows evidence of a difference of salary and hours worked for male and female respondents that have little to do with job status. Interestingly, since Asian and Pacific Islander respondents had the highest rate of unemployment of ethnicity subgroups in this study, the average earnings show evidence of selection bias due to the high mean and high standard deviation in relation to Black, White, and Hispanic respondents.

Table 23

Average 1999 Earnings by Gender and Ethnicity

Respondent Characteristic	*M*	95% *CI**	*SD*
Gender			
Male	$30,200	[$30, $31]	$22,700
Female	$20,600	[$20, $21]	$15,000
Ethnicity			
American Indian or Alaska Native	$19,200	[$16, $22]	$14,700
Asian or Pacific Islander	$28,300	[$26, $31]	$25,200
Black, not Hispanic	$21,500	[$21, $22]	$13,300
White, not Hispanic	$25,900	[$25, $26]	$20,000
Hispanic or Latino	$22,700	[$22, $24]	$20,300
More than one race	$22,000	[$20, $24]	$18,400

Note. All earnings levels are rounded to hundreds of dollars for readability.
* *CI* = Confidence interval for mean is rounded to the nearest thousand.

The table below displays average hours worked by gender and ethnicity. The table shows that female respondents worked approximately eight hours less per week than the average male respondent. Given that salary and pay are related, this table shows evidence that female respondents got fewer hours of work even though the employment status table above shows that female respondents were equally represented in the full-time working subgroup.

Table 24

Average Hours Worked per Week by Gender and Ethnicity

Respondent Characteristic	*M*	95% *CI*	*SD*
Gender			

Male	42.1	[41.9, 42.4]	16.9
Female	34.3	[34.1, 34.6]	17.7
Ethnicity			
American Indian or Alaska Native	36.1	[34.1, 38.1]	19.5
Asian or Pacific Islander	35.1	[34.2, 35.9]	20.1
Black, not Hispanic	38.3	[37.7, 38.9]	17.4
White, not Hispanic	38.6	[38.4, 38.8]	17.5
Hispanic or Latino	36.3	[35.8, 36.7]	17.8
More than one race	36.2	[35.1, 37.3]	19.1

Upon conducting a series of logistic regressions using the respondent's yearly income in 1999 as the dependent variable, I found no evidence supporting test performance growth influencing a respondent's income levels in 1999. Interestingly, this finding is inconsistent with the findings for health outcomes. Even upon disaggregation, there is no evidence to support that large-scale test performance growth influences income later in the respondent's life.

Table 25

Logistic Regression Models for Yearly Income in 1999 for Gender and Ethnicity

	Math		Reading	
	OR	Wald	*OR*	Wald
Overall				
Low Growth*		1.96		3.48
Typical Growth	*NS*	1.96	*NS*	0.07
High Growth	*NS*	0.47	*NS*	2.20
Ethnicity				
Asian or Pacific Islander				
Low Growth*		2.91		3.79
Typical Growth	*NS*	0.23	*NS*	2.14
High Growth	*NS*	2.63	*NS*	3.65
Black, not Hispanic				
Low Growth*		1.19		0.37
Typical Growth	*NS*	0.93	*NS*	0.08
High Growth	*NS*	0.80	*NS*	0.13
White, not Hispanic				
Low Growth*		2.55		2.89
Typical Growth	*NS*	2.54	*NS*	0.09

High Growth	*NS*	0.47	*NS*	1.71
Hispanic or Latino				
Low Growth*		1.04		0.25
Typical Growth	*NS*	0.28	*NS*	0.13
High Growth	*NS*	1.04	*NS*	0.02
Gender				
Male				
Low Growth*		0.13		1.80
Typical Growth	*NS*	0.07	*NS*	0.14
High Growth	*NS*	0.12	*NS*	0.94
Female				
Low Growth*		1.96		0.22
Typical Growth	*NS*	1.95	*NS*	0.99
High Growth	*NS*	0.29	*NS*	0.14

* Low growth is reference category for each model
NS = Odds ratio (OR) is not significant

Research Question 2: Are students with higher test performance growth more satisfied with their jobs?

Upon conducting a bivariate correlation of all available factors in the dataset, two of the following independent factors displayed as statistically significant and accounted for approximately 1% of variation in job satisfaction ratings while in school. Of the variables, whether the respondent was a college graduate and whether the respondent was in academic school in 2000 were sufficiently correlated with the job satisfaction composite.

Recall that the job satisfaction composite was generated as a sum of eight job satisfaction indicators. In Table 26, each job satisfaction indicator was cross-tabulated with the demographic factors of gender and ethnicity. In each cell, the percentage of satisfied respondents is shown with a parenthetical delta from the total respondent pool. The table below shows that in total, respondents were most satisfied with job security and least satisfied with promotion opportunities. For each of the categories of job satisfaction, female respondents were less satisfied except in the area where past training was useful for their current job. The biggest differences in satisfaction had to do with promotion opportunities, pay, and fringe benefits.

For the subgroups split by ethnicity, the size of the delta implied a deviation of job satisfaction from the aggregate population. American Indian respondents stated that they are 12% less satisfied with fringe benefits, while being 7% more satisfied with promotion opportunities and 6% more satisfied with work importance. For fringe benefits, American Indian, Black, and multiracial respondents were least satisfied with fringe benefits. Of all subgroups, Black, non-Hispanic respondents had the lowest job satisfaction responses for all eight categories.

Table 26

Job Satisfaction in 2000 by Ethnicity and Gender

Characteristic	Fringe benefits % (Δ%)	Further training % (Δ%)	Job security % (Δ%)	Overall job % (Δ%)	Pay % (Δ%)	Promotion % (Δ%)	Past training % (Δ%)	Work importance % (Δ%)
Total Respondent Pool	76	77	87	85	72	70	79	83
Gender								
Male (%)	78 (2)	79 (1)	88 (1)	87 (1)	75 (3)	73 (3)	79 (0)	84 (1)
Female (%)	74 (-2)	76 (-1)	87 (-1)	84 (-1)	69 (-3)	68 (-3)	79 (0)	82 (-1)
Ethnicity								
American Indian or Alaska Native (%)	64 (-12)	79 (2)	82 (-5)	86 (1)	71 (-1)	77 (7)	81 (2)	89 (6)
Asian or Pacific Islander (%)	76 (0)	79 (2)	86 (-1)	85 (0)	70 (-2)	71 (1)	78 (-2)	77 (-6)
Black, not Hispanic (%)	69 (-7)	70 (-7)	80 (-7)	76 (-9)	62 (-10)	64 (-7)	74 (-6)	78 (-5)
White, not Hispanic (%)	77 (2)	78 (1)	89 (2)	86 (1)	73 (1)	71 (1)	81 (1)	84 (1)
Hispanic or Latino (%)	74 (-1)	76 (-2)	85 (-2)	86 (0)	72 (0)	72 (1)	78 (-1)	82 (-1)
More than one race (%)	70 (-6)	75 (-3)	82 (-5)	76 (-9)	69 (-3)	65 (-6)	71 (-8)	77 (-6)

For the covariates, college graduation and not being in an academic school was associated with a positive delta for all job satisfaction measures. Being a college graduate creates the large gains in job satisfaction for fringe benefits and the opportunities for further training. Being in an academic school has the largest negative association with promotional opportunities and work importance.

Table 27

Job Satisfaction in 2000 for College Graduates and Those in Academic School

		College Graduate % (Δ%)		In Academic School % (Δ%)	
Satisfaction Category	% Satisfied	No (*n* = 7,025)	Yes (*n* = 3,802)	No (*n* = 8,846)	Yes (*n* = 1,979)
Fringe benefits	76%	73 (-3)	82 (6)	77 (1)	70 (-6)
Further training	77%	74 (-3)	82 (5)	78 (1)	74 (-3)
Job security	87%	86 (-1)	89 (2)	88 (1)	84 (-3)
Overall job satisfaction	85%	83 (-2)	88 (3)	86 (1)	81 (-4)
Pay	72%	72 (0)	73 (1)	73 (1)	67 (-5)
Promotion opportunity	70%	69 (-2)	74 (3)	72 (2)	63 (-8)
Use of past training	79%	77 (-2)	83 (4)	80 (1)	76 (-3)
Work importance	83%	82 (-1)	84 (1)	85 (2)	76 (-7)

I conducted the logistic regression models consistently with the previous dependent variables. The results for job satisfaction were similar to yearly earnings in that most associations were not significant. With that said, upon evaluating the subgroups, I found a surprising direction for two of the three significant odds-ratios. First, high growth in math performance seems to positively influence White, non-Hispanic respondents to be more satisfied with their jobs. The reverse is true for Black respondents and female respondents. Those respondents who had high reading performance growth in the aforementioned categories saw a drop in their job satisfaction later in life.

Table 28

Logistic Regression Models for Job Satisfaction for Gender and Ethnicity

	Math		Reading	
	OR	Wald	*OR*	Wald
Overall				
Low Growth*		2.61		3.75

Typical Growth	*NS*	0.00	*NS*	0.78
High Growth	*NS*	1.96	*NS*	3.74
Ethnicity				
Asian or Pacific Islander				
Low Growth*		1.72		0.87
Typical Growth	*NS*	0.30	*NS*	0.21
High Growth	*NS*	0.45	*NS*	0.86
Black, not Hispanic				
Low Growth*		1.80		4.60
Typical Growth	*NS*	0.35	*NS*	0.68
High Growth	*NS*	0.67	0.65	4.59
White, not Hispanic				
Low Growth*		6.16		0.94
Typical Growth	*NS*	0.03	*NS*	0.15
High Growth	1.15	4.37	*NS*	0.93
Hispanic or Latino				
Low Growth*		2.07		1.34
Typical Growth	*NS*	2.06	*NS*	0.90
High Growth	*NS*	0.39	*NS*	1.07
Gender				
Male				
Low Growth*		1.38		2.41
Typical Growth	*NS*	1.37	*NS*	2.33
High Growth	*NS*	0.42	*NS*	0.19
Female				
Low Growth*		3.46		7.15
Typical Growth	*NS*	0.42	*NS*	0.14
High Growth	*NS*	3.42	0.86	4.39

* Low growth is reference category for each model
NS = Odds ratio (*OR*) is not significant

Data Analysis for Societal Outcomes

Research Question 1: Are students with higher test performance growth more likely to vote?

Voting behavior was measured by a proxy variable of whether the respondent voted in the last 24 months prior to the 2000 survey. Upon conducting a Pearson bivariate correlation with the aforementioned variables, three inputs were flagged as significantly correlating with voting behavior. These factors included (1) whether or not the respondent was a college graduate in 2000, (2) the socioeconomic status (SES) composite of the respondent in 1988.

Table 29

Frequencies for Testing Variables as a Function of Voting in 1999 or 1998

	Did not vote		Voted	
	M	*SD*	*M*	*SD*
Student Growth Percentile (SGP)				
Math (8^{th} - 12^{th} Grades)	51	29	52	29
Reading (8^{th} - 12^{th} Grades)	50	29	50	29

In the table below, a cross-tabulation of the covariates and the voting indicator is shown. Most notably, college graduates and those respondents at the highest SES quartile in 1988 had the highest voting percentages. Interestingly, the age of the respondent did not seem to represent a linear relation. The first three quartiles of age voted at the same rates, while the oldest quartile voted at the lowest rate.

Table 30

Voting Indicator as a Function of College Graduation, SES in 1988, and Respondent Age

	Voted in 1998 or 1999 election
College graduate (*N* = 10,827)	
No	39%
Yes, graduated	51%
SES in 1988 (*N* = 10,827)	
Quartile 1 (low)	34%
Quartile 2	42%
Quartile 3	45%
Quartile 4 (high)	49%
Age (*N* = 10,585)	
Quartile 1 (youngest)	46%
Quartile 2	45%
Quartile 3	45%

Quartile 4 (oldest)	34%

Table 31

Voting Indicators as a Function of Gender and Ethnicity

	Voted in 1998 or 1999 election
Gender	
Male	56%
Female	59%
Ethnicity	
American Indian or Alaska Native	65%
Asian or Pacific Islander	42%
Black, not Hispanic	60%
White, not Hispanic	61%
Hispanic or Latino	49%
More than one race	55%

The table below shows the results of logistic regression models run using the same technique as prior dependent variables. The table shows that for all subgroups test performance growth does not significantly impact voting habits later in life. Interestingly, there is one exception for Black respondents. For respondents who identified themselves as Black, typical growth in math performance lowered the odds-ratio of voting later in life by 37%.

Table 32

Logistic Regression Models for Voting Habits for Gender and Ethnicity

	Math		Reading	
	OR	Wald	*OR*	Wald
Overall				
Low Growth*		0.40		0.55
Typical Growth	*NS*	0.39	*NS*	0.01
High Growth	*NS*	0.05	*NS*	0.48
Ethnicity				
Asian or Pacific Islander				

Low Growth*		2.48		2.56
Typical Growth	*NS*	0.11	*NS*	1.88
High Growth	*NS*	1.17	*NS*	0.03
Black, not Hispanic				
Low Growth*		4.97		1.22
Typical Growth	*0.63*	4.97	*NS*	0.93
High Growth	*NS*	0.86	*NS*	0.80
White, not Hispanic				
Low Growth*		0.56		0.12
Typical Growth	*NS*	0.03	*NS*	0.00
High Growth	*NS*	0.52	*NS*	0.07
Hispanic or Latino				
Low Growth*		1.20		0.59
Typical Growth	*NS*	0.41	*NS*	0.53
High Growth	*NS*	0.21	*NS*	0.02
Gender				
Male				
Low Growth*		0.53		0.67
Typical Growth	*NS*	0.06	*NS*	0.49
High Growth	*NS*	0.50	*NS*	0.49
Female				
Low Growth*		1.05		2.95
Typical Growth	*NS*	0.82	*NS*	0.83
High Growth	*NS*	0.71	*NS*	2.95

* Low growth is reference category for each model.
NS = Odds ratio (*OR*) is not significant.

Research Question 2: Are students with higher test performance growth more likely to spend time socializing with others?

Upon conducting a bivariate correlation including the dependent variable, social interaction level, I flagged two variables as significantly correlated with social interaction signals: whether the respondent was (1) a college graduate in 2000, and (2) the most valuable player (MVP) on a sports team in 1992. Recall that the social interaction levels were coded from the median of the social interaction index that contained the summation of hours volunteered, days attending organized religious services, days going to plays or concerts, and days participating in group sports or recreational activities.

Table 33

Mean Time of Social Involvement and Standard Deviations as a Function of 4-year College Degree

	Graduated College[a]		Not College Graduate[b]	
Social Involvement Category	*M*	*SD*	*M*	*SD*
Volunteer (hours/week)	1.39	3.07	1.13	3.42
Organized religion (days/month)	2.61	3.92	2.4	3.95
Attend plays, concerts (days/month)	1.53	2.16	1.1	1.9
Participate in group sports (days/month)	2.5	4.71	2.58	4.96

[a]*n* = 3,766. [b]*n* = 6,838.

Table 34

Mean Time of Social Involvement and Standard Deviations as a Function of Being Named MVP on a Sports Team in 1992

	Named MVP[a]		Not named MVP[b]	
Social Involvement Category	*M*	*SD*	*M*	*SD*
Volunteer (hours/week)	1.61	3.65	1.20	3.22
Organized religion (days/month)	2.48	3.53	2.51	3.94
Attend plays, concerts (days/month)	1.47	2.04	1.25	2.02
Participate in group sports (days/month)	4.77	6.42	2.25	4.47

[a]*n* = 1,038. [b]*n* = 8,287.

A cross-tabulation was conducted for each social involvement indicator and respondent demographics of gender and ethnicity. Male respondents participated more frequently in sports and, similarly, for the other three indicators. Of the subgroups split by ethnicity, Black respondents attended religious events most frequently, almost four days a month, while Asian respondents attended religious activities least often, slightly less than two days a month. All other social involvement indicators are similar between ethnic subgroups.

Table 35

Mean Time of Social Involvement and Standard Deviations as a Function of Gender and Ethnicity

	n	Volunteer[a] *M* (*SD*)	Religion[b] *M* (*SD*)	Concerts[b] *M* (*SD*)	Sports[b] *M* (*SD*)
Gender					
Male respondents	4,952	1.4 (3.9)	2.2 (3.8)	1.3 (2.2)	3.5 (5.5)
Female respondents	5,652	1.1 (2.7)	2.7 (4.1)	1.2 (1.8)	1.8 (4.1)
Ethnicity					
American Indian or Alaska Native respondents	91	1.5 (3.4)	2.3 (3.5)	1.6 (2.8)	2.9 (4.9)
Asian or Pacific Islander respondents	558	1.2 (2.8)	1.8 (3.4)	1.3 (1.9)	2.7 (5.2)
Black, not Hispanic respondents	888	1.3 (3.3)	3.9 (5.4)	1.2 (1.8)	2.9 (5.5)
White, not Hispanic respondents	7,320	1.2 (3.3)	2.4 (3.7)	1.2 (2.0)	2.5 (4.7)
Hispanic or Latino respondents	1,404	1.2 (3.1)	2.4 (3.9)	1.3 (2.2)	2.8 (5.2)
Respondents that are more than one race	293	1.6 (4.4)	2.2 (4.0)	1.1 (1.6)	2.8 (5.0)

[a] Volunteer hours were tracked as the average hours per week respondent volunteered over the last year (2000).
[b] Organized religious attendance, concert and play attendance, and group sports participation are measured in days per month.

Table 36

Input and Characteristic Factors as a Function of Social Involvement Composite Category

Input Factor	*n*	% high social involvement
Gender		
Male respondent	5,056	56%
Female respondent	5,771	49%
Ethnicity		
American Indian or Alaska Native respondent	94	56%
Asian or Pacific Islander respondent	558	47%
Black, not Hispanic respondent	895	63%
White, not Hispanic respondent	7,334	52%
Hispanic or Latino respondent	1,405	52%
Respondent of more than one race	295	50%
College graduate		

Yes, graduated	3,802	58%
No	7,025	48%
Named MVP in 1992		
Yes, named MVP	1,061	68%
No	8,441	51%

The table below shows a series of logistic regression models conducted on the dependent variable that represented the amount of time a respondent spends socializing with others at age 26. The data show that overall, only typical growth in math performance is positively associated with frequent social interaction later in life. This series of models was run consistently with all prior outcomes. The data show that social interaction later in life is predictive for those respondents with typical math performance growth. Upon disaggregating the data, I found a consistent significance for two subgroups: White respondents and male respondents. For those subgroups, typical math performance growth increased the likelihood of high social interaction later in life.

Table 37

Logistic Regression Models for Social Interaction for Gender and Ethnicity

	Math		Reading	
	OR	Wald	*OR*	Wald
Overall				
Low Growth*		5.53		1.23
Typical Growth	1.15	5.51	*NS*	0.51
High Growth	*NS*	1.58	*NS*	0.15
Ethnicity				
Asian or Pacific Islander				
Low Growth*		0.79		2.16
Typical Growth	*NS*	0.20	*NS*	1.42
High Growth	*NS*	0.79	*NS*	0.00
Black, not Hispanic				
Low Growth*		1.04		1.26
Typical Growth	*NS*	0.16	*NS*	0.45
High Growth	*NS*	1.03	*NS*	0.26

White, not Hispanic				
Low Growth*		3.91		0.51
Typical Growth	1.15	0.16	*NS*	0.50
High Growth	*NS*	1.03	*NS*	0.09
Hispanic or Latino				
Low Growth*		2.90		0.89
Typical Growth	*NS*	2.42	*NS*	0.18
High Growth	*NS*	1.87	*NS*	0.89
Gender				
Male				
Low Growth*		4.99		0.03
Typical Growth	1.19	3.85	*NS*	0.03
High Growth	*NS*	0.01	*NS*	0.00
Female				
Low Growth*		1.45		2.47
Typical Growth	*NS*	1.11	*NS*	1.25
High Growth	*NS*	1.00	*NS*	0.14

* Low growth is reference category for each model.
NS = Odds ratio (*OR*) is not significant.

Research Question 3: Are students with higher test performance growth more likely to spend time getting information about the outside world?

Upon conducting a bivariate correlation including the outcome variable, informational involvement composite, I flagged four variables as significantly correlated with informational involvement: (1) whether the respondent was a college graduate in 2000, (2) the socio-economic status (SES) composite of the respondent in 1988, (3) whether the respondent was in academic school, and (4) whether the respondent was ever in an Advanced Placement (AP) program.

Respondents who were in academic school in 2000 tended to go to the library, access the Internet, and read books more than respondents who were not in an academic school. Also, respondents who were in the AP program tended to read books and access the Internet for information more frequently than respondents who never participated in the AP program. In regard to SES, respondents who had higher SES, on average, accessed the Internet and read books more frequently than those who had lower SES. Interestingly, the reverse is true for watching news on television. Respondents with a higher SES quartile tended to watch TV news more frequently.

Table 38

Means and Standard Deviations of College Graduation, School Status in 2000, AP Program Status in 1992, and SES in 1988 as a Function of Informational Involvement Variables in Days per Week

	n	Go to Library *M* (*SD*)	Access Internet *M* (*SD*)	Read Books *M* (*SD*)	Read Newspapers *M* (*SD*)	Watch TV News *M* (*SD*)
College graduate						
Yes, graduated	6,820	0.4 (0.8)	2.3 (2.5)	2.6 (2.5)	4.0 (2.3)	4.8 (2.4)
No	3,770	0.4 (0.8)	3.3 (2.4)	3.3 (2.4)	4.1 (2.3)	4.5 (2.3)
Currently in school (2000)						
Yes, in academic school	1,952	0.7 (1.2)	3.5 (2.4)	3.7 (2.5)	4.2 (2.3)	4.5 (2.4)
No	8,637	0.3 (0.7)	2.5 (2.5)	2.7 (2.4)	4.0 (2.3)	4.7 (2.3)
Was in AP program						
Yes, in AP program	3,749	0.4 (0.9)	3.2 (2.5)	3.3 (2.5)	4.1 (2.3)	4.5 (2.4)
No	5,615	0.4 (0.8)	2.5 (2.5)	2.6 (2.4)	4.1 (2.3)	4.8 (2.3)
Socio-economic status						
Quartile 1 (low)	2,557	0.4 (0.9)	1.9 (2.4)	2.5 (2.4)	4.0 (2.3)	5.0 (2.3)
Quartile 2	2,529	0.4 (0.8)	2.5 (2.5)	2.7 (2.5)	4.0 (2.3)	4.9 (2.3)
Quartile 3	2,627	0.4 (0.8)	2.9 (2.5)	2.9 (2.5)	4.0 (2.3)	4.6 (2.3)
Quartile 4 (high)	2,877	0.4 (0.8)	3.3 (2.5)	3.4 (2.4)	4.2 (2.3)	4.3 (2.4)

Table 39

Means and Standard Deviations of Gender and Ethnicity as a Function of Informational Involvement Variables in Days per Week

	n	Go to Library *M* (*SD*)	Access Internet *M* (*SD*)	Read Books *M* (*SD*)	Read Newspapers *M* (*SD*)	Watch TV News *M* (*SD*)
Gender						

Male respondent	4,948	0.3 (0.8)	3.0 (2.6)	2.4 (2.4)	4.3 (2.3)	4.6 (2.4)
Female respondent	5,642	0.5 (0.9)	2.4 (2.5)	3.2 (2.5)	3.9 (2.3)	4.8 (2.3)
Ethnicity						
American Indian or Alaska Native respondent	94	0.6 (1.2)	2.0 (2.5)	2.7 (2.3)	4.2 (2.4)	4.9 (2.5)
Asian or Pacific Islander respondent	557	0.4 (0.9)	3.6 (2.5)	2.7 (2.4)	4.1 (2.3)	4.6 (2.3)
Black, not Hispanic respondent	886	0.6 (1.0)	2.2 (2.4)	2.9 (2.4)	4.3 (2.3)	5.4 (2.1)
White, not Hispanic respondent	7,307	0.3 (0.7)	2.7 (2.5)	2.9 (2.5)	4.0 (2.3)	4.5 (2.4)
Hispanic or Latino respondent	1,402	0.5 (1.0)	2.3 (2.5)	2.8 (2.5)	4.0 (2.3)	5.1 (2.2)
Respondent of more than one race	294	0.5 (1.0)	2.5 (2.5)	3.0 (2.5)	4.2 (2.3)	4.9 (2.4)

Table 40

Composite Mean of Days per Week Respondent Obtained Outside Information in 2000 by College Graduation Status*

Control Variables	*n*	*M*	95% *CI*	*SD*
College graduate				
Yes, graduated	6,820	14.3	[14.1, 14.4]	5.8
No	3,770	15.5	[15.3, 15.7]	5.6
Currently in school (2000)				
Yes, in academic school	1,952	14.3	[14.1, 14.4]	5.6
No	8,637	16.6	[16.3, 16.8]	5.8
Was in AP program (1992)				
Yes, in AP program	3,749	14.3	[14.2, 14.5]	5.7
No	5,615	15.4	[15.3, 15.6]	5.7
Socio-economic status (1988)				
Quartile 1 (low)	2,557	13.8	[13.6, 14.0]	5.7
Quartile 2	2,529	14.5	[14.2, 14.7]	5.9
Quartile 3	2,627	14.8	[14.5, 15.0]	5.6
Quartile 4 (high)	2,877	15.6	[15.4, 15.8]	5.7

* The composite mean is the summation of five informational indicators of access per week: library use, Internet access for information, reading books, reading newspapers, and watching news on TV.

Table 41

Composite Mean of Days per Week Respondent Obtained Outside Information in 2000 by Gender and Ethnicity*

Characteristic Variables	*n*	*M*	95% *CI*	*SD*

Gender				
Male respondent	4,948	14.6	[14.5, 14.8]	5.7
Female respondent	5,642	14.7	[14.6, 14.9]	5.8
Ethnicity				
American Indian or Alaska Native respondent	94	14.4	[13.2, 15.6]	5.8
Asian or Pacific Islander respondent	557	15.5	[15.0, 16.0]	5.8
Black, not Hispanic respondent	886	15.4	[15.0, 15.8]	6.0
White, not Hispanic respondent	7,307	14.5	[14.4, 14.6]	5.7
Hispanic or Latino respondent	1,402	14.7	[14.4, 15.0]	5.8
Respondent of more than one race	294	15.0	[14.4, 15.7]	5.8

* The composite mean is the summation of five informational indicators of access per week: library use, Internet access for information, reading books, reading newspapers, and watching news on TV.

Table 42

Composite Mean of Days per Week Respondent Obtained Outside Information in 2000 by Testing Variables*

Testing Variables	*n*	*M*	95% *CI*	*SD*
Math (1988-1992)				
Low Growth	2,895	14.6	[14.4, 14.8]	5.7
Typical Growth	2,267	14.7	[14.5, 15.0]	5.6
High Growth	3,078	14.8	[14.6, 15.0]	5.8
Reading (1988-1992)				
Low Growth	2,878	14.7	[14.5, 15.0]	5.7
Typical Growth	2,395	14.6	[14.3, 14.8]	5.8
High Growth	2,972	14.8	[14.6, 15.0]	5.7

* The composite mean is the summation of five informational indicators of access per week: library use, Internet access for information, reading books, reading newspapers, and watching news on TV.

The table below represents the results from the logistic regression runs on the composite dependent variable that represents the respondent's activities related to obtaining information from the outside world. Running the analysis consistently with the prior dependent variables, I found that test performance growth had, by and large, no effect on whether the respondent was actively obtaining information from the outside world. One subgroup, however, saw a significant negative impact from typical growth on the large-scale math tests. Black respondents who showed typical growth on math performance were 43% less likely to be actively obtaining information from the outside world compared to their peers.

Table 43

Logistic Regression Models for Informational Involvement for Gender and Ethnicity

	Math		Reading	
	OR	Wald	*OR*	Wald
Overall				
Low Growth*		0.04		0.37
Typical Growth	*NS*	0.00	*NS*	0.16
High Growth	*NS*	0.00	*NS*	0.35
Ethnicity				
Asian or Pacific Islander				
Low Growth*		0.15		0.34
Typical Growth	*NS*	0.15	*NS*	0.33
High Growth	*NS*	0.04	*NS*	0.07
Black, not Hispanic				
Low Growth*		7.35		2.67
Typical Growth	0.57	6.89	*NS*	1.96
High Growth	*NS*	3.32	*NS*	0.02
White, not Hispanic				
Low Growth*		1.12		0.38
Typical Growth	*NS*	0.44	*NS*	0.20
High Growth	*NS*	0.15	*NS*	0.35
Hispanic or Latino				
Low Growth*		3.10		1.85
Typical Growth	*NS*	0.03	*NS*	0.08
High Growth	*NS*	2.59	*NS*	1.70
Gender				
Male				
Low Growth*		1.47		0.52
Typical Growth	*NS*	1.45	*NS*	0.34
High Growth	*NS*	0.28	*NS*	0.41
Female				
Low Growth*		0.88		0.02
Typical Growth	*NS*	0.85	*NS*	0.01
High Growth	*NS*	0.33	*NS*	0.02

* Low growth is reference category for each model
NS = Odds ratio (OR) is not significant

Summary of Findings

Based on the data presented in this chapter, it is clear that there are inconsistent effects of large-scale assessment performance growth on long-term outcomes. Upon conducting subgroup analysis, trends emerged that were not present when viewing the population in aggregate. I discuss the findings in three phases. First, I discuss the aggregate influence of performance growth on outcomes. Afterwards, the subgroups split by ethnicity and gender were considered as the underlying trend of the overall findings.

In aggregate, performance growth in early years can be predictive of health outcomes. Specifically, having low growth increased a respondent's likelihood of smoking cigarettes by approximately 20% and by 15% for math and reading, respectively. For binge drinking, math was not a significant predictor, yet low growth in reading increased the odds of a respondent binge drinking by about 15% to 20%. Career-related outcomes were not significantly related to performance growth. Further, societal outcomes including voting habits, social interaction, and information acquisition were by-and-large not influenced by performance growth. There was one exception, however, for social interaction. Respondents who had typical growth in their math assessments were 15% more likely to be highly socially engaged.

The following summaries regarding ethnicity and gender are split by each subgroup. This is consistent with the methodology used for the regression model splits. The logistic regression was run independently for each subgroup. Therefore, the following results should be understood as the effect of large-scale assessment growth within each subgroup. This distinction is important since subgroups were not compared, as the usual method of dummy coding requires. Rather, each logistic regression was run for effects within the subgroups.

For all of the long-term outcomes, the performance growth of one subgroup, Asian or Pacific Islander respondents, showed absolutely no impact for all seven long-term outcomes. For Black respondents, performance growth indicated limited effects for long-term outcomes. Only growth in mathematics was indicative of outcomes later in life. Black respondents who had high growth on reading tests were 35% less likely to be highly satisfied with their jobs. Further, Black respondents who exhibited typical growth in math were 37% less likely to vote in their local elections and 43% less likely to be highly involved with accessing information later in life. For White respondents, four of the seven outcomes saw predictive powers from performance growth. Indeed, White respondents who had high reading performance growth demonstrated a 22% drop in binge drinking. For that same subgroup of respondents, low math performance growth indicated approximately a 20% rise in the chance of smoking cigarettes daily. White respondents who had high growth on math tests were 15% more likely to be highly satisfied with their jobs. White respondents with typical growth on mathematics tests saw a 15% increase in the likelihood of identifying themselves as highly social later in life. Hispanic or Latino respondents displayed only one outcome for which performance growth was predictive. For respondents with low growth in reading, the probability of smoking cigarettes daily increased by approximately 60%.

Binge drinking and social involvement were the only outcomes for which male respondents found performance growth predictive. For binge drinking, low growth in reading increased the chance of binge drinking by approximately 20%. Further, male respondents who experienced typical growth in math test performance saw a 19% increase in being highly socially involved. For female respondents, performance growth was predictive of binge drinking, cigarette use, and job satisfaction. Female respondents with high growth on math tests were 21% less likely to participate in binge drinking. Further, female respondents who performed with low growth on math tests were approximately 72% more likely to smoke cigarettes. Additionally, female respondents who exhibited high growth in reading saw a 14% drop in the odds of being highly satisfied with their jobs.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

Overview of the Study

In this study, I employed logistic regression analysis to discover the predictive power of early large-scale assessment growth on later life outcomes, using data from the NELS:88/00 study. More specifically, I examined the relationship of large-scale assessment performance, and the subsequent growth in performance, of students in the eighth grade between the years of 1988 and 1992 with their life outcomes at the age of 26. Outcomes included specific behaviors and attainments that were assessed in the NELS:88/00 dataset and were representative of three major life domains: health, career, and societal involvement. Measured outcomes under the domain of health include habits of excessive drinking and smoking. Outcomes under the domain of career include annual income earned and overall job satisfaction. Outcomes under the domain of societal involvement include voting habits, social involvement, and frequency in obtaining information from the outside world. These outcomes were chosen because they also represent the greater ambitions of academia and the development of well-adjusted adults.

Conclusions on the Predictive Power of Large-Scale Assessments Performance on Health Outcomes

Health Outcome: Binge Drinking at Age 26

According to my analysis, math growth is not predictive of the rate of binge drinking at age 26. Reading performance growth is predictive for respondents in aggregate. Upon disaggregating the population into subgroups, I found that only White respondents and male respondents had reading growth as predictive for binge drinking later in life. High reading growth is associated with a 23% decrease in the probability of a White respondent binge drinking. Further, male respondents who exhibited low growth in reading performance saw an approximately 20% increase in binge drinking behavior later in life.

Background variables were the most significant predictors of whether a respondent will binge drink. According to logistic regression odds-ratios, male respondents are over three times more likely to binge drink than female respondents. Furthermore, respondents whose socioeconomic status was in the third quarter were approximately 27% more likely to binge drink in 2000. These results indicate that the relationship between the socioeconomic status of the respondents in 1988 and binge drinking later in life was not linear. Respondents who attended schools with high poverty populations were approximately 10% to 40% less likely to binge drink than their peers who went to schools with populations that were higher in poverty level. This finding implies that respondents who have higher financial means early on are more likely to binge drink later in their life.

Health Outcome: Smoking Cigarettes at Age 26

According to the results detailed in Chapter IV, low performance growth had a significant impact on the odds that the participants smoke cigarettes daily. Specifically, respondents who displayed low growth in math and reading saw increased probabilities of 20% and 15%, respectively, for being a daily smoker. When I disaggregated the population into subgroups, the same trend did not hold. Male respondents, as well as Asian and Black respondents, did not see significant impact from large-scale assessment performance growth. Low growth in math increased the odds of smoking by about 20% and 27% for White respondents and female respondents, respectively. Furthermore, low reading growth was associated with an approximately 40% increase for Hispanic or Latino respondents.

Consistent with the findings of Lantz et al. (1998), I found that the number of years of schooling an individual receives relates to his or her cigarette smoking habits at age 26. Respondents who were college graduates were 65% less likely to smoke cigarettes daily. College graduation was by far the strongest predictor included in the model with a χ^2 (1, 7701) = 246.4. For the full SPSS outputs, consult Appendix E.

The findings associated with cigarette use support the signaling theory of education. That is, a respondent's growth in skills in math or reading, as determined by large-scale assessments, is a significantly weaker predictor of his or her use of cigarettes than college graduation status. Furthermore, even though performance growth is a stronger predictor than whether the respondent repeated a grade, the magnitude of the predictive value provided by college graduation status dwarfs the predictive power of performance growth. As such, the amount that is learned during school is a weak predictor of smoking habits in later life when compared to college graduation status. The signals of educational attainment, and not growth on performance for large-scale assessments, significantly predict the choices that a respondent makes in regard to smoking at age 26. Respondents who have higher degrees are less likely to engage in cigarette use.

The relation between degrees attained and cigarette use implies one of two possibilities: (1) there are skills and dispositions that a respondent obtains from more years of education that lower the likelihood of cigarette use or that (2) respondents who obtain higher academic degrees have the predisposed skills and dispositions to lower the likelihood of cigarette use.

The Predictive Power of Large-Scale Assessment Growth on Career Outcomes

Career Outcome: Yearly Income at Age 25

Test performance growth did not significantly impact yearly income later in life. In fact, this is one of the few outcomes for which performance growth showed no impact even when I disaggregated data into subgroups. The findings that I present contradict the existing body of research regarding the impact of large-scale assessment performance on later income (Hanushek, 2009b; Krueger, 2003; Lazear, 2003; Mulligan, 1999; Murnane et al., 2000). I found no significant impact of performance growth on income later in life. The aforementioned research may be skewed due to the difficulty of separating performance on large-scale assessments and the test taker's background characteristics. However, due to the nature of Student Growth Percentiles (SGP), I was able to control for initial test performance when comparing respondents.

Using growth measures instead of performance subverts several studies that make claims about the absolute relationship of education, test performance, and later earnings in life, particularly the findings from Haveman and Wolfe's (1995) study. Although aggregate data hold true for some groups, it is not true for all. For example, college graduation does not have the same benefits for all respondent subgroups. In aggregate, a respondent increases the odds of having a higher than median salary by 250% if he or she is a college graduate. However, when considering male respondents only, that same increase is cut dramatically, only increasing the odds of having a higher than median salary by about 55%. In contrast, for female respondents, this same graduation status increases the odds of having a salary higher than the median by about 360%. Indeed, distinctive demographic subgroups have surprisingly dissimilar predictors of later income based on college education.

Haveman and Wolf found that once a student's ability and education level is controlled, parent variables have little influence on later income. In contrast, I found that the socioeconomic status (SES) of the family in 1988 has a similar complexity in predictive measure to college graduation. Interestingly, SES has little or no predictive value for all ethnic groups except for White respondents. For White respondents, high SES increases the odds of earning above the median salary by about 50%.

This study, furthermore, adds to findings from Lazear's (2003) study. He estimates 12% income gains for each standard deviation in test score gains. Using the same dataset at Lazear's, the findings in this study further break down and show a more nuanced impact of large-scale assessment performance on later earnings. An analysis contests that the Simpson's paradox had skewed the results of Lazear's study due to the use of aggregates of both test scores and populations, as mentioned previously. Indeed, not only does the use of performance instead of growth measures conflate the impact of large-scale assessments on later outcomes, but also the aggregate measures disguise the true predictive impact of variables on later life outcomes.

Career Outcome: Job Satisfaction at Age 26

Similar to earnings, I found that large-scale assessment performance growth is not predictive of high self-rated job satisfaction. Only for three subgroups does high growth matter. In two of the cases, I found the reverse of the expected value for high growth, which would logically be viewed as a positive experience in a student's life. Indeed, Black respondents and female respondents saw a dramatic drop of 35% and 14% in the likelihood that they would be highly satisfied with their jobs later in life if they had high growth in reading.

On the other hand, graduating from college increases the odds by about 33% that a respondent will be highly satisfied with his or her job. At the same time, being enrolled in an academic school in 2000 decreased the likelihood that a respondent would be satisfied with his or her job by 43%. Interestingly, Black respondents and Asian respondents saw no significant impact from college graduation on being highly satisfied with their jobs. A further complication in disaggregating the population found that only Black respondents saw no negative impact from attending an academic school on their likelihood of job satisfaction. These findings indicate some support for signaling theory and shows unambiguous evidence for the presence of Simpson's paradox. There are clear benefits of college graduation for certain subgroups, and not for others. Furthermore, skills displayed from large-scale assessment performance have little to no impact on job satisfaction other than some counterintuitive negative associations for high performance growth in reading. This finding implies that the signal of graduating college, and not math and reading test performance determined early on, provides some, but not all, respondents with the opportunity to obtain jobs that respondents find more satisfying on various levels including, but not limited to, pay, fringe benefits, security, importance, and promotional opportunities.

The Predictive Power of Large-Scale Assessment Performance Growth on Societal Outcomes

Societal Outcome: Voting Behavior in 2000

In this study, voting in local elections in 24 months prior to the 2000 interview was used as a proxy measure for voting behavior. Large-scale assessment growth variables had no predictive power to the voting indicator, except for one subgroup described in the previous chapter. Higher socioeconomic status in 1988 increased the likelihood of voting by about 40%. When I split age into quarters, only the oldest group was 23% less likely to vote. College graduation status increased the likelihood of voting by 29%. Thus, I find that large-scale assessment performance growth in math and reading do not significantly impact voting habits when compared to background variables and the signal of college graduation. This implies that the cognitive abilities, as measured by the large-scale assessments in math and reading, do not predict voting habits later in life. At the same time, college graduation, a signal of educational attainment, shows that respondents who graduated from college may be predisposed to voting more frequently than their peers who did not graduate from college.

Interestingly, as was consistent with prior findings, not all subgroups saw the same impact from background variables. Asian respondents were the only group for whom college graduation status was not a significant predictor of voting habits later in life. Indeed, the predictive value of the other background variables showed surprising inconsistency for all subgroups. For a detailed report of all logistic regression outputs, see Appendix E.

Societal Outcome: Social Involvement in 2000

Large-scale assessment performance growth in math and reading, measured early on, was not an important predictor, overall, of whether a respondent was highly involved in community activities at age 26. Interestingly, for subgroups, typical growth in math was predictive for only male respondents and White respondents. On the other hand, if a respondent was named MVP on a sports team at age 17, he or she was 93% more likely to be highly socially involved at age 26. Further, if the respondent was a college graduate, he or she was about 37% more likely to be highly socially involved at age 26.

It is expected that respondents who were named MVP would be more likely to be highly socially involved later in life. The behaviors that are required to be named MVP are necessarily highly social in nature. Unexpectedly, when I disaggregated the results based on ethnicity, I found that the two background variables that I controlled for, college graduation and MVP status, were both not significant for Asian respondents and Black respondents for predicting highly social behaviors later in life.

Societal Outcome: Informational Involvement in 2000

In aggregate, informational involvement, as measured by the composite of library visits, books read, accessing Internet for information, reading newspapers, and watching TV news, was not predicted by growth in large-scale assessment performance. Consistent with the findings for social involvement, there was one exception for a subgroup mentioned in the previous chapter. College graduation increased the odds of being highly involved in informational acquisition by 24%, being in an academic school increases the odds by about 88%, and being enrolled in an AP program in high school increases these odds by 21%.

The only independent variable that was predictive for all subgroups was whether the respondent was in an academic school in 2000, which consistently indicated that the respondent was highly involved in obtaining information from the outside world. This finding is intuitive and significantly affects the odds of the respondent's informational involvement. In contrast, the other similar indicator, college graduation status, was not a consistently significant predictor for all subgroups. Indeed, although overall high informational involvement can be predicted by college graduation, female respondents, Asian respondents, Black respondents, and Hispanic respondents saw no effect from the signals of college graduation and being enrolled in an AP program. This finding implies that the signals obtained through school and through enrollment in an AP program do not affect all subgroups identically.

Recommendations for Policymakers

This aim of this research was to find out the predictive value of large-scale assessments using student growth. Currently, a majority of states do not employ multiple measures of student success and have an overreliance on single high-stakes measures. I recommend using multiple measures of student success to add to the snapshot created by criterion-based measures. Tracking the growth of a student over the course of many years through large-scale assessments can help provide a clearer picture of the student's mastery of a subject.

Furthermore, tracking longitudinal data nationally is commonplace for U.S. government agencies including, but not limited to, the Internal Revenue Service and the National Security Agency. I recommend that along with tracking large-scale assessment performance growth, we track and make available to educators timely performance data to help educators make data-driven decisions on behalf of every student.

Recommendations for the Institute of Educational Sciences (IES)

IES is the agency that provided the data for this research by tracking eighth grade students for 12 years. I found two significant limitations while conducting the analyses. First, NELS did not track Student Growth Percentiles (SGP), which are an important measure for student growth on exams. Adding this measure will help future researchers determine the value of improving students' exam performance in relation to their peers. Currently, the performance results of test performance are highly correlated to background variables. This makes disentangling performance from other measures a difficult statistical task for researchers. By including a measure of student improvement, which is one of the primary goals of educators, future studies will include a useful metric currently in place in several states throughout the country.

The second recommendation is to track participants longer in their studies. Although the income levels at age 25 were enough for projections for this study, the participants have barely joined the workforce; and the advantages of college graduation, and perhaps performance on large-scale assessments, is understated as a result. Since college graduates were in school for four years neither earning a salary nor gaining experience, salaries are still skewed at age 25. I recommend creating an ongoing study for future research where the participants are tracked into their mature life and tracking the future generation that would include their children. This will allow researchers in the future to measure the true impact of educational programs on the populations that they are meant to serve.

Recommendations for Future Research

To trigger a shift in the culture of education research towards outcomes that matter to students and society, this study begins a philosophical shift of measuring accountability via long term outcomes: judging advantages and disadvantages by what happens to students when school ends and the real test begins. The existing body of research surrounding student performance on large-scale assessments misses the fundamentals of education. When students go to school each day, the commonplace thoughts of those in charge are to improve student performance and measure student growth. However, student and teacher performance does not always fit neatly into one category or another. Whether we measure improvements or performance, existing accountability systems tend to measure how much students learn without incorporating why students learn. Thus, authentic learning situations can often take a back seat to tomorrow's mandated high-stakes test. As such, I encourage researchers to focus on outcomes beyond a single measure of success and view those single measures as a single marker in a long marathon that lasts a lifetime.

Going beyond the classroom is where one will find what is most important:

real-world outcomes encompassing employment, health, and social engagement. The chief concern surrounding student test performance should also engage policymakers and education professionals in the conversation of what outcomes we are seeking from a good education. As longitudinal student data becomes more readily available to researchers, I recommend that we continue tracking large-scale gains in education and value those gains over short-term and superficial measurements of learning that are used today. For this, researchers will have to bear in mind that a good education is not an end, but rather the beginning of a lifelong love of learning. Success should be viewed, not as a number that a student achieves at age 16, but as lifelong improvements in lifestyle, community engagement, and beyond.

REFERENCES

Achilles, C. M. (1999). Let's put kids first, finally: Getting class size right. Thousand Oaks, CA: Corwin Press.

Adebayo, B. (2008). Cognitive and non-cognitive factors: Affecting the academic performance and retention of conditionally admitted freshmen. *Journal of College Admission*, *200*, 15–21.

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college*. Washington, DC: U.S. Department of Education.

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., & Syme, S. L. (1994). Socioeconomic status and health: The challenge of the gradient. *American Psychologist*, *49*(1), 15.

Alderman, H., Behrman, J. R., Ross, D. R., & Sabot, R. (1996). The returns to endogenous human capital in Pakistan's rural wage labour market. *Oxford Bulletin of Economics and Statistics*, *58*(1), 29–55.

Allen, S. G. (2001). Technology and the wage structure. *Journal of Labor Economics*, *19*, 440–483.

Altonji, J. G. (1995). The effects of high school curriculum on education and labor market outcomes. *Journal of Human Resources*, *30*, 409–438.

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing & student learning. *Education Policy Analysis Archives*, *10*, 18.

Amrein-Beardsley, A., Berliner, D. C., & Rideau, S. (2010). Cheating in the first, second, and third degree: Educators' responses to high-stakes testing. *Education Policy Analysis Archives*, *18*(14), 1-33.

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, *114*(2), 533–575.

Arrow, K. J. (1997). The benefits of education and the formation of preferences. In J. R. Behrman & N. Stacey (Eds.), *The social benefits of education* (pp. 11–16). Ann Arbor, MI: University of Michigan Press.

Ashton, P. (1985). Motivation and the teacher's sense of efficacy. *Research on Motivation in Education*, *2*, 141–174.

Bachman, J. G., Wallace, J. M., O'Malley, P. M., Johnston, L. D., Kurth, C. L., & Neighbors, H. W. (1991). Racial/ethnic differences in smoking, drinking, and illicit drug use among American high school seniors, 1976-89. *American Journal of Public Health*, *81*(3), 372–377.

Backlund, E., Sorlie, P. D., & Johnson, N. J. (1999). A comparison of the relationships of education and income with mortality: the national longitudinal mortality study. *Social Science & Medicine*, *49*(10), 1373–1384.

Baker, B. D., Oluwole, J. O., & Green, P. C., III. (2013). The legal consequences of mandating high-stakes decisions based on low quality information: Teacher evaluation in the Race-to-the-Top era. *Education Policy Analysis Archives*, *21*.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L. (2010). *Problems with the use of student test scores to evaluate teachers* (Vol. 278). Washington, DC: Economic Policy Institute.

Becker, G. S. (1993). *Human capital: A theoretical and empirical analysis, with* *special reference to education* (3rd ed.). Chicago, IL: University of Chicago Press.

Becker, G. S., & Mulligan, C. B. (1997). The endogenous determination of time preference. *The Quarterly Journal of Economics*, *112(*3), 729–758.

Behrman, J. R. (1997). Conceptual and measurement issues. In J. R. Behrman & N. Stacey (Eds.), *The social benefits of education* (pp. 17–68). Ann Arbor, MI: University of Michigan Press.

Behrman, J. R., & Stacey, N. (Eds.). (1997). *The social benefits of education*. Ann Arbor, MI: University of Michigan Press.

Berkman, L. F., Glass, T., Brissette, I., & Seeman, T. E. (2000). From social integration to health: Durkheim in the new millennium. *Social Science & Medicine*, *51*(6), 843–857.

Betebenner, D. W. (2007). *Estimation of student growth percentiles for the Colorado student assessment program*. Dover, NH: National Center for the Improvement of Educational Assessment.

Betebenner, D. W. (2011*). A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth projections/trajectories.* Dover, NH: The National Center for the Improvement of Educational Assessment.

Betebenner, D. W., & Linn, R. L. (2009, December 1). *Growth in student achievement: Issues of measurement, longitudinal data analysis, and accountability*. Paper presented at the Exploratory Seminar: Measurement Challenges within the Race to the Top Agenda, Princeton, NJ.

Betebenner, D. W., Van Iwaarden, A., Domingue, B., & Shang, Y. (2013*). SGP: An R package for the calculation and visualization of student growth percentiles & percentile growth trajectories* (pp. 1–68). R package version 1.0-3.0. Retrieved from http://schoolview.github.com/SGP/

Bishop, J. H. (1989a). *Achievement, test scores and relative wages* (CAHRS Working Paper No. 89-22). Ithaca, NY: Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies.

Bishop, J. H. (1989b). Why the apathy in American high schools? *Educational Researcher*, *18*(1), 6–42.

Blau, F. D., & Kahn, L. M. (2005). Do cognitive test scores explain higher U.S. wage inequality? *Review of Economics and Statistics*, *87*(1), 184–193.

Bowles, S., & Gintis, H. (2000). Does schooling raise earnings by making people smarter. *Meritocracy and Economic Inequality*, 118–36.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy*, *1*(2), 176–216.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*(1), 371–399.

Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & García Coll, C. (2001). The home environments of children in the United States, Part I: Variations by age, ethnicity, and poverty status. *Child Development*, *72(*6), 1844–1867.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2013). School leaders matter.

Education Next, *13*(1), 62–69.

Brewer, D. J., Eide, E. R., & Ehrenberg, R. G. (1999). Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college type on earnings. *Journal of Human Resources*, 104–123.

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children*, 55–71.

Brophy, J. E., & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences*. Oxford, England: Holt, Rinehart & Winston.

Caplan, N. (1992). Indochinese refugee families and academic achievement. *Scientific American*, *266*(2), 36–42.

Caplan, N. S., Choy, M. H., & Whitmore, J. K. (1991). *Children of the boat people: A study of educational success*. Ann Arbor, MI: University of Michigan Press.

Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics*, *3*, 1801–1863.

Carnevale, A. P., & Desrochers, D. M. (2001). *Connecting education standards and employment: Course-taking patterns of young workers*. Washington, DC: The American Diploma Project.

Carson, C. C., Huelskamp, R. M., & Woodall, T. D. (1993). Perspectives on education in America. *Journal of Educational Research*, *86*, 259–265.

Cawley, J., Heckman, J., & Vytlacil, E. (2001). Three observations on wages and measured cognitive ability. *Labour Economics*, *8*(4), 419–442.

Chapman, C., Laird, J., & KewalRamani, A. (2013). *Trends in High School Dropout and Completion Rates in the United States: 1972-2009* (Compendium

Report). NCES 2012-006. Washington, DC: National Center for Education Statistics.

Chatterji, M., Seaman, P. T., & Singell, L. D., Jr. (2003). A test of the signalling hypothesis. *Oxford Economic Papers*, *55*(2), 191–215.

Chen, C., & Wei, Y. (2005). Computational issues for quantile regression. *Sankhyā: The Indian Journal of Statistics*, 399–417.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood. *National Bureau of Economic Research Working Paper Series, 17699.*

Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, *357*(4), 370–379.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, *41*(4), 778–820.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, S95–S120.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington DC: U.S. Government Printing Office.

Courtney, K. E., & Polich, J. (2009). Binge drinking in young adults: Data, definitions, and determinants. *Psychological Bulletin*, *135*(1), 142.

Croninger, R., & Lee, V. (2001). Social capital and dropping out of high school: Benefits to at-risk students of teachers' support and guidance. *The Teachers*

College Record, *103*(4), 548–581.

Crum, R. M., Ensminger, M. E., Ro, M. J., & McCord, J. (1998). The association of educational achievement and school dropout with risk of alcoholism: a twenty-five-year prospective study of inner-city children. *Journal of Studies on Alcohol and Drugs*, *59*(3), 318.

Curtin, T. R., Ingels, S. J., Wu, S., Heuer, R., & Owings, J. (2002). *Base-year to fourth follow-up data file user's manual*. Washington, DC: National Center for Education Statistics.

Cutler, D. M., & Lleras-Muney, A. (2006). Education and health. *National Bureau of Economic Research Working Paper Series, 12352, 37.*

Cutler, D. M., & Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics*, *29(*1), 1–28. doi:10.1016/j.jhealeco.2009.10.003

Dale, S. B., & Krueger, A. B. (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *The Quarterly Journal of Economics*, *117*(4), 1491–1527.

Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state policy evidence* (No. R-99-1). Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.

Dee, T. S. (2004). Are there civic returns to education? *Journal of Public Economics*, *88*(9), 1697–1720.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, *30*(3), 418–446.

doi:10.1002/pam.20586

Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, *35*(2), 252–279.

Del Boca, F. K., & Darkes, J. (2003). The validity of self‑reports of alcohol consumption: State of the science and challenges for research. *Addiction*, *98*, 1–12. doi:10.1046/j.1359-6357.2003.00586.x

Dewey, J. (1903a). Democracy in education. *The Elementary School Teacher*, *4*(4), 193–204.

Dewey, J. (1903b). *Ethical principles underlying education*. Chicago, IL: University of Chicago Press.

Dewey, J. (1938). Experience and Education. New York, NY: Macmillan.

Duncan, G. J., Yeung, W. J., Brooks-Gunn, J., & Smith, J. R. (1998). How much does childhood poverty affect the life chances of children? *American Sociological Review*, 406–423.

Dweck, C. S. (2008). Mindsets and math/science achievement. Paper presented at the Carnegie Corporation of New York-Institute for Advanced Study Commission on Mathematics and Science Education. New York, NY: Carnegie Corporation of New York, Institute for Advanced Study, Commission on Mathematics and Science Education.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040.

Dweck, C. S. (2007). Boosting achievement with messages that motivate. *Education*

Canada, *47*(2), 6–10.

Ehrenberg, R. G., & Brewer, D. J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review*, *14*(1), 1–21.

Eisenkopf, G. (2010). Peer effects, motivation, and learning. *Economics of* Education *Review*, *29*(3), 364–374.

Eisenstadt, J. M. (1978). Parental loss and genius. *American Psychologist*, *33*(3), 211.

Elo, I. T., & Preston, S. H. (1996). Educational differentials in mortality: United States, 1979–1985. *Social Science & Medicine*, *42*(1), 47–57.

Embree, B. G., & Whitehead, P. C. (1993). Validity and reliability of self-reported drinking behavior: dealing with the problem of response bias. *Journal of Studies on Alcohol and Drugs*, *54*(3), 334.

Farebrother, R. W. (1974). Algorithm AS 79: Gram-Schmidt regression. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 23*(3), 470–476.

Feddersen, T. J., & Pesendorfer, W. (1996). The swing voter's curse. *The American Economic Review*, 408–424.

Felder, S. (2006). The gender longevity gap: Explaining the difference between singles and couples. *Journal of Population Economics*, *19*(3), 543–557.

Feldhusen, J. F., & Moon, S. M. (1992). Grouping gifted students: Issues and concerns. *Gifted Child Quarterly*, *36*(2), 63–67.

Fine, M. (1991). *Framing dropouts: Notes on the politics of an urban public high school*. New York, NY: SUNY Press.

Finn, J. D., & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, *27*(3), 557–577.

Finnie, R., & Meng, R. (2002). Minorities, cognitive skills and incomes of Canadians. *Canadian Public Policy/Analyse De Politiques*, 257–273.

Freeman, R. B. (1997). Working for nothing: The supply of volunteer labor. *Journal of Labor Economics*, *15*(Part 2: Essays in Honor of Yoram Ben- Porath), S140–S166.

Fullan, M. (2008). *The six secrets of change: What the best leaders do to help their organizations survive and thrive*. San Fransico, CA: Jossey-Bass.

Gamoran, A. (1994). *Impact of academic coursework on labor market outcomes for youth who do not attend college: A research review*. Washington, DC: U.S. Department of Education.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*(6), 1539–1552. doi:10.1037/a0025510

Gee, J. P. (1999). Critical issues: Reading and the new literacy studies: Reframing the national academy of sciences report on reading. *Journal of Literacy Research*, *31*(3), 355–374.

Gibson, J. (2001). Unobservable family effects and the apparent external benefits of education. *Economics of Education Review*, *20*(3), 225–233.

Gladwell, M. (2013). *David and Goliath: Underdogs, misfits, and the art of battling giants*. New York, NY: Little, Brown and Company.

Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative routes to teaching: The impacts of Teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management*, *25*(1), 75–96.

Goldhaber, D. D., & Brewer, D. J. (1996). Evaluating the effect of teacher degree level on educational performance. *Education Economics*, *7*.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, *22*(2), 129–145.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics*, *89*(1), 134–150.

Goodman, P. (1964). *Compulsory miseducation*. New York, NY: Horizon Press.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, *80*(2), 223–255.

Grossman, M., & Kaestner, R. (1997). Effects of education on health. In J. R. Behrman & N. Stacey (Eds.), *The social benefits of education* (pp. 69–124). Ann Arbor, MI: University of Michigan Press.

Gruenewald, P. J., & Johnson, F. W. (2006). The stability and reliability of self-reported drinking measures. *Journal of Studies on Alcohol and Drugs*, *67*(5), 738.

Guerra, L. G., & Verghese, A. (1992). The costs of poor health habits. *JAMA: The Journal of the American Medical Association*, *267*(18), 2534–2534.

Hansen, W. L. (1963). Total and private rates of return to investment in schooling.

Journal of Political Economy, *71*(2), 128–140.

Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review*, 280–288.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, *24*(3), 1141–1177.

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher, 18*(4), 45–62.

Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy*, 84–117.

Hanushek, E. A. (1994). *Making schools work: Improving performance and controlling costs*. Washington, DC: Brookings Institution Press.

Hanushek, E. A. (1995). Education production functions. *International Encyclopedia of Economics of Education*, 277–282.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, *19*(2), 141–164.

Hanushek, E. A. (1998). Conclusions and controversies about the effectiveness of school resources. *Economic Policy Review*, *4*(1).

Hanushek, E. A. (1999). The evidence on class size. In S. Mayer and P. E. Peterson (Eds.), *Earning and Learning: How Schools Matter* (pp.131–168). Washington, DC: Brookings Institution.

Hanushek, E. A. (2003). The failure of input‑based schooling policies. *The Economic Journal*, *113*(485), F64–F98.

Hanushek, E. A. (2005). The economics of school quality. *German Economic Review*, *6*(3), 269–286.

Hanushek, E. A. (2006). *Courting failure: How school finance lawsuits exploit judges' good intentions and harm our children*. Stanford, CA: Hoover Press.

Hanushek, E. A. (2009a). Teacher deselection. In D. Goldhaber and J. Hathaway (Eds.), *Creating a New Teaching Profession* (pp. 165–80). Washington, DC: Urban Institute Press.

Hanushek, E. A. (2009b). The economic value of education and cognitive skills. In G. Sykes, B. Schneider, and D. N. Plank (Eds.), *Handbook of Education Policy Research* (pp. 39–56). New York, NY: Routledge.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality*. (Working paper). Cambridge, MA: National Bureau of Economic Research.

Hanushek, E. A., & Raymond, M. E. (2006). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, *24*(2), 297–327.

Hanushek, E. A., Rivkin, S. G., Figlio, D., & Jacob, B. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, *100*(2), 267–271.

Hanushek, E. A., Rivkin, S. G., & Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *Review of Economics and Statistics*, *78*(4), 611–627.

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, *27*(1), 4–9.

Haveman, R., & Wolfe, B. (1995). The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature*, *33*(4), 1829–1878.

Hayghe, H. V. (1991). Volunteers in the U.S.: Who donates the time? *Monthly Labor Review*, 114, 17.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, *24*, 411–482.

Helliwell, J. F., & Putnam, R. D. (2007). Education and social capital. *Eastern Economic Journal*, 33.

Hill, C. R., & Stafford, F. P. (1974). Allocation of time to preschool children and educational opportunity. *Journal of Human Resources*, 323–341.

Hinchey, P. H. (2010). *Getting teacher assessment right: What policymakers can learn from research*. Boulder, CO: National Education Policy Center.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, *7*, 24-52.

Horton, H. D., Hixson, L., & Monnat, S. M. (2000). Educational attainment and mortality differentials. In P. Peterson, E. L. Baker, & B. McGraw (Eds.), *International Encyclopedia of Education* (Vol. 1, pp. 602–609). Oxford, England: Elsevier.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. National Bureau of Economic Research Working Paper Series, 7867.

Huang, J., Maassen van den Brink, H., & Groot, W. (2009). A meta-analysis of the effect of education on social capital. *Economics of Education Review*, *28*(4), 454–464.

Hungerford, T., & Solon, G. (1987). Sheepskin effects in the returns to education. *The Review of Economics and Statistics*, 175–177.

Ingels, S. J. (1994). *National education longitudinal study of 1988. Second follow-up: Student component data file user's manual*. Washington, DC: National Center for Education Statistics.

Iremonger, L. (1970). *The fiery chariot: A study of British prime ministers and the search for love*. London, England: Martin Secker & Warburg, Ltd.

Jackson, A. W., & Davis, G. A. (2000). *Turning points 2000*. New York, NY: Teachers College Press

Jacob, B. A. (2002). *The impact of high-stakes testing on student achievement: Evidence from Chicago*. Unpublished Manuscript, Harvard University, Cambridge, MA.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, *89*(5-6), 761–796. doi:10.1016/j.jpubeco.2004.08.004

Jaffe, D. H., Eisenbach, Z., Neumark, Y. D., & Manor, O. (2006). Effects of husbands' and wives' education on each other's mortality. *Social Science & Medicine*, *62*(8), 2014–2023. doi:10.1016/j.socscimed.2005.08.030

Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. *Advances in Experimental Social Psychology*, *28*, 281–388.

Kamins, M. L., & Dweck, C. S. (1999). Person versus process praise and criticism: Implications for contingent self-worth and coping. *Developmental Psychology*, *35*(3), 835.

Kane, T. J., & Rouse, C. E. (1995). Labor-market returns to two-and four-year college. *The American Economic Review*, *85*(3), 600–614.

Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *The Journal of Economic Perspectives*, *16*(4), 91-114.

Kane, T. J., & Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. National Bureau of Economic Research Working Paper Series, 14607.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615–631.

Kaufman, P. (1992). *Characteristics of at-risk students in NELS: 88. National Education Longitudinal Study of 1988: Contractor Report*. (No. 92-042). Washington DC: National Center for Education Statistics.

Kenkel, D. S. (1991). Health behavior, health knowledge, and schooling. *Journal of Political Economy*, 287–305.

Kenkel, D. S. (1993). Drinking, driving, and deterrence: The effectiveness and social costs of alternative policies. *Journal of Law and Economics*, *36*(2), 877–913.

Kim, J. (2008). Intercohort trends in the relationship between education and health: Examining physical impairment and depressive symptomatology. *Journal of Aging and Health*, *20*(6), 671–693.

King, B. A., Dube, S. R., & Tynan, M. A. (2012). Current tobacco use among adults in the United States: Findings from the national adult tobacco survey. *American Journal of Public Health*, *102*(11), e93–e100. doi:10.2105/AJPH.2012.301002

Kitagawa, E. M., & Hauser, P. M. (1968). Education differentials in mortality by cause of death: United States, 1960. *Demography*, *5*(1), 318–353.

Kitagawa, E. M., & Hauser, P. M. (1973). *Differential mortality in the United States: A study in socioeconomic epidemiology*. Cambridge, MA: Harvard University Press.

Koenker, R., & Ng, P. (2012). *SparseM: Sparse Linear Algebra. R package version 0.96*. Retrieved from http://CRAN.R-project.org/package=SparseM

Korenman, S., Miller, J. E., & Sjaastad, J. E. (1995). Long-term poverty and child development in the United States: Results from the NLSY. *Children and Youth Services Review*, *17*(1), 127–155.

Koretz, D. M. (2002). Limitations in the use of achievement tests as measures of educators' productivity. *Journal of Human Resources*, 752–777.

Koretz, D. M. (2008). A measured approach. *American Educator*, *32*(2), 18–39.

Koretz, D. M. (2009). *Measuring up*. Cambridge, MA: Harvard University Press.

Koretz, D. M., Barron, S., Mitchell, K. J., & Stecher, B. M. (1996). *Perceived effects* *of the Kentucky instructional results information system (KIRIS).* Santa Monica, CA: RAND.

Koretz, D. M., McCaffrey, D. F., & Hamilton, L. S. (2001). *Toward a framework for validating gains under high-stakes conditions* (No. 551). Los Angeles, CA: The Regents of the University of California.

Koretz, D. M., Linn, R. L., Dunbar, S. B., & Shepard, L. A. (1991). *The effect of high-stakes testing on achievement: Preliminary findings about generalization across tests*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, Educational Resources Information Center.

Kroenker, R. (2000). Galton, Edgeworth, Frisch, and prospects for quantile regression in econometrics. *Journal of Econometrics*, *95*, 347–374.

Kronholz, J. (2012). Academic value of non-academics: The case for keeping extracurriculars. *Education Next*, *12*(1), 8-14.

Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, *114*(2), 497–532.

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, *113*(485), F34–F63. doi:10.1111/1468-0297.00098

Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college‐test taking and middle school test results: Evidence from Project STAR. *The Economic Journal*, *111*(468), 1–28.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy*

Analysis, *24*(1), 37–62.

Lantz, P. M., House, J. S., Lepkowski, J. M., Williams, D. R., Mero, R. P., & Chen, J. (1998). Socioeconomic factors, health behaviors, and mortality. *JAMA: The Journal of the American Medical Association*, *279*(21), 1703–1708.

Lazear, E. P. (2001). Educational production. *The Quarterly Journal of Economics*, *116*(3), 777–803.

Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review*, 10(2), 179–214.

Lee, Y. (2000). The predictive value of self-assessed general, physical, and mental health on functional decline and mortality in older adults. *Journal of Epidemiology and Community Health*, *54*(2), 123–129.

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2005). *SPSS for intermediate statistics: Use and interpretation* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Levine, M., & Levine, A. (2013). Holding accountability accountable: A cost-benefit analysis of achievement test scores. *American Journal of Orthopsychiatry*, *83*(1), 17–26. doi:10.1111/ajop.12014

Lintonen, T., Ahlström, S., & Metso, L. (2004). The reliability of self-reported drinking in adolescence. *Alcohol and Alcoholism*, *39*(4), 362–368.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *The Review of Economic Studies*, *72(*1), 189–221.

Lleras-Muney, A., & Lichtenberg, F. R. (2002). The effect of education on medical technology adoption: Are the more educated more likely to use new drugs.

National Bureau of Economic Research Working Paper Series, 9185, 2–47.

Lucas, R. E. (1977). Hedonic wage equations and psychic wages in the returns to schooling. *The American Economic Review*, 549–558.

Lynch, S. M. (2003). Cohort and life-course patterns in the relationship between education and health: A hierarchical approach. *Demography*, *40*(2), 309–331.

Mann, H. (1855). *Lectures on education*. Boston, MA: Ide & Dutton.

Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, *12*(3), 591–612.

Mason, C. H., & Perreault, W. D., Jr. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research*, 268–280.

Massy, W. F. (1965). Principal components regression in exploratory statistical research. *Journal of the American Statistical Association*, *60*(309), 234–256.

Maynard, R. A., & McGrath, D. (1997). Family structure, fertility, and child welfare. In J. R. Behrman & N. Stacey (Eds.), *The social benefits of education* (pp. 125–174). Ann Arbor, MI: University of Michigan Press.

McIntosh, S. (2012). *State high school exit exams*. Washington, DC: Center for Educational Policy.

McIntosh, S., & Vignoles, A. (2001). Measuring and assessing the impact of basic skills on labour market outcomes. *Oxford Economic Papers*, *53*(3), 453–481.

Mechanic, D. (2002). Disadvantage, inequality, and social policy. *Health Affairs*, *21*(2), 48–59.

Milligan, K., Moretti, E., & Oreopoulos, P. (2004). Does education improve citizenship? Evidence from the United States and the United Kingdom.

Journal of Public Economics, *88*(9), 1667–1695.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, *13*(2), 125–145.

Mulligan, C. B. (1999). Galton versus the human capital approach to inheritance. *Journal of Political Economy*, *107*(S6), S184–S224.

Murnane, R. J. (1985). Do effective teachers have common characteristics: Interpreting the quantitative research evidence. Paper presented at the National Research Council Conference on Teacher Quality in Science and Mathematics, Washington, DC.

Murnane, R. J., & Phillips, B. R. (1981). Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teaching performance. *Economics of Education Review*, *1*(4), 453–465.

Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *The Review of Economics and Statistics*, *77*, 251–266.

Murnane, R. J., Willett, J. B., Duhaldeborde, Y., & Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management*, *19*(4), 547–568.

Musick, M. A., & Wilson, J. (2007). *Volunteers: A social profile*. Bloomington, IN: Indiana University Press.

National Commission on Excellence in Education. (1983). *A nation at risk*. Washington, DC: U.S. Government Printing Office.

Neal, D., & Schanzenbach, D. W. (2007). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics*, *92*(2), 263–283.

Newhouse, J. P., Rand Corporation. Insurance Experiment Group. (1993). *Free for all?: Lessons from the RAND health insurance experiment*. Cambridge, MA: Harvard University Press.

Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness. *Educational Policy Analysis Archives*, *18*.

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.

Nie, N., & Hillygus, D. S. (2001). Education and democratic citizenship. In D. Ravitch & J. P. Viteritti (Eds.), *Making good citizens: Education and civil society*. Binghamton, NY: Yale University Press.

Olson, L. (2001). Study questions reliability of single year test-score gains. *Education Week*, *20*(37).

Owens, J. (2004). A review of the social and non-market returns to education. *Education and Learning Wales*, 2–22.

Parsons, J., Kaczala, C. M., & Meece, J. L. (1982). Socialization of achievement attitudes and beliefs: Classroom influences. *Child Development*, 322–339.

Patrick, D. L., Cheadle, A., Thompson, D. C., Diehr, P., Koepsell, T., & Kinne, S. (1994). The validity of self-reported smoking: A review and meta-analysis. *American Journal of Public Health*, *84*(7), 1086–1093.

Phelps, R. P. (2005). *Kill the messenger: The war on standardized testing*. New Brunswick, NJ: Transaction Publishers.

Portes, A. (1998). Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology*, *24*, 1–24.

Preston, S. H., & Elo, I. T. (2003). Are educational differentials in adult mortality increasing in the United States? *Journal of Aging and Health*, *7*(4), 476–496. Retrieved from http://jah.sagepub.com/content/7/4/476.short

Putnam, R. D. (1995). Bowling alone: America's declining social capital. *Journal of Democracy*, *6*(1), 65–78.

Ravitch, D. (2011). *The death and life of the great American school system*. New York, NY: Basic Books.

Ravitch, D. (2013). *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. New York, NY: Knopf.

Rist, R. C. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review*, *40*(3), 411–451.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458. doi:10.1111/j.1468-0262.2005.00584.x

Rock, D. A., Pollack, J., & Quinn, P. (1995). *Psychometric report for the NELS:88 base year through second follow-up*. Washington, DC: National Center for Education Statistics.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, *94*(2), 247–252.

Roeser, R. W., Eccles, J. S., & Sameroff, A. J. (1998). Academic and emotional functioning in early adolescence: Longitudinal relations, patterns, and prediction by experience in middle school. *Development and Psychopathology*, *10*(02), 321–352.

Rose, H., & Betts, J. R. (2004). The effect of high school courses on earnings. *Review of Economics and Statistics*, *86*(2), 497–513.

Rosenzweig, M. R., & Schultz, T. P. (1982). Market opportunities, genetic endowments, and intrafamily resource distribution: Child survival in rural India. *The American Economic Review*, *72*(4), 803–815.

Ross, C. E., & Wu, C. L. (2007). The links between education and health. *American Sociological Review*, *60*, 719–745.

Rumberger, R. (2011). Solving the nation's dropout crisis. *Education Week*, *28*.

Ryan, R. M., Fauth, R. C., & Brooks-Gunn, J. (2006). Childhood poverty: Implications for school readiness and early childhood education. In O. N. Saracho (Ed.), *Handbook of research on the education of young children* (2nd ed., pp. 323–346). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Sacerdote, B. (2004). What happens when we randomly assign children to families? National Bureau of Economic Research Working Paper Series, 10894.

Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11*(1), 57–67.

Schafer, W. D., Lissitz, R. W., Zhu, X., Zhang, Y., Hou, X., & Li, Y. (2012). Evaluating teachers and schools using student growth models. *Practical*

Assessment, Research & Evaluation, *17*.

Schochet, P. Z., & Chiang, H. (2010). Error rates in measuring teacher and student performance based on student test score gains (NCEE 2010-4004). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, *51*(1), 1–17.

Schwartz, F. (1981). Supporting or subverting learning: Peer group patterns in four tracked schools. *Anthropology & Education Quarterly*, *12*(2), 99–121.

Scott, L. A., Ingels, S. J., & Ownings, J. A. (2007). *Interpreting 12th-graders' NAEP-scaled mathematics performance using high school predictors and postsecondary outcomes from the National Education Longitudinal Study of 1988 (NELS:88),* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Scott, L. A., Rock, D. A., Pollack, J., & Ingels, S. J. (1995). *Two years later: Cognitive gains and school transitions of NELS: 88 eighth graders*. (No. NCES 95-436) Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Sewell, W. H., & Hauser, R. M. (1975*). Education, occupation, and earnings: Achievement in the early career*. New York, NY: Academic Press.

Shields, C. M. (1996). To group or not to group academically talented or gifted students? *Educational Administration Quarterly*, *32*(2), 295–323.

Simonton, D. K. (2012). Teaching creativity: Current findings, trends, and controversies in the psychology of creativity. *Teaching of Psychology*, *39*(3), 217–222. doi:10.1177/0098628312450444

Simonton, D. K., Colangelo, N., Assouline, S. G., & Ambroson, D. L. (1993). Genius and giftedness: Parallels and discrepancies. Paper presented at the Talent Development: National Research Symposium, Dayton, OH.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society*. Series B (Methodological), 238–241.

Singh, G. K., & Siahpush, M. (2002). Ethnic-immigrant differentials in health behaviors, morbidity, and cause-specific mortality in the United States: An analysis of two national data bases. *Human Biology*, *74*, 83–109.

Smeeding, T. (1983). The size distribution of wage and nonwage compensation: employer cost versus employee value. In J. E. Triplett (Ed.), *The measurement of labor cost* (pp. 237–286). Chicago, IL: University of Chicago Press.

Smith, V. K. (1997). Feedback effects and environmental resources. In J. R. Behrman & N. Stacey (Eds.), *The social benefits of education*. Ann Arbor, MI: University of Michigan Press.

Son, J., & Wilson, J. (2012). Using normative theory to explain the effect of religion and education on volunteering. *Sociological Perspectives*, *55*(3), 473–499.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355–374.

Stecher, B. M., & Barron, S. I. (1999). *Quadrennial milepost accountability testing in Kentucky*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation (CSE), Graduate School of Education & Information Studies, University of California.

Stecher, B. M., & Mitchell, K. J. (1995). *Portfolio-driven reform: Vermont teachers' understanding of mathematical problem solving and related changes in classroom practice.* Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Graduate School of Education & Information Studies, University of California.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797.

Tanner, D. (1998). The social consequences of bad research. *Phi Delta Kappan*, *79*(5), 345–349.

Tanner, D., & Tanner, L. N. (2006). *Curriculum development: Theory into practice* (4th ed.). Saddle River, NJ: Pearson Merrill/Prentice Hall.

Thorndike, E. L. (1927). The refractory period in associative processes. *Psychological Review*, *34*(3), 234.

Thorndike, E. L. (1932). *The fundamentals of learning*. New York, NY: Teachers College Bureau of Publications.

Tienken, C. H. (2011). Structured inequity: The intersection of socioeconomic status and the standard error of measurement of state mandated high school test results. *Blazing New Trails: Preparing Leaders to Improve Access*. Lancaster: DEStech Publications, 257-285.

Tienken, C. H., & Orlich, D. C. (2013). *The school reform landscape: Fraud, myth, and lies*. Lanham, MD: Rowman & Littlefield.

Van Ments, M. (1999). *The effective use of role play* (2nd ed.). London, England: Kogan Page Limited.

Vardardottir, A. (2013). Peer effects and academic achievement: A regression discontinuity approach. *Economics of Education Review*, *36*, 108–121. doi:10.1016/j.econedurev.2013.06.011

Viscusi, W. K. (1992). *Smoking: Making the risky decision*. New York, NY: Oxford University Press.

Wechsler, H., Dowdall, G. W., Davenport, A., & Castillo, S. (1995). Correlates of college student binge drinking. *American Journal of Public Health*, *85*(7), 921–926.

Weinstein, R. S. (1989). Perceptions of classroom processes and student motivation: Children's views of self-fulfilling prophecies. *Research on Motivation in Education*, *3*, 187–221.

Weiss, A. (1995). Human capital vs. signaling explanations of wages. *The Journal of Economic Perspectives*, *9*(4), 133–154.

Welch, F. (1970). Education in production. *Journal of Political Economy*, *78*(1), 35–59.

Wilczenski, F. L., Bontrager, T., Ventrone, P., & Correia, M. (2001). Observing collaborative problem-solving processes and outcomes. *Psychology in the Schools*, *38*(3), 269–281.

Wolfe, B. L., & Haveman, R. H. (2002). Social and nonmarket benefits from education in an advanced economy. *Federal Reserve Bank of Boston, 47*, 97-142.

Wolfe, B. L., & Zuvekas, S. (1995). Nonmarket outcomes of schooling. *International Journal of Educational Research*, *27*(6), 491-501.

Wolfinger, R. E., & Rosenstone, S. J. (1980). *Who votes?* New Haven, CT: Yale University Press.

Zellmer, M. B., Frontier, A., & Pheifer, D. (2006). What are NCLB's instructional costs? *Educational Leadership*, *64*(3), 43.

Appendix A – Technical Specifications

Technique for Calculating Student Growth Percentiles (SGP)

The NELS:88/00 dataset consisted of 12,144 students in the total dataset. After using the F4PNLWT weight described above, 1,317 students were eliminated from the sample due to having a weight of 0. A total of 10,827 students were counted in the dataset. To compute SGPs for each student, I used the *R* software package developed by Betebenner et al. (2013). This package was developed specifically for computing SGP over multiple years of data. There are two technical limitations to using SGP on the NELS:88/00 dataset: (1) I am not aware of any technique that allows weights to be entered directly when calculating SGPs, and (2) the size of the dataset with applied weights creates computational complexity issues that cannot be resolved with regular computers.

Using SPSS, I exported a Microsoft Excel-formatted subset of the NELS:88/00 data. See the table below for the list of variables and their descriptions. There were 12 test score variables that represented three years of testing for subjects. Student ID was used as the primary key to merge the computational results back to the SPSS dataset. F4PNLWT, as described previously, was the weight applied to maintain the representative nature of the underlying student population after subsampling.

Table 44

Variables Exported To Run SGP Computations

Variable	Description
STU_ID	Student ID
F4PNLWT	Panel weight for students who completed all 5 follow-ups
BY2XRTH	1988 reading score, 8th Grade
F12XRTH	1990 reading score
F22XRTH	1992 reading score
BY2XMTH	1988 mathematics score, 8th Grade
F12XMTH	1990 mathematics score
F22XMTH	1992 mathematics score
BY2XSTH	1988 science score, 8th Grade
F12XSTH	1990 science score
F22XSTH	1992 science score
BY2XHTH	1988 history score, 8th Grade
F12XHTH	1990 history score
F22XHTH	1992 history score

Computational Complexity Issues

The weight of the dataset was normalized using a technique described below due to issues of computational complexity. The 10,827 participants represented 2,921,547 individuals who were in eighth grade in 1988. The SGP technique used by Betebenner et al. (2013) directly uses quantile regression (Koenker & Ng, 2012) in the software package *R*. Koenker (2000) estimates the computational complexity for n observations with p parameters to be $O(np^3\log^2 n)$. For n = 2,921,547, for the number of participants and p = 3, for number of exams, the computational time would be unreasonable on a standard computer. In line with similar estimations (Chen & Wei, 2005; Kroenker, 2000), I found the computational time for three exams to be $O(n^2)$ upon completing a battery of sample calculations. When I entered the sample data and entered a polynomial line of best fit, the projected time for completing the computation on the computer used for this research would take approximately 43 days for each subject. See Figure 5 below for the estimation of computation time.

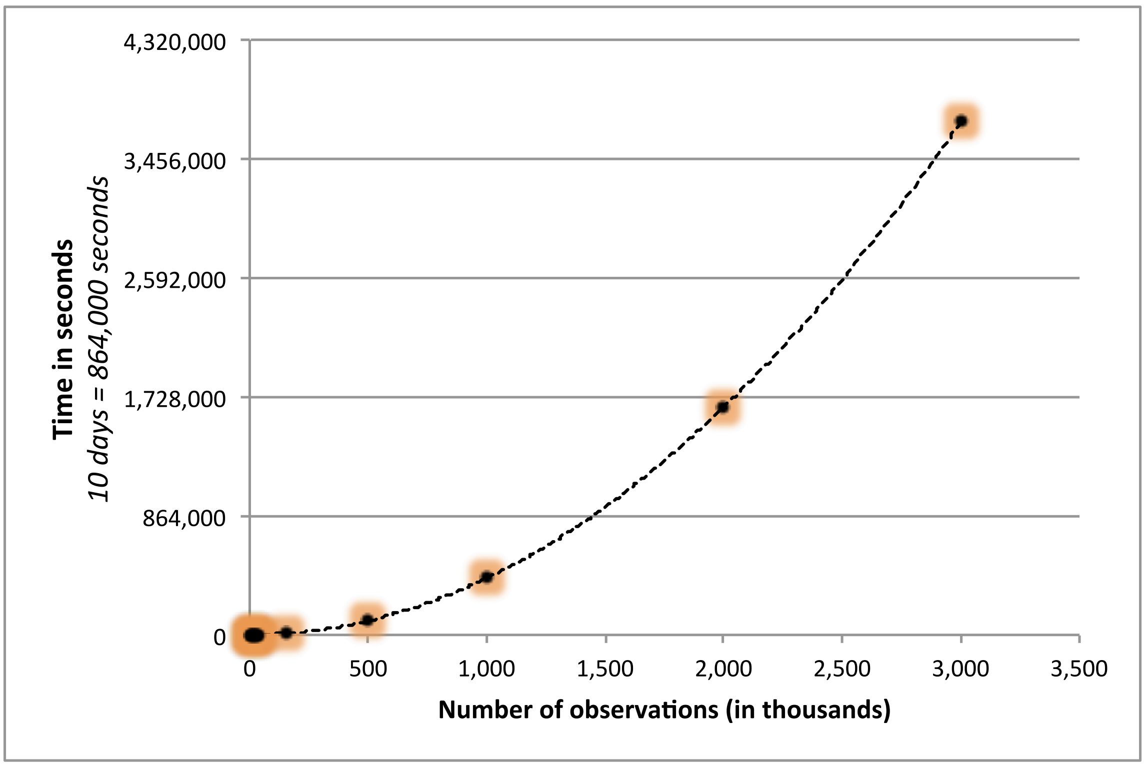


Figure 4. Estimated time of computation for calculating SGP on the weighted NELS dataset.

Computing Student Growth Percentiles Using NELS

In light of the computational time associated with the fully weighted dataset, I selected to normalize the weights to minimize the number of observations. To do this, I used the aforementioned exported dataset in Microsoft Excel and transformed the weight to the nearest 20. This is a form of rounding using the following formula for each participant: =ROUND([@F4PNLWT]/20,0)*20. This rounding technique projected the population to 2,916,560, which was 0.2% less than the original estimated population using F4PNLWT. Furthermore, since each weight was now divisible by 20, this technique reduced the number of observations needed to 145,828. This minimized the computation time from 43 days to 2 hours per subject.

Using a counting technique described here in Microsoft Excel, I generated the appropriate number of students for the rounded weight. In sum, this technique duplicated each participant the adjusted weight vector amount of times; that is, imagine a participant represents 100 people. I added the equivalent of 100 duplicates of this participant. Keep in mind that this technique mimics the weights that were assigned for the NELS:88/00 to within 0.2% accuracy, as descried previously.

I added a variable counter that acted as an incremental primary key. This variable increased by one for each sequential participant. Specifically, the first participant had a value of 1, the second participant had a value of 2, and so on in this manner. I created a separate table and used variables and formulas described in the table below. The table contains the appropriate number of duplicates to represent the weight F4PNLWT.

Table 45

Variables Used To Generate Appropriate NELS:88/00 Weights

Variable	Pseudo Formula	Description
Count	=IF(Last_Count<=Last_Counter,Last_Count+1,Last_Count)	Uniquely identifies the participant to look up all 3 test scores per subject.
STU_ID	=VLOOKUP(Count,NELS,2,FALSE)&"X"&Counter	Student ID is used to map student back to original dataset in SPSS. Concatenation for X is used as a delimiter, and count allows unique identifiers for each student as required by the SGP R software package.
SGPAdjustedWeight	=VLOOKUP(Count,NELS,17)	Adjusted weight for given participant
Counter	=IF(Count=Last_Count,Counter+1,1)	Increments this counter until it reaches weight.
Vector of Exams	=IF(Count=0,"NA",VLOOKUP(Count,NELS,3,FALSE))	Looks up test scores based on count.

Using the cells associated with formulas in Table 45, I created 145,828 entries that represented a national population of 2,916,560 eighth graders in 1988. I saved the document as a tab-separated text file and ran the script described in Table 46 in the *R* software package.

Table 46

Script To Apply Student Growth Percentiles to the NELS:88/00 Dataset

Algorithmic Description	Pseudo Code
Step 1: Open SGP package	`> require(SGP)`
Step 2: Import adjusted NELS:88/00 data in tab separated format into *nels* array	`> nels <- read.csv("nels-reading.txt", sep = "")`
Step 3: Apply SGP to nels array and save into array nelsWithSGP	`> nelsWithSGP <- studentGrowthPercentiles (panel.data=nels, grade.progression=8:12)`
Step 5: Create comma separated format with data from nelsWithSGP	`> sgpCSV <- rbind(nelsWithSGP $SGPercentiles$READING.1992)`
Step 6: Save comma-separated dataset for importing to SPSS NELS:88/00	`> write.csv(sgpCSV, file="sgp_reading.csv")`

Coding of Independent Variables

Social and characteristic factors included measures that described the student's personal and family characteristics in eighth grade in 1988. With the exception of the SEX and F4RACEM variables, all variables in table 4 were collected during the base year. The variable F4RACEM was a derivation of previous responses to align with federal standards for categorizing ethnicity and allowed the identification of individuals who were of mixed heritage to self-identify appropriately (Curtin et al., 2002, p. 235). The variable SEX was derived mainly from the second follow up.

Table 47

Summary of Variables That Measure Characteristic and Social Factors

Variable	Label	Coded Values
MonthsOld_Grade8	Age in months in 8th grade	No coding
SEX	Composite Sex	1 = Male 2 = Female
F4RACEM	Ethnicity by category	1 = American Indian or Alaska Native
		2 = Asian or Pacific Islander
		3 = Black, not Hispanic
		4 = White, not Hispanic
		5 = Hispanic or Latino
		6 = More than one race
BYRISK	Risk of dropping out of school	0 = No risk factors
		...
		5 = Five risk factors
BYSES	Socio-economic status composite	No coding
BYFAMINC	Yearly family income 1987	0 = None
		1 = Less than $1,000
		...
		14 = $200,000 or more

Breakdown and Computation of Variables

Computation of age.

In order to create an age in months, I converted the month and year of birth to months, using the script below in SPSS on the NELS:88 dataset:

```
RECODE BIRTHMO (1='01') (2='02') (3='03') (4='04') (5='05') (6='06') (7='07') (8='08') (9='09')
(10='10') (11='11') (12='12') INTO String_Month.
RECODE BIRTHYR (72='72') (73='73') (74='74') (75='75') INTO String_Year.
STRING  DOB (A12).
IF  (BIRTHMO <= 12)
   DOB=CONCAT(RTRIM(String_Year),RTRIM(String_Month)).
```

The resulting computation for age entering eighth grade was as follows:

```
IF (BIRTHMO <= 12)
          MonthsOld_Grade8 = (88 - String_Year)*12 + 9 - String_Month.
VARIABLE LABELS  MonthsOld_Grade8 'Age in months in 8th grade'.
```

Risk of dropping out of school.

This composite variable was created using variables from the student's family characteristics in 1988, the base year of the longitudinal study, to assess the respondent's risk of dropping out of school. These factors were specifically identified as associated with at-risk students who tend to have lower performance on standardized exams (Kaufman, 1992). This variable was derived using flags for each variable in the table below and then using sums to create cumulative factors for dropping out of school.

Table 48

Breakdown of Composite Variable Measuring Risk of Dropping Out of School

Variable Name	Description	Flag Condition
BYFCOMP	Family Composition Composite	Single parent family
BYPARED	Parent Education Level Composite	Neither parent Graduated HS
BYP6	Number Siblings Dropped Out	Sibling dropped out
BYS41	Time Spent With No Adults	3 or more hours no adult
BYLEP	Limited English Proficiency Composite	Limited English Proficiency
BYFAMINC	Family Income in 1987	Income less than $15,000

School-level factors.

Various measures of the participants' school characteristics were collected over the first three data collection waves. I used the latest available measures of the region, type, size, and other characteristics described in the table below of schools the student attended from the end of eighth grade to the end of high school.

Table 49

Summary of Variables That Measure School-level Factors

Variable	Label	Coded Values
G12REGON	of student's school in 12th grade	1 = Northeast
		2 = Midwest
		3 = South
		4 = West
G12CTRL1	School classification reported by school in 12th grade	1 = Public
		2 = Catholic
		3 = Private Non-Catholic Religious
		4 = Private Non-Religious
		5 = Other Private
G10URBAN	Type of school district, diocese, county in 10th grade	1 = Urban
		2 = Suburban
		3 = Rural
G10ENROL	School enrollment in 10th grade	1 = 1 - 99
		...
		7 = 700 +
G8LUNCH	Percent free lunch in school in 8th grade	0 = None
		1 = 1% - 5%
		...
		7 = 76% - 100%
G8TYPE	Grade span of school in 8th grade	1 = P, K, Or 1-8
		2 = P, K, Or 1-12
		3 = 6, 7 Or 8-12
		4 = 3, 4, Or 5-8
		5 = 6 - 8

Student performance.

Student performance was measured in four ways in this study: (1) the curricular choices the student made through rigor, enrichment, and workload, (2) the signals of school performance as measured by grades and completion of school degrees, (3) the criterion- and norm-referenced student performance as measured by large-scale assessments, and (4) the growth of performance on large-scale assessments as measured by SGP. See the table below for a complete list of variables in the four categories.

Table 50

Summary of Variables That Measure Student Performance

Variable	Label	Coded values
F22XRPRO	Criterion-measured reading performance {see Ingels:1994tr p. H-34}	0 = below level 1 1 = level 1 2 = level 2 3 = level 3
F22XMPRO	Criterion-measured math performance {see Ingels:1994tr p. H-34}	0 = below level 1 1 = level 1 2 = level 2 3 = level 3 4 = level 4 5 = level 5
MathSGPCategory	SGP: math growth category	1 = low growth 2 = typical growth 3 = high growth
ReadingSGPCategory	SGP: reading growth category	1 = low growth 2 = typical growth 3 = high growth
F22XMTH	Norm-measured math performance	No coding
F22XRTH	Norm-measured reading performance	No coding
F2RHENG2	Grades in English for 12th grade	No coding
F2RHMAG2	Grades in math for 12th grade	No coding
BYGRADS	Grades composite up to and including 8th grade	No coding
Workload	Course workload	1 = low workload 2 = typical workload 3 = high workload
Rigor	Rigor of Schedule	1 = low rigor 2 = typical rigor 3 = high rigor
Enrichment	Enrichment of Schedule	1 = low enrichment 2 = typical enrichment 3 = high enrichment
EDU_LevelAchieve	Education: level achieved	1 = No schooling complete 2 = GED completed 3 = High school diploma completed 4 = Some PSE completed 5 = Associates or certificate completed 6 = Bachelors completed 7 = Masters or above completed
F3EVDOST	Ever dropped out flag	0 = never dropped out of HS 1 = dropped out HS one+

Alternative measures of student performance beyond large-scale assessments were used whenever possible to control for the aforementioned selection bias inherent with following a cohort of students through high school. Since this would include an upward estimate of completing high school, the level of achievement and classroom performance allowed controls for the differing experience levels of the eighth graders in 1988.

Curricular choice.

In this study, the strength of a student's schedule was judged by rigor, enrichment, and workload. Using the courses in the table below, a student could increase the strength of schedule by taking more challenging courses, participating in extracurricular activities, or by taking more courses before graduation.

Table 51

Breakdown of Variables That Measure Curricular Workload

Variable Name	Label
F1S22C	How much coursework in Algebra I
F1S22E	How much coursework in Algebra II
F1S24H	How much coursework in Art
F1S23C	How much coursework in Biology
F1S23E	How much coursework in Chemistry
F1S24O	How much coursework in Computer Science
F1S23D	How much coursework in Earth Science
F1S24A	How much coursework in English
F1S24G	How much coursework in Foreign Language
F1S22A	How much coursework in General math
F1S23A	How much coursework in General science
F1S22D	How much coursework in Geometry
F1S24I	How much coursework in Music
F1S24L	How much coursework in Physical Education
F1S23B	How much coursework in Physical Science
F1S23G	How much coursework in Physics
F1S22B	How much coursework in Pre-Algebra
F1S22G	How much coursework in Pre-Calculus
F1S22F	How much coursework in Trigonometry

Computation of curricular workload.

The composite of coursework completed by 10th grade was computed using the following script in SPSS:

```
COMPUTE WorkLoad=F1S22C + F1S22E + F1S24H + F1S23C + F1S23E + F1S24O + F1S23D +
F1S24A + F1S24G + F1S22A + F1S23A + F1S22D + F1S24I + F1S24L + F1S23B + F1S23G +
F1S22B + F1S22G + F1S22F.
```

Computation of curricular enrichment.

Table 52

Breakdown of Variables That Measure Curricular Enrichment

Variable Name	Label
F1S24H	How much coursework in art
F1S24I	How much coursework in music
F2S29G	Named most valuable player on sport team
F2S29A	Elected officer of school class
F2S29H	Received a community service award
F2S29D	Received recognition for good attendance

The composite of non-core classes and achievements by 10th and 12th grades was computed using the following script in SPSS:

```
COMPUTE Enrichment= F1S24H + F1S24I + F2S29G + F2S29A + F2S29H + F2S29D.
```

Computation of curricular rigor.

Table 53

Breakdown of Variables and Weights That Measure Curricular Rigor

Variable Name	Label	Rigor Weight
F1S22C	How much coursework in Algebra I	-1
F1S22E	How much coursework in Algebra II	1
F1S23C	How much coursework in Biology	1
F1S23E	How much coursework in Chemistry	1
F1S24G	How much coursework in Foreign Language	1
F1S22A	How much coursework in General Math	-1
F1S23A	How much coursework in General Science	-1
F1S22D	How much coursework in Geometry	1
F1S23G	How much coursework in Physics	2
F1S22B	How much coursework in Pre-Algebra	-2
F1S22G	How much coursework in Pre-Calculus	2
F1S22F	How much coursework in Trigonometry	2
F2S13E	Advanced Placement Program	2
BYS66B	In Advanced, Enriched, Accelerated Soc. Studies	2
BYS66A	In Advanced, Enriched, Accelerated English	2
BYS66D	In Advanced, Enriched, Accelerated Math	2
BYS66C	In Advanced, Enriched, Accelerated Science	2
F2S29B	R Won An Academic Honor	2
F2S29F	Received Recognition For Writing Essay/Poem	2

I determined these weights using two criteria. If a student completed a grade-appropriate course, then the weight was 1. If a student completed a course that was considered a remedial course, or one level below remedial, then the weight was -1. This reduced the rigor of the overall course load due to the imbalance of taking a course that does not challenge the student appropriately at his or her grade level. For each course that was advanced level for the grade, the weight of 2 was applied. See table above for a full breakdown of weights and description of curricular rigor. The rigor of the curriculum from eighth to twelfth grades was computed using the following script in SPSS:

```
COMPUTE ScheduleRigor=F1S22C * -1 + F1S22E * 1 + F1S23C * 1 + F1S23E * 1 + F1S24G * 1 +
F1S22A * -1 + F1S23A * -1 + F1S22D * 1 + F1S23G * 2 + F1S22B * -2 + F1S22G * 2 + F1S22F *
2 + F2S13E * 2 + BYS66B * 2 + BYS66A * 2 + BYS66D * 2 + BYS66C * 2 + F2S29B * 2 +
F2S29F * 2 .
```

Large-scale assessment performance.

The IRT-theta *t*-test score was scaled to have a mean of 50 and a standard deviation of 10 across each subject area for all three years of testing. These variables, along with proficiency scores in mathematics, reading, and science were designed to be comparable longitudinally (Ingels, 1994, p. 34).

The large-scale assessment measurements used in this study were status and growth measures. The study used the four cognitive batteries that I limited to reading and mathematics. SGPs were used to measure growth as a value from 0 - 99. This growth calculation was described in the previous section and was categorized by *low growth* as 0-35, *typical growth* as 36-65, and *high growth* as 66-99 (Betebenner, 2007, 2011).

The criterion growth measurement compared year-to-year proficiency gains that were included in the NELS 88:2000 dataset for the cognitive batteries for reading, mathematics, and science, but not for history. The gains were measured by comparing year-to-year growth in each proficiency level. The comparison was made through referencing the ordinal growth levels. There are three levels of proficiency in reading and five in mathematics.

Education level.

The level of education completed was coded using the formula below in SPSS. Dummy variables were used in this calculation for ease of reading this script. Note the overlapping nature of educational levels was accounted for by using the max function for high school diplomas and some post-secondary experience. A student who completed some post-secondary work was considered more advanced than someone who completed only a high school diploma.

```
COMPUTE EDU_LevelAchieve=EDU_NotComplete + EDU_GED * 2 + MAX(EDU_HSDiploma *
3, EDU_SomePSE * 4) + EDU_Certification * 5 + EDU_Bachelor * 6 + EDU_Master * 7.
VARIABLE LABELS  EDU_LevelAchieve 'Education: Level Achieved'.
```

Dependent Variables

Table 54

Summary of Variables That Measure Health Outcomes

Variable	Label
F4IDRINK	Daily consumption over the last 30 days
F4IBINGE	Five or more drinks in last two weeks
DailyCigarettes	Daily Use of Cigarettes

Table 55

Summary of Variables That Measure Societal Outcomes

Variable	**Label**
INVOLVE_Social	Involvement: Social Integration Composite
INVOLVE_Information	Involvement: Informational Integration Composite
INVOLVE_Civic	Involvement: Civic Engagement Composite

Computation of Composite Variable of Social Involvement

Since the variables were stored by either month or hours worked per week, volunteer hours were coded with weights to obtain a common measure of days per month. For the purpose of analysis, volunteer hours are equated to work hours, where seven hours a day is considered full workday. Therefore, the transformed weight for the HRSVLNTR is COMPUTE HRSVLNTR=(HRSVLNTR * 4) / 7. The composite INVOLVE_Social variable was a proxy measurement for days per month the respondent interacted socially within his or her community.

Table 56

Breakdown of Variables That Measure Social Integration

Variable Name	Description	Weight
F4IRELIG	Attend religious activities (days per month)	1
F4ICULT	Attend plays and concerts (days per month)	1
F4ISPORT	Play organized sports (days per month)	1
HRSVLNTR	Volunteered (days per month)	.571

Associated SPSS script:

```
COMPUTE INVOLVE_Social=HRSVLNTR + F4IRELIG + F4ICULT + F4ISPORT.
VARIABLE LABELS  INVOLVE_Social 'Involvement: Social Integration Composite'.
```

Computation of Composite Variable of Informational Involvement

Since the variables were stored by either month or week, they were coded with weights to obtain a common measure of days per month. The composite variable was a proxy measurement for days per month the respondent obtained information through external means.

Table 57

Breakdown of Variables That Measure Informational Engagement

Variable Name	Description	Weight
F4IMAGS	Read newspapers or magazines (days per month)	4
F4IBOOKS	Read books (days per month)	4
F4INET	Use Internet to obtain information (days per month)	4
F4ITVNEW	Watch news on TV (days per month)	4
F4ILIBRY	Go to public library (days per month)	1

Associated SPSS script:

```
COMPUTE INVOLVE_Information=F4ILIBRY + F4IINET + F4IBOOKS + F4IMAGS +
F4ITVNEW.
VARIABLE LABELS  INVOLVE_Information 'Involvement: Informational Integration Composite'.
```

Computation of Composite Variable of Job Satisfaction

Using eight measures of job satisfaction detailed in the table below, I created a composite value of job satisfaction.

Table 58

Breakdown of Variables That Measure Job Satisfaction

Variable	Label
F4BSFRG	Job satisfaction-fringe benefits (2000)
F4BSED2	Job satisfaction-further training (2000)
F4BSSEC	Job satisfaction-job security (2000)
F4BSOVR	Job satisfaction-overall satisfaction (2000)
F4BSPAY	Job satisfaction-pay (2000)
F4BSPRO	Job satisfaction-promotion opportunity (2000)
F4BSED1	Job satisfaction-use of past training (2000)
F4BSIMP	Job satisfaction-work importance (2000)

Associated SPSS script:

```
RECODE F4BSFRG F4BSED2 F4BSSEC F4BSOVR F4BSPAY F4BSPRO F4BSED1 F4BSIMP
(2=0).
DO IF  (F4BSOVR >= 0).
COUNT JobSatisfaction=F4BSFRG F4BSED2 F4BSSEC F4BSOVR F4BSPAY F4BSPRO
F4BSED1 F4BSIMP(1).
VARIABLE LABELS  JobSatisfaction 'Career: Job satisfaction composite'.
END IF.
```

Appendix B – Technical Specifications for Health Outcomes

Binge Drinking

Upon conducting a bivariate Pearson correlation of all available factors in the dataset, three of the following independent variables displayed both statistical significance and accounted for at least 1% of variation in binge drinking responses. Gender, with $R = -0.259$, $p < .01$; yearly family income in 1987, with $R = 0.104$, $p < .01$; and percent free lunch in school, with $R = 0.097$, $p < .01$

The output in the table below reveals that male respondents were more likely to binge drink than female respondents. Further, as family income in 1987 increased, the percentage of respondents who binge drink also increased. The reverse was true for the percentage of students receiving free lunch in the respondent's school. This variable was a proxy measure of average population income in the school. The variables of family income and of school lunch are strongly associated, with $R = .41$ and $p < .01$. As a result, there was reason to suspect interaction effects between the respondent's family income in 1987 and free lunch percent in the respondent's eighth grade school.

Table 59

Frequencies for Predictor Variables as a function of Binge Drinking in 2000

Factor	Binge Drinking in 2000 No	Yes	% Yes	Statistic (p < .001)
Gender				$\chi^2(1) = 709.7$
Male	3,039	1,905	38.5	
Female	4,772	888	15.7	
Yearly Family Income (1987)				$\chi^2(7) = 108.0$
\$0 - \$24,999	2,739	767	21.9	
\$25,000-\$74,999	2,880	1,044	26.6	
\$75,000 or more	1,508	745	33.1	
% Free Lunch in School (8th Grade)				$\chi^2(14) = 118.3$
0-10 (Higher Income)	2,918	1,269	30.3	
11-30	2,398	879	26.8	
30-100 (Lower Income)	2372	586	19.8	

Note. Categories for Yearly Family Income and % Free Lunch are compressed for readability.

Table 60

Frequencies for Testing Variables as a Function of Binge Drinking in 2000

	Binge Drinking in 2000			
	No		Yes	
	M	*SD*	*M*	*SD*
Student Growth Percentile (SGP)				
Math	50	29	52	30
Reading	52	29	48	30

Upon conducting a bivariate correlation among all factors in the regression model, math growth, as indicated previously, was uncorrelated with binge drinking. Reading growth was only weakly negatively correlated with binge drinking, which implies that respondents who improved at a higher rate in relation to their peers, tended to drink in excess slightly less frequently.

Table 61

Correlations for Binge Drinking in 2000 and Predictor Variables

Measure	1	2	3	4	5
1. Binge Drinking (2000)					
2. Gender	.26				
3. Family income (1987)	10	.05			
4. School % free lunch (1988)	.10	03	.41		
5. Reading Growth (1988-92)	.04	03	08	.07	
6. Math Growth (1988-92)	00^{NS}	.10	11	.09	22

Note. Gender is coded as 1 = male, 2 = female. All correlations are significant at p < 0.01 level unless superscripted.
^{NS} Correlation is not significant.
‡ Correlation is significant at the 0.05 level (2-tailed).

Use of Cigarettes

Upon conducting a bivariate Pearson correlation of all available factors in the dataset, two of the following independent factors were displayed as both statistically significant, accounting for at least 1% of variation in cigarette smoking responses. Highest diploma achieved was highly correlated with many other variables such as English and Math grades (R = .59, and R = .50, respectively), whether the student ever dropped out (R = .53), and socioeconomic status (R = .47). When creating the regression models, I selected the two variables below into the logistic regression model to eliminate the confounding effects of the diploma achievement variable. The two variables selected were highest diploma achieved, with R = -.26, p < .01, and ever held back a grade, with R = .09, p < .01

As indicated in the table below, respondents who repeated a grade prior to 1988, prior to eighth grade, tended to smoke approximately 11% more than their peers who did not repeat a grade. The covariate of highest diploma achieved showed a decline in the percentage of respondents who smoked daily as the degree level declined. The most significant drop occurred between the GED and high school diploma levels.

Table 62

Frequencies for Predictor Variables as a Function of Daily Cigarette Use in 2000

	Daily Smoker in 2000			
Factor	No	Yes	% Yes	Statistic ($p < .001$)
Repeated Grade				$\chi^2(1) = 86.6$
No	6703	1903	22.1	
Yes, repeated	935	471	33.5	
Highest Diploma (2000)				$\chi^2(5) = 757.9$
No diploma	225	255	53.1	
GED	264	248	48.4	
High school diploma	3029	1231	28.9	
Associates/Certificate	1220	377	23.6	
Bachelors	2913	406	12.2	
Masters or above	409	37	8.3	

Upon conducting a bivariate correlation of all variables used in the regression model, there was evidence of interaction effects between test performance, highest diploma achieved, and whether the student repeated a grade.

Table 63

Correlations for Cigarette Use in 2000 and Predictor Variables

Measure	1	2	3	4
1. Daily smoker (2000)	–			
2. Highest diploma	-.26	–		
3. Repeated a grade	.09	-.28	–	
4. Reading Growth (1988-92)	-.05	.13	-.04	–
5. Math Growth (1988-92)	-.08	.17	-.06	.22

Note. Held back a grade is coded as 0 = No, 1 = Yes, left back. Highest diploma is coded 0 = no diploma, 1 = GED, 2 = HS diploma, 3 = associates/certificate, 4 = bachelors, 5 = masters or above. All correlations are significant at $p < 0.001$ level unless superscripted.

^{NS} Correlation is not significant.

‡ Correlation is significant at the 0.05 level (2-tailed).

Appendix C – Technical Specifications for Career Outcomes

Yearly Income Level

Upon conducting a bivariate Pearson correlation of all available factors in the dataset, five of the following independent factors displayed as both statistically significant and accounted for at least 1% of variation in 1999 income responses. When creating the regression models, I selected the five variables below into the logistic regression models. The two characteristic variables, gender and family income, were selected due to the high correlation values and independence from the test scores. The three covariates of college degree, hours worked, and whether the respondent attended school were crucial in controlling appropriately for the confounding effects to income. The respondents were mostly aged 25 or 26, and if they were still attending higher education or were working at an entry level position that offered fewer hours could be reflected by the four years of college that the respondent attended. Specifically, a respondent who had a college degree could potentially have four fewer years of work experience than his or her counterpart who worked right after high school, thus resulting in a lower salary and causing a downward bias on college degree estimations. Controlling for these factors allowed the isolation of testing for yearly earnings in 1999 based on the exams that the respondent took as a student from 1988 to 1992.

Bivariate correlations were run against the earnings outcome. Of the background variables, gender accounted for 7% of the variance in the adjusted salary. The strongest predictor of earning was hours worked, which accounted for 20% of the variance in earnings in 1999.

Job Satisfaction

Table 64

Correlations for Job Satisfaction Index in 2000 and Predictor Variables

Measure	1	2	3	4
1. Job Satisfaction Index (2000)	–			
2. College graduate	.07	–		
3. In academic school	-.10	.11	–	
4. Reading growth (1988-92)	-.02[NS]	.12	.04	–
5. Math growth (1988-92)	.02[‡]	.18	.06	.22

Note. College Graduate is coded 0 = No, 1 = Yes, graduated. In academic school is coded 0 = No, 1 = Yes, in school. All correlations are significant at $p < 0.01$ level unless superscripted.

[NS] Correlation is not significant.

[‡] Correlation is significant at the 0.05 level (2-tailed).

Appendix D – Technical Specifications for Societal Outcomes

Voting Habits

Table 65

Correlations for Voting Indicators and Predictor Variables

Measure	1	2	3	4	5
1. Voted in 1998 or 1999	–				
2. College graduate	.11	–			
3. SES Composite	.11	.46	–		
4. Age in months	-.09	-.21	-.26	–	
5. Reading growth (1988-92)	.01^{NS}	.12	.11	-.08	–
6. Math growth (1988-92)	.02^{NS}	.18	.13	-.09	.22

Note. Voting variables are coded 0 = No, 1 = Yes. College Graduate is coded 0 = No, 1 = Yes, graduated. All correlations are significant at $p < 0.01$ level unless superscripted.
^{NS} Correlation is not significant.
‡ Correlation is significant at the 0.05 level (2-tailed).

Social Involvement

A correlation was conducted for all variables included in the regression model. Interestingly, the MVP status of a respondent in 1992 was moderately correlated with reading performance in 1988. On the other hand, math performance in 1988 was completely uncorrelated with MVP status in 1992.

Table 66

Correlations for Voting Indicators and Predictor Variables

Measure	1	2	3	4	5
1. Social Involvement Composite	–				
2. College graduate	.14	–			
3. Named MVP (1992)	.11	.08	–		
4. Reading growth (1988-92)	.02^{NS}	.12	.05	–	
5. Math growth (1988-92)	.05	.18	.02^{NS}	.22	–

Note. College Graduate is coded 0 = No, 1 = Yes, graduated. Named MVP is coded 0 = No, 1 = Yes, named MVP. All correlations are significant at $p < 0.01$ level unless superscripted.
^{NS} Correlation is not significant.
‡ Correlation is significant at the 0.05 level (2-tailed).

Informational Involvement

Table 67

Correlations for Informational Involvement Composite and Predictor Variables

	1	2	3	4	5	6
1. Informational Involvement Composite	–					
2. In School (2000)	.16	–				
3. SES (1988)	.12	.16	–			
4. College graduate	.10	.11	.46	–		
5. In AP program (1992)	.10	.14	.28	.34	–	
7. Reading growth (1988-92)	.00[NS]	.04	.11	.12	.10	–
8. Math growth (1988-92)	.01[NS]	.06	.13	.18	.11	.22

Note. In School in 2000 is coded 0 = No, 1 = Yes, in academic school, College Graduate is coded 0 = No, 1 = Yes, graduated. Was in AP Program is coded 0 = No, 1 = Yes, was in AP program. All correlations are significant at $p < 0.01$ level unless superscripted.

[NS] Correlation is not significant.

‡ Correlation is significant at the 0.05 level (2-tailed).

Appendix E – SPSS Output for Logistic Regression Models

SPSS Output for Binge Drinking

Logistic Regression - Math Performance Growth

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	95% C.I.for OR Upper
Gender (1 = female)	1.25	.05	522.51	1	.00	3.47	3.12	3.87
Socio-economic status (1988) [quarter 1: low]*			10.46	3	.02			
Quarter 2	.04	.08	0.20	1	.66	1.04	0.88	1.22
Quarter 3	.23	.08	8.26	1	.00	1.26	1.08	1.47
Quarter 4 [high]	.14	.08	2.86	1	.09	1.15	0.98	1.35
School % free lunch (1988) [low poverty]*			55.86	2	.00			
11% - 30% [middle]	-.14	.06	5.26	1	.02	0.87	0.77	0.98
31% - 100% [high poverty]	-.54	.07	54.87	1	.00	0.58	0.50	0.67
Math SGP Category [low growth]*			1.39	2	.50			
Typical growth	-.04	.07	0.30	1	.59	0.96	0.84	1.10
High growth	-.07	.06	1.38	1	.24	0.93	0.82	1.05
Constant	-1.57	.09	328.29	1	.00	0.21		

* Reference group

Logistic Regression - Reading Performance Growth

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	95% C.I.for OR Upper
Gender (1 = female)	1.23	.05	514.98	1	.00	3.43	3.08	3.81
Socio-economic status (1988) [quarter 1: low]*			11.48	3	.01			
Quarter 2	.04	.08	0.24	1	.63	1.04	0.89	1.22
Quarter 3	.24	.08	9.13	1	.00	1.27	1.09	1.49
Quarter 4 [high]	.15	.08	3.14	1	.08	1.16	0.98	1.36
School % free lunch (1988) [low poverty]*			57.31	2	.00			
11% - 30% [middle]	-.14	.06	4.90	1	.03	0.87	0.77	0.98
31% - 100% [high poverty]	-.55	.07	56.07	1	.00	0.58	0.50	0.67
Reading SGP Category [low growth]*			13.72	2	.00			
Typical growth	-.15	.07	4.88	1	.03	0.86	0.76	0.98
High growth	-.23	.06	13.37	1	.00	0.79	0.70	0.90
Constant	-1.48	.09	288.55	1	.00	0.23		

Logistic Regression by Ethnicity - Math Performance Growth

Ethnicity = Asian or Pacific Islander

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	95% C.I.for OR Upper
Gender (1 = female)	1.48	.29	26.81	1	.00	4.39	2.51	7.68
Socio-economic status (1988) [quarter 1: low]*			1.68	3	.64			
Quarter 2	.34	.45	0.57	1	.45	1.41	0.58	3.41
Quarter 3	.40	.44	0.81	1	.37	1.49	0.63	3.53
Quarter 4 [high]	.04	.42	0.01	1	.93	1.04	0.46	2.35
School % free lunch (1988) [low poverty]*			3.39	2	.18			
11% - 30% [middle]	.19	.31	0.37	1	.54	1.21	0.66	2.23
31% - 100% [high poverty]	-.49	.37	1.70	1	.19	0.61	0.30	1.28

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
Math SGP Category [low growth]*			3.61	2	.16			
Typical growth	-.76	.40	3.60	1	.06	0.47	0.21	1.03
High growth	-.28	.31	0.82	1	.37	0.75	0.41	1.39
Constant	-2.08	.51	16.75	1	.00	0.12		

Ethnicity = 3 Black, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	1.78	.27	42.18	1	.00	5.91	3.46	10.11
Socio-economic status (1988) [quarter 1: low]*			2.12	3	.55			
Quarter 2	-.39	.34	1.37	1	.24	0.67	0.35	1.30
Quarter 3	-.10	.35	0.09	1	.77	0.90	0.46	1.78
Quarter 4 [high]	.13	.35	0.14	1	.71	1.14	0.57	2.28
School % free lunch (1988) [low poverty]*			5.27	2	.07			
11% - 30% [middle]	-.36	.35	1.11	1	.29	0.69	0.35	1.37
31% - 100% [high poverty]	-.71	.31	5.17	1	.02	0.49	0.27	0.91
Math SGP Category [low growth]*			2.26	2	.32			
Typical growth	-.14	.32	0.18	1	.67	0.87	0.46	1.64
High growth	.31	.29	1.19	1	.28	1.37	0.78	2.39
Constant	-2.41	.40	35.41	1	.00	0.09		

Ethnicity = 4 White, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	1.20	.06	370.65	1	.00	3.33	2.94	3.76
Socio-economic status (1988) [quarter 1: low]*			5.56	3	.14			
Quarter 2	.08	.10	0.59	1	.44	1.08	0.89	1.32
Quarter 3	.21	.10	4.54	1	.03	1.23	1.02	1.49
Quarter 4 [high]	.17	.10	2.97	1	.08	1.19	0.98	1.44
School % free lunch (1988) [low poverty]*			26.67	2	.00			
11% - 30% [middle]	-.15	.07	4.42	1	.04	0.86	0.75	0.99
31% - 100% [high poverty]	-.47	.09	26.64	1	.00	0.63	0.52	0.75
Math SGP Category [low growth]*			0.31	2	.86			
Typical growth	-.02	.08	0.05	1	.82	0.98	0.84	1.14
High growth	-.04	.07	0.30	1	.58	0.96	0.83	1.11
Constant	-1.50	.10	209.54	1	.00	0.22		

Ethnicity = 5 Hispanic or Latino

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	1.23	.16	59.37	1	.00	3.43	2.51	4.69
Socio-economic status (1988) [quarter 1: low]*			5.51	3	.14			
Quarter 2	-.19	.21	0.86	1	.35	0.83	0.55	1.24
Quarter 3	.36	.22	2.85	1	.09	1.44	0.94	2.20
Quarter 4 [high]	.08	.25	0.10	1	.76	1.08	0.66	1.76
School % free lunch (1988) [low poverty]*			3.94	2	.14			
11% - 30% [middle]	-.05	.22	0.05	1	.83	0.95	0.62	1.46
31% - 100% [high poverty]	-.35	.20	3.08	1	.08	0.71	0.48	1.04
Math SGP Category [low growth]*			2.02	2	.36			
Typical growth	.06	.19	0.10	1	.75	1.06	0.73	1.55
High growth	-.20	.19	1.11	1	.29	0.82	0.57	1.18
Constant	-1.54	.23	43.93	1	.00	0.21		

Logistic Regression by Ethnicity - Reading Performance Growth

Ethnicity = 2 Asian or Pacific Islander

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	1.45	.29	25.41	1	.00	4.26	2.42	7.48
Socio-economic status (1988) [quarter 1: low]*			1.39	3	.71			
Quarter 2	.32	.46	0.47	1	.49	1.38	0.55	3.42
Quarter 3	.42	.45	0.89	1	.35	1.53	0.63	3.68
Quarter 4 [high]	.10	.43	0.06	1	.81	1.11	0.48	2.56
School % free lunch (1988) [low poverty]*			4.62	2	.10			
11% - 30% [middle]	.29	.31	0.88	1	.35	1.33	0.73	2.43
31% - 100% [high poverty]	-.54	.38	1.98	1	.16	0.59	0.28	1.23
Reading SGP Category [low growth]*			0.03	2	.98			
Typical growth	.00	.36	0.00	1	1.00	1.00	0.50	2.01
High growth	-.05	.33	0.02	1	.88	0.95	0.50	1.81
Constant	-2.43	.52	22.02	1	.00	0.09		

Ethnicity = 3 Black, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	1.76	.27	41.29	1	.00	5.82	3.40	9.95
Socio-economic status (1988) [quarter 1: low]*			2.01	3	.57			
Quarter 2	-.38	.34	1.30	1	.25	0.68	0.35	1.32
Quarter 3	-.07	.35	0.04	1	.85	0.94	0.47	1.85
Quarter 4 [high]	.13	.35	0.13	1	.72	1.14	0.57	2.28
School % free lunch (1988) [low poverty]*			6.39	2	.04			
11% - 30% [middle]	-.40	.34	1.37	1	.24	0.67	0.34	1.31
31% - 100% [high poverty]	-.78	.31	6.29	1	.01	0.46	0.25	0.84
Reading SGP Category [low growth]*			0.64	2	.73			
Typical growth	-.24	.30	0.64	1	.42	0.79	0.44	1.42
High growth	-.08	.30	0.06	1	.80	0.93	0.51	1.68
Constant	-2.19	.40	29.85	1	.00	0.11		

Ethnicity = 4 White, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	1.19	.06	367.02	1	.00	3.29	2.91	3.71
Socio-economic status (1988) [quarter 1: low]*			6.61	3	.09			
Quarter 2	.08	.10	0.66	1	.42	1.09	0.89	1.32
Quarter 3	.22	.10	5.32	1	.02	1.25	1.03	1.51
Quarter 4 [high]	.19	.10	3.59	1	.06	1.21	0.99	1.46
School % free lunch (1988) [low poverty]*			26.27	2	.00			
11% - 30% [middle]	-.14	.07	4.00	1	.05	0.87	0.76	1.00
31% - 100% [high poverty]	-.47	.09	26.26	1	.00	0.63	0.53	0.75
Reading SGP Category [low growth]*			12.03	2	.00			
Typical growth	-.12	.08	2.45	1	.12	0.89	0.76	1.03
High growth	-.25	.07	12.03	1	.00	0.78	0.68	0.90
Constant	-1.40	.10	181.20	1	.00	0.25		

Ethnicity = 5 Hispanic or Latino

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	1.23	.16	59.52	1	.00	3.42	2.50	4.67
Socio-economic status (1988) [quarter 1: low]*			5.48	3	.14			
Quarter 2	-.21	.21	1.03	1	.31	0.81	0.54	1.21
Quarter 3	.35	.22	2.63	1	.10	1.42	0.93	2.16
Quarter 4 [high]	.03	.25	0.02	1	.90	1.03	0.63	1.68
School % free lunch (1988) [low poverty]*			4.17	2	.12			
11% - 30% [middle]	-.06	.22	0.08	1	.78	0.94	0.61	1.44
31% - 100% [high poverty]	-.36	.20	3.34	1	.07	0.69	0.47	1.03
Reading SGP Category [low growth]*			2.26	2	.32			
Typical growth	-.28	.19	2.26	1	.13	0.75	0.52	1.09
High growth	-.13	.19	0.49	1	.49	0.88	0.61	1.26
Constant	-1.43	.24	35.47	1	.00	0.24		

Logistic Regression by Gender - Math Performance Growth

Gender = 1 Male

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Socio-economic status (1988) [quarter 1: low]*			1.60	3	.66			
Quarter 2	.07	.10	0.42	1	.52	1.07	0.87	1.31
Quarter 3	.13	.10	1.59	1	.21	1.14	0.93	1.39
Quarter 4 [high]	.07	.10	0.47	1	.50	1.07	0.88	1.32
School % free lunch (1988) [low poverty]*			25.32	2	.00			
11% - 30% [middle]	-.06	.08	0.54	1	.46	0.94	0.80	1.10
31% - 100% [high poverty]	-.44	.09	22.80	1	.00	0.64	0.53	0.77
Math SGP Category [low growth]*			0.85	2	.65			
Typical growth	.08	.09	0.85	1	.36	1.09	0.91	1.29
High growth	.04	.08	0.21	1	.65	1.04	0.89	1.22
Constant	-.41	.11	15.16	1	.00	0.66		

Gender = 2 Female

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Socio-economic status (1988) [quarter 1: low]*			14.20	3	.00			
Quarter 2	-.02	.14	0.03	1	.86	0.98	0.75	1.27
Quarter 3	.37	.13	8.27	1	.00	1.45	1.12	1.86
Quarter 4 [high]	.26	.13	3.85	1	.05	1.30	1.00	1.68
School % free lunch (1988) [low poverty]*			33.90	2	.00			
11% - 30% [middle]	-.27	.10	7.31	1	.01	0.76	0.63	0.93
31% - 100% [high poverty]	-.70	.12	33.61	1	.00	0.50	0.39	0.63
Math SGP Category [low growth]*			6.62	2	.04			
Typical growth	-.21	.11	3.79	1	.05	0.81	0.66	1.00
High growth	-.24	.10	5.49	1	.02	0.79	0.65	0.96
Constant	-1.46	.13	129.31	1	.00	0.23		

Logistic Regression by Gender - Reading Performance Growth

Gender = 1 Male

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
Socio-economic status (1988) [quarter 1: low]*			2.41	3	.49			
Quarter 2	.08	.10	0.55	1	.46	1.08	0.88	1.32
Quarter 3	.16	.10	2.39	1	.12	1.17	0.96	1.43
Quarter 4 [high]	.09	.10	0.71	1	.40	1.09	0.89	1.34
School % free lunch (1988) [low poverty]*			26.43	2	.00			
11% - 30% [middle]	-.06	.08	0.53	1	.47	0.94	0.80	1.10
31% - 100% [high poverty]	-.45	.09	23.76	1	.00	0.63	0.53	0.76
Reading SGP Category [low growth]*			11.90	2	.00			
Typical growth	-.21	.08	6.41	1	.01	0.81	0.68	0.95
High growth	-.26	.08	10.41	1	.00	0.77	0.66	0.90
Constant	-.24	.10	5.36	1	.02	0.79		

Gender = 2 Female

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Socio-economic status (1988) [quarter 1: low]*			13.51	3	.00			
Quarter 2	-.02	.14	0.02	1	.88	0.98	0.75	1.28
Quarter 3	.36	.13	8.07	1	.00	1.44	1.12	1.85
Quarter 4 [high]	.25	.13	3.48	1	.06	1.28	0.99	1.66
School % free lunch (1988) [low poverty]*			34.10	2	.00			
11% - 30% [middle]	-.26	.10	6.63	1	.01	0.77	0.64	0.94
31% - 100% [high poverty]	-.70	.12	33.97	1	.00	0.49	0.39	0.63
Reading SGP Category [low growth]*			3.79	2	.15			
Typical growth	-.04	.11	0.13	1	.72	0.96	0.78	1.18
High growth	-.19	.10	3.39	1	.07	0.83	0.67	1.01
Constant	-1.51	.13	134.06	1	.00	0.22		

SPSS Output for Cigarette Use

Logistic Regression - Math Performance Growth

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-1.05	.07	246.35	1	.00	0.35	0.31	0.40
Repeated a grade (1 = yes, repeated)	0.23	.08	8.12	1	.00	1.26	1.08	1.48
Math SGP Category [low growth]*			15.48	2	.00			
Typical growth	-0.21	.07	8.81	1	.00	0.81	0.70	0.93
High growth	-0.24	.07	13.10	1	.00	0.79	0.69	0.90
Constant	-0.83	.05	286.82	1	.00	0.44		

Logistic Regression - Reading Performance Growth

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-1.07	.07	259.06	1	.00	0.34	0.30	0.39
Repeated a grade (1 = yes, repeated)	0.24	.08	9.01	1	.00	1.27	1.09	1.49
Reading SGP Category [low growth]*			7.14	2	.03			
Typical growth	-0.16	.07	5.38	1	.02	0.85	0.74	0.98
High growth	-0.15	.07	4.99	1	.03	0.86	0.75	0.98
Constant	-0.87	.05	295.62	1	.00	0.42		

Logistic Regression by Ethnicity - Math Performance Growth

Ethnicity = 2 Asian or Pacific Islander

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-1.08	.31	12.45	1	.00	0.34	0.19	0.62
Repeated a grade (1 = yes, repeated)	-1.88	1.04	3.27	1	.07	0.15	0.02	1.17
Math SGP Category [low growth]*			1.26	2	.53			
Typical growth	-0.50	.46	1.21	1	.27	0.61	0.25	1.48
High growth	-0.13	.35	0.14	1	.71	0.88	0.44	1.75
Constant	-1.03	.31	11.34	1	.00	0.36		

Ethnicity = 3 Black, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-0.83	.38	4.66	1	.03	0.44	0.21	0.93
Repeated a grade (1 = yes, repeated)	0.36	.33	1.19	1	.27	1.43	0.75	2.73
Math SGP Category [low growth]*			0.93	2	.63			
Typical growth	0.33	.35	0.93	1	.34	1.40	0.71	2.75
High growth	0.19	.34	0.30	1	.58	1.21	0.62	2.36
Constant	-2.21	.26	71.15	1	.00	0.11		

Ethnicity = 4 White, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-1.12	.07	223.33	1	.00	0.33	0.28	0.38
Repeated a grade (1 = yes, repeated)	0.35	.10	12.85	1	.00	1.42	1.17	1.72
Math SGP Category [low growth]*			14.56	2	.00			
Typical growth	-0.21	.08	6.74	1	.01	0.81	0.69	0.95
High growth	-0.28	.08	13.25	1	.00	0.75	0.65	0.88
Constant	-0.66	.06	139.80	1	.00	0.52		

Ethnicity = 5 Hispanic or Latino

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-0.79	.25	9.84	1	.00	0.46	0.28	0.74
Repeated a grade (1 = yes, repeated)	0.05	.23	0.04	1	.85	1.05	0.66	1.65
Math SGP Category [low growth]*			1.67	2	.43			
Typical growth	-0.21	.22	0.94	1	.33	0.81	0.53	1.24
High growth	-0.25	.21	1.44	1	.23	0.78	0.52	1.17
Constant	-1.28	.15	68.98	1	.00	0.28		

Logistic Regression by Ethnicity - Reading Performance Growth

Ethnicity = 2 Asian or Pacific Islander

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-1.08	.31	12.35	1	.00	0.34	0.19	0.62
Repeated a grade (1 = yes, repeated)	-1.88	1.04	3.26	1	.07	0.15	0.02	1.17
Reading SGP Category [low growth]*			0.51	2	.77			
Typical growth	-0.26	.41	0.42	1	.51	0.77	0.35	1.70
High growth	-0.22	.36	0.36	1	.55	0.80	0.39	1.64
Constant	-1.04	.31	11.52	1	.00	0.35		

Ethnicity = 3 Black, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-0.77	.38	4.05	1	.04	0.46	0.22	0.98
Repeated a grade (1 = yes, repeated)	0.36	.33	1.20	1	.27	1.43	0.75	2.73
Reading SGP Category [low growth]*			1.98	2	.37			
Typical growth	-0.49	.35	1.97	1	.16	0.61	0.31	1.21
High growth	-0.19	.34	0.31	1	.58	0.83	0.43	1.61
Constant	-1.86	.23	67.03	1	.00	0.16		

Ethnicity = 4 White, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-1.15	.07	239.92	1	.00	0.32	0.27	0.37
Repeated a grade (1 = yes, repeated)	0.37	.10	14.68	1	.00	1.45	1.20	1.75
Reading SGP Category [low growth]*			4.21	2	.12			
Typical growth	-0.14	.08	3.06	1	.08	0.87	0.74	1.02
High growth	-0.14	.08	3.09	1	.08	0.87	0.75	1.02
Constant	-0.71	.06	149.88	1	.00	0.49		

Ethnicity = 5 Hispanic or Latino

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-0.79	.25	9.97	1	.00	0.45	0.28	0.74
Repeated a grade (1 = yes, repeated)	0.04	.23	0.02	1	.88	1.04	0.65	1.64
Reading SGP Category [low growth]*			7.74	2	.02			
Typical growth	-0.41	.21	3.91	1	.05	0.66	0.44	1.00
High growth	-0.55	.21	6.67	1	.01	0.58	0.38	0.88
Constant	-1.13	.15	60.57	1	.00	0.32		

Logistic Regression by Gender - Math Performance Growth

Gender = 1 Male

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-1.09	.10	126.77	1	.00	0.34	0.28	0.41
Repeated a grade (1 = yes, repeated)	0.19	.11	3.29	1	.07	1.21	0.98	1.50
Math SGP Category [low growth]*			3.30	2	.19			
Typical growth	-0.14	.10	1.88	1	.17	0.87	0.71	1.06
High growth	-0.16	.10	2.94	1	.09	0.85	0.70	1.02
Constant	-0.75	.08	98.44	1	.00	0.47		

Gender = 2 Female

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-0.99	.09	114.87	1	.00	0.37	0.31	0.45
Repeated a grade (1 = yes, repeated)	0.21	.13	2.86	1	.09	1.24	0.97	1.59
Math SGP Category [low growth]*			18.10	2	.00			
Typical growth	-0.30	.10	9.15	1	.00	0.74	0.61	0.90
High growth	-0.38	.10	15.39	1	.00	0.69	0.57	0.83

Constant	-0.89	.06	187.89	1	.00	0.41		

Logistic Regression by Gender - Reading Performance Growth

Gender = 1 Male

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-1.10	.10	129.51	1	.00	0.33	0.28	0.40
Repeated a grade (1 = yes, repeated)	0.20	.11	3.52	1	.06	1.22	0.99	1.50
Reading SGP Category [low growth]*			2.82	2	.24			
Typical growth	-0.12	.10	1.51	1	.22	0.88	0.72	1.08
High growth	-0.15	.09	2.46	1	.12	0.86	0.72	1.04
Constant	-0.76	.07	115.54	1	.00	0.47		

Gender = 2 Female

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	-1.04	.09	127.54	1	.00	0.35	0.30	0.42
Repeated a grade (1 = yes, repeated)	0.24	.13	3.70	1	.05	1.27	1.00	1.63
Reading SGP Category [low growth]*			3.92	2	.14			
Typical growth	-0.18	.10	3.29	1	.07	0.84	0.69	1.01
High growth	-0.15	.10	2.41	1	.12	0.86	0.71	1.04
Constant	-0.96	.07	180.93	1	.00	0.38		

SPSS Output for Yearly Income

Logistic Regression - Math Performance Growth

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	-1.05	.05	443.64	1	.00	0.35	0.32	0.38
Socio-economic status (1988) [quarter 1: low]*			32.78	3	.00			
Quarter 2	0.28	.07	14.81	1	.00	1.32	1.15	1.52
Quarter 3	0.37	.07	26.19	1	.00	1.45	1.26	1.67
Quarter 4 [high]	0.38	.08	25.05	1	.00	1.47	1.26	1.71
In Academic School (1 = yes)	-0.84	.06	170.87	1	.00	0.43	0.38	0.49
College graduate (1 = yes, graduate)	0.91	.06	253.02	1	.00	2.49	2.22	2.78
Math SGP Category [low growth]*			1.96	2	.38			
Typical growth	0.09	.06	1.96	1	.16	1.09	0.97	1.23
High growth	0.04	.06	0.47	1	.49	1.04	0.93	1.17
Constant	0.00	.07	0.00	1	.96	1.00		

* Reference category

Logistic Regression - Reading Performance Growth

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	-1.06	.05	450.03	1	.00	0.35	0.31	0.38
Socio-economic status (1988) [quarter 1: low]*			32.45	3	.00			
Quarter 2	0.27	.07	14.15	1	.00	1.31	1.14	1.51
Quarter 3	0.37	.07	25.59	1	.00	1.44	1.25	1.66
Quarter 4 [high]	0.39	.08	25.29	1	.00	1.47	1.27	1.71
In Academic School (1 = yes)	-0.84	.06	171.01	1	.00	0.43	0.38	0.49
College graduate (1 = yes, graduate)	0.93	.06	266.54	1	.00	2.54	2.27	2.84
Reading SGP Category [low growth]*			3.48	2	.18			

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
Typical growth	0.02	.06	0.07	1	.79	1.02	0.90	1.15
High growth	-0.09	.06	2.20	1	.14	0.92	0.82	1.03
Constant	0.06	.07	0.79	1	.37	1.06		

Logistic Regression by Ethnicity - Math Performance Growth

Ethnicity = 2 Asian or Pacific Islander

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	-0.68	.23	8.52	1	.00	0.50	0.32	0.80
Socio-economic status (1988) [quarter 1: low]*			1.70	3	.64			
Quarter 2	-0.20	.38	0.28	1	.59	0.82	0.39	1.71
Quarter 3	0.27	.36	0.53	1	.47	1.30	0.64	2.66
Quarter 4 [high]	0.02	.33	0.00	1	.95	1.02	0.53	1.97
In Academic School (1 = yes)	-1.79	.26	48.99	1	.00	0.17	0.10	0.28
College graduate (1 = yes, graduate)	0.76	.25	9.19	1	.00	2.14	1.31	3.51
Math SGP Category [low growth]*			2.91	2	.23			
Typical growth	0.16	.33	0.23	1	.63	1.17	0.62	2.21
High growth	0.46	.29	2.63	1	.10	1.59	0.91	2.78
Constant	0.44	.34	1.62	1	.20	1.55		

Ethnicity = 3 Black, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	-0.83	.19	19.77	1	.00	0.43	0.30	0.63
Socio-economic status (1988) [quarter 1: low]*			8.95	3	.03			
Quarter 2	0.09	.24	0.13	1	.72	1.09	0.68	1.73
Quarter 3	0.72	.25	8.22	1	.00	2.06	1.26	3.38
Quarter 4 [high]	0.27	.29	0.87	1	.35	1.31	0.75	2.29
In Academic School (1 = yes)	-0.05	.23	0.05	1	.83	0.95	0.61	1.49
College graduate (1 = yes, graduate)	1.11	.22	25.68	1	.00	3.05	1.98	4.69
Math SGP Category [low growth]*			1.19	2	.55			
Typical growth	0.22	.23	0.93	1	.33	1.25	0.80	1.95
High growth	0.20	.22	0.80	1	.37	1.22	0.79	1.87
Constant	-0.57	.22	6.68	1	.01	0.57		

Ethnicity = 4 White, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	-1.12	.06	353.41	1	.00	0.33	0.29	0.37
Socio-economic status (1988) [quarter 1: low]*			19.96	3	.00			
Quarter 2	0.32	.09	11.55	1	.00	1.37	1.14	1.65
Quarter 3	0.32	.09	12.05	1	.00	1.38	1.15	1.65
Quarter 4 [high]	0.42	.10	18.75	1	.00	1.53	1.26	1.85
In Academic School (1 = yes)	-0.83	.08	111.32	1	.00	0.44	0.37	0.51
College graduate (1 = yes, graduate)	0.82	.07	146.64	1	.00	2.28	1.99	2.60
Math SGP Category [low growth]*			2.55	2	.28			
Typical growth	0.12	.07	2.54	1	.11	1.13	0.97	1.30
High growth	0.05	.07	0.47	1	.49	1.05	0.91	1.20
Constant	0.08	.09	0.90	1	.34	1.09		

Ethnicity = 5 Hispanic or Latino

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	-0.97	.14	45.62	1	.00	0.38	0.28	0.50
Socio-economic status (1988) [quarter 1: low]*			5.32	3	.15			
Quarter 2	0.26	.18	1.99	1	.16	1.29	0.91	1.84

	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Quarter 3	0.37	.20	3.46	1	.06	1.45	0.98	2.16
Quarter 4 [high]	-0.06	.24	0.07	1	.79	0.94	0.59	1.49
In Academic School (1 = yes)	-0.93	.18	25.50	1	.00	0.39	0.27	0.56
College graduate (1 = yes, graduate)	1.20	.18	42.71	1	.00	3.33	2.32	4.78
Math SGP Category [low growth]*			1.04	2	.59			
Typical growth	-0.10	.18	0.28	1	.59	0.91	0.64	1.29
High growth	-0.17	.17	1.04	1	.31	0.84	0.60	1.17
Constant	0.02	.16	0.02	1	.90	1.02		

Logistic Regression by Ethnicity - Reading Performance Growth

Ethnicity = 2 Asian or Pacific Islander

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	-0.72	.24	9.30	1	.00	0.49	0.31	0.77
Socio-economic status (1988) [quarter 1: low]*			1.92	3	.59			
Quarter 2	-0.28	.38	0.55	1	.46	0.75	0.36	1.58
Quarter 3	0.21	.37	0.34	1	.56	1.24	0.60	2.53
Quarter 4 [high]	0.02	.34	0.01	1	.94	1.02	0.53	1.98
In Academic School (1 = yes)	-1.69	.25	47.42	1	.00	0.18	0.11	0.30
College graduate (1 = yes, graduate)	0.84	.26	10.89	1	.00	2.33	1.41	3.84
Reading SGP Category [low growth]*			3.79	2	.15			
Typical growth	-0.47	.32	2.14	1	.14	0.63	0.33	1.17
High growth	-0.55	.29	3.65	1	.06	0.58	0.33	1.01
Constant	1.08	.34	9.78	1	.00	2.93		

Ethnicity = 3 Black, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	-0.81	.19	18.75	1	.00	0.44	0.31	0.64
Socio-economic status (1988) [quarter 1: low]*			8.61	3	.03			
Quarter 2	0.04	.24	0.03	1	.87	1.04	0.65	1.66
Quarter 3	0.70	.25	7.57	1	.01	2.00	1.22	3.29
Quarter 4 [high]	0.23	.29	0.66	1	.42	1.26	0.72	2.21
In Academic School (1 = yes)	-0.07	.23	0.08	1	.77	0.94	0.60	1.47
College graduate (1 = yes, graduate)	1.18	.22	29.13	1	.00	3.27	2.13	5.03
Reading SGP Category [low growth]*			0.37	2	.83			
Typical growth	0.06	.22	0.08	1	.78	1.06	0.69	1.63
High growth	-0.08	.23	0.13	1	.72	0.92	0.59	1.45
Constant	-0.44	.21	4.61	1	.03	0.64		

Ethnicity = 4 White, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	-1.13	.06	362.82	1	.00	0.32	0.29	0.36
Socio-economic status (1988) [quarter 1: low]*			20.33	3	.00			
Quarter 2	0.32	.09	11.48	1	.00	1.37	1.14	1.65
Quarter 3	0.32	.09	12.17	1	.00	1.38	1.15	1.65
Quarter 4 [high]	0.43	.10	19.22	1	.00	1.53	1.27	1.86
In Academic School (1 = yes)	-0.82	.08	109.07	1	.00	0.44	0.38	0.51
College graduate (1 = yes, graduate)	0.85	.07	157.53	1	.00	2.33	2.04	2.66
Reading SGP Category [low growth]*			2.89	2	.24			
Typical growth	0.02	.07	0.09	1	.76	1.02	0.88	1.18
High growth	-0.09	.07	1.71	1	.19	0.91	0.80	1.05
Constant	0.15	.09	3.03	1	.08	1.16		

Ethnicity = 5 Hispanic or Latino

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Gender (1 = female)	-0.95	.14	43.76	1	.00	0.39	0.29	0.51
Socio-economic status (1988) [quarter 1: low]*			4.69	3	.20			
Quarter 2	0.24	.18	1.73	1	.19	1.27	0.89	1.80

	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Quarter 3	0.37	.20	3.30	1	.07	1.44	0.97	2.14
Quarter 4 [high]	-0.03	.24	0.02	1	.88	0.97	0.61	1.54
In Academic School (1 = yes)	-0.95	.18	26.41	1	.00	0.39	0.27	0.56
College graduate (1 = yes, graduate)	1.21	.18	42.99	1	.00	3.35	2.33	4.80
Reading SGP Category [low growth]*			0.25	2	.88			
Typical growth	0.06	.17	0.13	1	.72	1.07	0.76	1.50
High growth	-0.02	.17	0.02	1	.90	0.98	0.70	1.37
Constant	-0.08	.16	0.25	1	.62	0.92		

Logistic Regression by Gender - Math Performance Growth

Gender = 1 Male

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Socio-economic status (1988) [quarter 1: low]*			11.70	3	.01			
Quarter 2	0.21	.10	4.20	1	.04	1.24	1.01	1.52
Quarter 3	0.35	.10	11.32	1	.00	1.42	1.16	1.74
Quarter 4 [high]	0.26	.11	5.69	1	.02	1.30	1.05	1.61
In Academic School (1 = yes)	-1.19	.09	177.90	1	.00	0.30	0.25	0.36
College graduate (1 = yes, graduate)	0.44	.09	26.38	1	.00	1.55	1.31	1.83
Math SGP Category [low growth]*			0.13	2	.94			
Typical growth	0.02	.09	0.07	1	.79	1.03	0.85	1.23
High growth	0.03	.08	0.12	1	.73	1.03	0.87	1.22
Constant	0.29	.09	11.19	1	.00	1.34		

Gender = 2 Female

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Socio-economic status (1988) [quarter 1: low]*			27.66	3	.00			
Quarter 2	0.36	.11	11.56	1	.00	1.43	1.17	1.77
Quarter 3	0.41	.11	15.09	1	.00	1.51	1.23	1.86
Quarter 4 [high]	0.57	.11	26.86	1	.00	1.77	1.42	2.19
In Academic School (1 = yes)	-0.48	.09	28.75	1	.00	0.62	0.52	0.74
College graduate (1 = yes, graduate)	1.27	.08	274.01	1	.00	3.57	3.07	4.15
Math SGP Category [low growth]*			1.96	2	.37			
Typical growth	0.12	.09	1.95	1	.16	1.13	0.95	1.34
High growth	0.05	.08	0.29	1	.59	1.05	0.89	1.23
Constant	-1.39	.09	232.56	1	.00	0.25		

Logistic Regression by Gender - Reading Performance Growth

Gender = 1 Male

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Socio-economic status (1988) [quarter 1: low]*			11.14	3	.01			
Quarter 2	0.20	.10	3.69	1	.05	1.22	1.00	1.50
Quarter 3	0.34	.10	10.74	1	.00	1.41	1.15	1.72
Quarter 4 [high]	0.26	.11	5.61	1	.02	1.30	1.05	1.61
In Academic School (1 = yes)	-1.19	.09	176.73	1	.00	0.31	0.26	0.36
College graduate (1 = yes, graduate)	0.45	.08	28.55	1	.00	1.57	1.33	1.86
Reading SGP Category [low growth]*			1.80	2	.41			
Typical growth	0.03	.09	0.14	1	.71	1.03	0.87	1.23
High growth	-0.08	.08	0.94	1	.33	0.92	0.78	1.09
Constant	0.33	.09	14.61	1	.00	1.39		

Gender = 2 Female

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
Socio-economic status (1988) [quarter 1: low]*			28.29	3	.00			

	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Quarter 2	0.36	.11	11.47	1	.00	1.43	1.16	1.76
Quarter 3	0.41	.11	14.96	1	.00	1.51	1.22	1.85
Quarter 4 [high]	0.58	.11	27.63	1	.00	1.78	1.44	2.21
In Academic School (1 = yes)	-0.49	.09	29.15	1	.00	0.62	0.52	0.73
College graduate (1 = yes, graduate)	1.30	.08	286.82	1	.00	3.65	3.14	4.24
Reading SGP Category [low growth]*			3.01	2	.22			
Typical growth	0.00	.09	0.00	1	.99	1.00	0.84	1.19
High growth	-0.13	.08	2.22	1	.14	0.88	0.75	1.04
Constant	-1.30	.09	197.88	1	.00	0.27		

SPSS Output for Job Satisfaction

Logistic Regression - Math Performance Growth

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
College graduate (1 = yes, graduate)	0.28	.05	34.77	1	.00	1.32	1.21	1.45
In Academic School (1 = yes)	-0.56	.06	97.46	1	.00	0.57	0.51	0.64
Math SGP Category [low growth]*			2.61	2	.27			
Typical growth	0.00	.06	0.00	1	.97	1.00	0.89	1.12
High growth	0.08	.05	1.96	1	.16	1.08	0.97	1.20
Constant	0.29	.04	51.14	1	.00	1.34		

Logistic Regression - Reading Performance Growth

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
College graduate (1 = yes, graduate)	0.30	.05	41.56	1	.00	1.36	1.24	1.49
In Academic School (1 = yes)	-0.57	.06	99.04	1	.00	0.57	0.51	0.63
Reading SGP Category [low growth]*			3.75	2	.15			
Typical growth	-0.05	.06	0.78	1	.38	0.95	0.85	1.06
High growth	-0.10	.05	3.74	1	.05	0.90	0.81	1.00
Constant	0.36	.04	75.07	1	.00	1.43		

Logistic Regression by Ethnicity - Math Performance Growth

Ethnicity = 2 Asian or Pacific Islander

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
College graduate (1 = yes, graduate)	0.26	.21	1.65	1	.20	1.30	0.87	1.95
In Academic School (1 = yes)	-0.92	.22	17.54	1	.00	0.40	0.26	0.61
Math SGP Category [low growth]*			1.72	2	.42			
Typical growth	0.16	.29	0.30	1	.58	1.17	0.67	2.06
High growth	-0.16	.24	0.45	1	.50	0.85	0.53	1.37
Constant	0.26	.23	1.32	1	.25	1.30		

Ethnicity = 3 Black, not Hispanic

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
College graduate (1 = yes, graduate)	0.10	.18	0.28	1	.60	1.10	0.77	1.57
In Academic School (1 = yes)	-0.29	.20	2.03	1	.15	0.75	0.51	1.11
Math SGP Category [low growth]*			1.80	2	.41			
Typical growth	-0.12	.20	0.35	1	.56	0.89	0.61	1.31
High growth	0.16	.19	0.67	1	.41	1.17	0.81	1.69
Constant	-0.03	.14	0.03	1	.85	0.97		

Ethnicity = 4 White, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.23	.06	16.82	1	.00	1.26	1.13	1.41
In Academic School (1 = yes)	-0.56	.07	63.99	1	.00	0.57	0.49	0.65
Math SGP Category [low growth]*			6.16	2	.05			
Typical growth	-0.01	.07	0.03	1	.86	0.99	0.86	1.13
High growth	0.14	.07	4.37	1	.04	1.15	1.01	1.30
Constant	0.39	.05	62.84	1	.00	1.48		

Ethnicity = 5 Hispanic or Latino

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.37	.16	5.26	1	.02	1.44	1.05	1.97
In Academic School (1 = yes)	-0.49	.16	9.86	1	.00	0.61	0.45	0.83
Math SGP Category [low growth]*			2.07	2	.36			
Typical growth	0.23	.16	2.06	1	.15	1.26	0.92	1.74
High growth	0.10	.15	0.39	1	.53	1.10	0.81	1.49
Constant	0.22	.12	3.45	1	.06	1.24		

Logistic Regression by Ethnicity - Reading Performance Growth

Ethnicity = 2 Asian or Pacific Islander

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.32	.21	2.30	1	.13	1.37	0.91	2.06
In Academic School (1 = yes)	-0.98	.22	20.43	1	.00	0.38	0.25	0.58
Reading SGP Category [low growth]*			0.87	2	.65			
Typical growth	-0.13	.28	0.21	1	.65	0.88	0.51	1.51
High growth	-0.23	.25	0.86	1	.36	0.79	0.49	1.29
Constant	0.33	.23	2.04	1	.15	1.39		

Ethnicity = 3 Black, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.16	.18	0.77	1	.38	1.17	0.82	1.68
In Academic School (1 = yes)	-0.32	.20	2.54	1	.11	0.73	0.49	1.08
Reading SGP Category [low growth]*			4.60	2	.10			
Typical growth	-0.15	.19	0.68	1	.41	0.86	0.59	1.24
High growth	-0.43	.20	4.59	1	.03	0.65	0.44	0.96
Constant	0.15	.14	1.26	1	.26	1.16		

Ethnicity = 4 White, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.26	.06	22.04	1	.00	1.30	1.17	1.45
In Academic School (1 = yes)	-0.56	.07	62.57	1	.00	0.57	0.50	0.66
Reading SGP Category [low growth]*			0.94	2	.62			
Typical growth	-0.03	.07	0.15	1	.70	0.97	0.85	1.11
High growth	-0.06	.06	0.93	1	.33	0.94	0.83	1.07
Constant	0.45	.05	78.52	1	.00	1.57		

Ethnicity = 5 Hispanic or Latino

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.38	.16	5.62	1	.02	1.46	1.07	2.00
In Academic School (1 = yes)	-0.51	.16	10.57	1	.00	0.60	0.44	0.82
Reading SGP Category [low growth]*			1.34	2	.51			
Typical growth	-0.15	.16	0.90	1	.34	0.86	0.63	1.17
High growth	-0.16	.16	1.07	1	.30	0.85	0.62	1.16
Constant	0.42	.12	12.83	1	.00	1.52		

Logistic Regression by Gender - Math Performance Growth

Gender = 1 Male

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.30	.07	18.21	1	.00	1.35	1.18	1.56
In Academic School (1 = yes)	-0.64	.08	58.28	1	.00	0.53	0.45	0.62
Math SGP Category [low growth]*			1.38	2	.50			
Typical growth	-0.10	.09	1.37	1	.24	0.90	0.76	1.07
High growth	-0.05	.08	0.42	1	.52	0.95	0.81	1.11
Constant	0.48	.06	58.18	1	.00	1.62		

Gender = 2 Female

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.28	.06	19.54	1	.00	1.33	1.17	1.51
In Academic School (1 = yes)	-0.50	.08	40.86	1	.00	0.61	0.52	0.71
Math SGP Category [low growth]*			3.46	2	.18			
Typical growth	0.05	.08	0.42	1	.52	1.05	0.90	1.22
High growth	0.14	.07	3.42	1	.06	1.15	0.99	1.32
Constant	0.15	.05	7.97	1	.00	1.16		

Logistic Regression by Gender - Reading Performance Growth

Gender = 1 Male

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.31	.07	19.66	1	.00	1.37	1.19	1.57
In Academic School (1 = yes)	-0.65	.08	60.80	1	.00	0.52	0.44	0.61
Reading SGP Category [low growth]*			2.41	2	.30			
Typical growth	-0.13	.08	2.33	1	.13	0.88	0.75	1.04
High growth	-0.03	.08	0.19	1	.66	0.97	0.83	1.13
Constant	0.48	.06	63.36	1	.00	1.61		

Gender = 2 Female

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.32	.06	24.96	1	.00	1.38	1.21	1.56
In Academic School (1 = yes)	-0.50	.08	40.48	1	.00	0.61	0.52	0.71
Reading SGP Category [low growth]*			7.15	2	.03			
Typical growth	0.03	.08	0.14	1	.71	1.03	0.89	1.20
High growth	-0.16	.07	4.39	1	.04	0.86	0.74	0.99
Constant	0.24	.06	17.55	1	.00	1.28		

SPSS Output for Voting Habits

Logistic Regression - Math Performance Growth

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
Age of respondent [quarter 1: youngest]			17.71	3	.00			
Quarter 2	0.00	.06	0.01	1	.94	1.00	0.89	1.12
Quarter 3	0.00	.06	0.00	1	.99	1.00	0.89	1.13
Quarter 4 [oldest]	-0.26	.07	13.49	1	.00	0.77	0.67	0.89
Socio-economic status (1988) [quarter 1: low]*			39.14	3	.00			
Quarter 2	0.34	.07	25.18	1	.00	1.41	1.23	1.61
Quarter 3	0.36	.07	27.19	1	.00	1.43	1.25	1.64
Quarter 4 [high]	0.40	.07	30.59	1	.00	1.49	1.29	1.72
College graduate (1 = yes, graduate)	0.25	.05	22.99	1	.00	1.29	1.16	1.43
Math SGP Category [low growth]*			0.40	2	.82			
Typical growth	-0.04	.06	0.39	1	.53	0.96	0.86	1.08
High growth	-0.01	.05	0.05	1	.83	0.99	0.89	1.10
Constant	-0.55	.07	60.34	1	.00	0.58		

Logistic Regression - Reading Performance Growth

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
Age of respondent [quarter 1: youngest]			18.12	3	.00			
Quarter 2	-0.01	.06	0.01	1	.92	0.99	0.88	1.12
Quarter 3	0.00	.06	0.00	1	.95	1.00	0.88	1.13
Quarter 4 [oldest]	-0.26	.07	14.00	1	.00	0.77	0.67	0.88
Socio-economic status (1988) [quarter 1: low]*			38.34	3	.00			
Quarter 2	0.34	.07	24.98	1	.00	1.41	1.23	1.61
Quarter 3	0.36	.07	26.93	1	.00	1.43	1.25	1.63
Quarter 4 [high]	0.39	.07	29.50	1	.00	1.48	1.28	1.71
College graduate (1 = yes, graduate)	0.26	.05	23.98	1	.00	1.29	1.17	1.43
Reading SGP Category [low growth]*			0.55	2	.76			
Typical growth	-0.01	.06	0.01	1	.91	0.99	0.89	1.11
High growth	-0.04	.05	0.48	1	.49	0.96	0.87	1.07
Constant	-0.55	.07	59.75	1	.00	0.58		

Logistic Regression by Ethnicity - Math Performance Growth

Ethnicity = 2 Asian or Pacific Islander

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
Age of respondent [quarter 1: youngest]			0.32	3	.96			
Quarter 2	-0.08	.27	0.09	1	.76	0.92	0.54	1.57
Quarter 3	-0.12	.30	0.17	1	.68	0.88	0.49	1.60
Quarter 4 [oldest]	-0.16	.32	0.26	1	.61	0.85	0.45	1.59
Socio-economic status (1988) [quarter 1: low]*			1.95	3	.58			
Quarter 2	0.26	.38	0.47	1	.49	1.29	0.62	2.70
Quarter 3	0.20	.37	0.29	1	.59	1.22	0.59	2.53
Quarter 4 [high]	0.44	.33	1.71	1	.19	1.55	0.80	2.98
College graduate (1 = yes, graduate)	0.15	.23	0.42	1	.52	1.16	0.74	1.83
Math SGP Category [low growth]*			2.48	2	.29			
Typical growth	0.10	.30	0.11	1	.74	1.10	0.61	1.99
High growth	-0.28	.26	1.17	1	.28	0.75	0.45	1.26
Constant	-0.96	.37	6.73	1	.01	0.38		

Ethnicity = 3 Black, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
Age of respondent [quarter 1: youngest]			10.05	3	.02			

	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Quarter 2	0.07	.23	0.09	1	.76	1.07	0.69	1.67
Quarter 3	0.22	.24	0.91	1	.34	1.25	0.79	1.99
Quarter 4 [oldest]	-0.53	.24	4.74	1	.03	0.59	0.37	0.95
Socio-economic status (1988) [quarter 1: low]*			7.12	3	.07			
Quarter 2	0.23	.21	1.23	1	.27	1.26	0.84	1.90
Quarter 3	0.18	.24	0.57	1	.45	1.19	0.75	1.89
Quarter 4 [high]	0.73	.28	7.00	1	.01	2.08	1.21	3.58
College graduate (1 = yes, graduate)	0.62	.21	9.11	1	.00	1.86	1.24	2.78
Math SGP Category [low growth]*			4.97	2	.08			
Typical growth	-0.46	.21	4.97	1	.03	0.63	0.42	0.95
High growth	-0.18	.20	0.86	1	.35	0.83	0.56	1.23
Constant	-0.11	.23	0.23	1	.63	0.89		

Ethnicity = 4 White, not Hispanic

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Age of respondent [quarter 1: youngest]			3.43	3	.33			
Quarter 2	0.02	.07	0.05	1	.83	1.02	0.88	1.17
Quarter 3	-0.01	.07	0.01	1	.91	0.99	0.86	1.15
Quarter 4 [oldest]	-0.13	.08	2.41	1	.12	0.88	0.74	1.03
Socio-economic status (1988) [quarter 1: low]*			23.91	3	.00			
Quarter 2	0.35	.09	16.29	1	.00	1.43	1.20	1.69
Quarter 3	0.37	.09	18.04	1	.00	1.45	1.22	1.72
Quarter 4 [high]	0.41	.09	19.90	1	.00	1.50	1.26	1.80
College graduate (1 = yes, graduate)	0.29	.06	21.24	1	.00	1.33	1.18	1.50
Math SGP Category [low growth]*			0.56	2	.76			
Typical growth	0.01	.07	0.03	1	.86	1.01	0.88	1.16
High growth	0.05	.06	0.52	1	.47	1.05	0.92	1.19
Constant	-0.62	.09	46.80	1	.00	0.54		

Ethnicity = 5 Hispanic or Latino

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Age of respondent [quarter 1: youngest]			9.34	3	.03			
Quarter 2	-0.27	.18	2.19	1	.14	0.77	0.54	1.09
Quarter 3	-0.05	.18	0.07	1	.80	0.95	0.66	1.37
Quarter 4 [oldest]	-0.55	.20	7.73	1	.01	0.58	0.39	0.85
Socio-economic status (1988) [quarter 1: low]*			8.20	3	.04			
Quarter 2	0.37	.17	4.55	1	.03	1.45	1.03	2.03
Quarter 3	0.46	.19	5.93	1	.01	1.59	1.10	2.31
Quarter 4 [high]	0.33	.22	2.35	1	.12	1.39	0.91	2.13
College graduate (1 = yes, graduate)	0.39	.17	5.21	1	.02	1.47	1.06	2.05
Math SGP Category [low growth]*			1.20	2	.55			
Typical growth	-0.11	.17	0.41	1	.52	0.90	0.64	1.25
High growth	0.07	.16	0.21	1	.65	1.08	0.79	1.47
Constant	-0.51	.18	8.37	1	.00	0.60		

Logistic Regression by Ethnicity - Reading Performance Growth

Ethnicity = 2 Asian or Pacific Islander

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Age of respondent [quarter 1: youngest]			0.51	3	.92			
Quarter 2	-0.12	.27	0.21	1	.65	0.88	0.52	1.51
Quarter 3	-0.19	.30	0.41	1	.52	0.82	0.45	1.50
Quarter 4 [oldest]	-0.15	.32	0.23	1	.63	0.86	0.46	1.61
Socio-economic status (1988) [quarter 1: low]*			2.43	3	.49			
Quarter 2	0.38	.38	0.99	1	.32	1.46	0.69	3.07
Quarter 3	0.28	.37	0.55	1	.46	1.32	0.63	2.75
Quarter 4 [high]	0.51	.34	2.24	1	.13	1.66	0.85	3.24
College graduate (1 = yes, graduate)	0.12	.24	0.27	1	.60	1.13	0.71	1.79
Reading SGP Category [low growth]*			2.56	2	.28			
Typical growth	0.40	.30	1.88	1	.17	1.50	0.84	2.67
High growth	0.05	.27	0.03	1	.86	1.05	0.62	1.78

	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Constant	-1.26	.38	11.09	1	.00	0.28		

Ethnicity = 3 Black, not Hispanic

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Age of respondent [quarter 1: youngest]			8.89	3	.03			
Quarter 2	0.03	.22	0.02	1	.88	1.03	0.67	1.61
Quarter 3	0.23	.23	0.92	1	.34	1.25	0.79	1.99
Quarter 4 [oldest]	-0.49	.24	4.15	1	.04	0.61	0.38	0.98
Socio-economic status (1988) [quarter 1: low]*			7.97	3	.05			
Quarter 2	0.27	.21	1.63	1	.20	1.31	0.87	1.97
Quarter 3	0.23	.24	0.93	1	.34	1.25	0.79	1.99
Quarter 4 [high]	0.77	.28	7.83	1	.01	2.17	1.26	3.73
College graduate (1 = yes, graduate)	0.63	.21	9.39	1	.00	1.87	1.25	2.80
Reading SGP Category [low growth]*			1.22	2	.54			
Typical growth	-0.19	.20	0.93	1	.33	0.83	0.56	1.22
High growth	-0.19	.21	0.80	1	.37	0.83	0.55	1.25
Constant	-0.21	.22	0.93	1	.34	0.81		

Ethnicity = 4 White, not Hispanic

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Age of respondent [quarter 1: youngest]			3.78	3	.29			
Quarter 2	0.01	.07	0.03	1	.86	1.01	0.88	1.16
Quarter 3	-0.01	.07	0.03	1	.86	0.99	0.85	1.14
Quarter 4 [oldest]	-0.14	.08	2.76	1	.10	0.87	0.74	1.03
Socio-economic status (1988) [quarter 1: low]*			23.57	3	.00			
Quarter 2	0.36	.09	16.52	1	.00	1.43	1.20	1.70
Quarter 3	0.37	.09	17.86	1	.00	1.44	1.22	1.71
Quarter 4 [high]	0.40	.09	19.20	1	.00	1.49	1.25	1.78
College graduate (1 = yes, graduate)	0.29	.06	22.83	1	.00	1.34	1.19	1.51
Reading SGP Category [low growth]*			0.12	2	.94			
Typical growth	0.00	.07	0.00	1	.96	1.00	0.87	1.14
High growth	0.02	.06	0.07	1	.78	1.02	0.90	1.15
Constant	-0.60	.09	44.41	1	.00	0.55		

Ethnicity = 5 Hispanic or Latino

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Age of respondent [quarter 1: youngest]			9.30	3	.03			
Quarter 2	-0.26	.18	2.02	1	.15	0.77	0.54	1.10
Quarter 3	-0.05	.19	0.08	1	.78	0.95	0.66	1.37
Quarter 4 [oldest]	-0.55	.20	7.81	1	.01	0.57	0.39	0.85
Socio-economic status (1988) [quarter 1: low]*			8.09	3	.04			
Quarter 2	0.34	.17	3.89	1	.05	1.40	1.00	1.97
Quarter 3	0.48	.19	6.20	1	.01	1.61	1.11	2.34
Quarter 4 [high]	0.34	.22	2.48	1	.12	1.41	0.92	2.15
College graduate (1 = yes, graduate)	0.39	.17	5.25	1	.02	1.47	1.06	2.05
Reading SGP Category [low growth]*			0.59	2	.74			
Typical growth	0.12	.16	0.53	1	.47	1.13	0.82	1.55
High growth	0.02	.16	0.02	1	.90	1.02	0.74	1.41
Constant	-0.56	.18	9.65	1	.00	0.57		

Logistic Regression by Gender - Math Performance Growth

Gender = 1 Male

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Age of respondent [quarter 1: youngest]			13.02	3	.00			
Quarter 2	0.08	.09	0.84	1	.36	1.09	0.91	1.30

	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Quarter 3	0.08	.09	0.81	1	.37	1.09	0.91	1.31
Quarter 4 [oldest]	-0.23	.10	5.28	1	.02	0.80	0.66	0.97
Socio-economic status (1988) [quarter 1: low]*			15.73	3	.00			
Quarter 2	0.27	.10	6.96	1	.01	1.31	1.07	1.60
Quarter 3	0.33	.10	10.24	1	.00	1.39	1.13	1.69
Quarter 4 [high]	0.40	.11	14.42	1	.00	1.50	1.22	1.85
College graduate (1 = yes, graduate)	0.19	.08	5.58	1	.02	1.20	1.03	1.40
Math SGP Category [low growth]*			0.53	2	.77			
Typical growth	0.02	.09	0.06	1	.81	1.02	0.86	1.21
High growth	0.06	.08	0.50	1	.48	1.06	0.90	1.24
Constant	-0.61	.11	31.78	1	.00	0.54		

Gender = 2 Female

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Age of respondent [quarter 1: youngest]			6.26	3	.10			
Quarter 2	-0.07	.08	0.73	1	.39	0.93	0.80	1.09
Quarter 3	-0.06	.08	0.54	1	.46	0.94	0.80	1.11
Quarter 4 [oldest]	-0.25	.10	6.22	1	.01	0.78	0.64	0.95
Socio-economic status (1988) [quarter 1: low]*			24.61	3	.00			
Quarter 2	0.40	.09	18.75	1	.00	1.49	1.24	1.78
Quarter 3	0.38	.09	16.62	1	.00	1.46	1.22	1.76
Quarter 4 [high]	0.40	.10	16.19	1	.00	1.49	1.23	1.81
College graduate (1 = yes, graduate)	0.30	.07	17.36	1	.00	1.34	1.17	1.55
Math SGP Category [low growth]*			1.05	2	.59			
Typical growth	-0.07	.08	0.82	1	.37	0.93	0.80	1.09
High growth	-0.06	.08	0.71	1	.40	0.94	0.81	1.09
Constant	-0.51	.09	29.95	1	.00	0.60		

Logistic Regression by Gender - Reading Performance Growth

Gender = 1 Male

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Age of respondent [quarter 1: youngest]			12.96	3	.00			
Quarter 2	0.08	.09	0.79	1	.37	1.09	0.91	1.30
Quarter 3	0.08	.09	0.83	1	.36	1.09	0.91	1.31
Quarter 4 [oldest]	-0.23	.10	5.28	1	.02	0.80	0.66	0.97
Socio-economic status (1988) [quarter 1: low]*			15.02	3	.00			
Quarter 2	0.27	.10	6.88	1	.01	1.31	1.07	1.60
Quarter 3	0.32	.10	9.93	1	.00	1.38	1.13	1.68
Quarter 4 [high]	0.39	.11	13.64	1	.00	1.48	1.20	1.83
College graduate (1 = yes, graduate)	0.19	.08	5.77	1	.02	1.21	1.04	1.41
Reading SGP Category [low growth]*			0.67	2	.72			
Typical growth	0.06	.08	0.49	1	.49	1.06	0.90	1.25
High growth	0.05	.08	0.49	1	.48	1.06	0.91	1.23
Constant	-0.62	.11	33.65	1	.00	0.54		

Gender = 2 Female

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
Age of respondent [quarter 1: youngest]			6.90	3	.08			
Quarter 2	-0.07	.08	0.73	1	.39	0.93	0.80	1.09
Quarter 3	-0.07	.08	0.76	1	.38	0.93	0.79	1.10
Quarter 4 [oldest]	-0.27	.10	6.87	1	.01	0.77	0.63	0.94
Socio-economic status (1988) [quarter 1: low]*			25.16	3	.00			
Quarter 2	0.40	.09	19.19	1	.00	1.49	1.25	1.79
Quarter 3	0.39	.09	17.11	1	.00	1.47	1.23	1.77
Quarter 4 [high]	0.40	.10	16.43	1	.00	1.49	1.23	1.81
College graduate (1 = yes, graduate)	0.30	.07	18.30	1	.00	1.35	1.18	1.55
Reading SGP Category [low growth]*			2.95	2	.23			
Typical growth	-0.07	.08	0.83	1	.36	0.93	0.80	1.09

High growth	-0.13	.08	2.95	1	.09	0.88	0.76	1.02
Constant	-0.48	.09	26.21	1	.00	0.62		

SPSS Output for Social Involvement

Logistic Regression - Math Performance Growth

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
College graduate (1 = yes, graduate)	0.31	.05	42.25	1	.00	1.37	1.24	1.50
Named MVP in 1992 (1 = named MVP)	0.66	.08	70.07	1	.00	1.93	1.65	2.25
Math SGP Category [low growth]*			5.53	2	.06			
Typical growth	0.14	.06	5.51	1	.02	1.15	1.02	1.29
High growth	0.07	.06	1.58	1	.21	1.07	0.96	1.19
Constant	-0.11	.04	6.94	1	.01	0.89		

Logistic Regression - Reading Performance Growth

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
College graduate (1 = yes, graduate)	0.32	.05	46.23	1	.00	1.38	1.26	1.52
Named MVP in 1992 (1 = named MVP)	0.66	.08	69.60	1	.00	1.93	1.65	2.25
Reading SGP Category [low growth]*			1.23	2	.54			
Typical growth	0.04	.06	0.51	1	.48	1.04	0.93	1.17
High growth	-0.02	.06	0.15	1	.69	0.98	0.88	1.09
Constant	-0.06	.04	1.75	1	.19	0.94		

Logistic Regression by Ethnicity - Math Performance Growth

Ethnicity = 2 Asian or Pacific Islander

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
College graduate (1 = yes, graduate)	0.08	.21	0.15	1	.70	1.08	0.72	1.63
Named MVP in 1992 (1 = named MVP)	-0.07	.36	0.04	1	.84	0.93	0.46	1.87
Math SGP Category [low growth]*			0.79	2	.67			
Typical growth	0.13	.29	0.20	1	.65	1.14	0.65	2.00
High growth	0.22	.25	0.79	1	.37	1.24	0.77	2.01
Constant	-0.28	.24	1.39	1	.24	0.76		

Ethnicity = 3 Black, not Hispanic

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
College graduate (1 = yes, graduate)	0.35	.20	2.98	1	.08	1.42	0.95	2.13
Named MVP in 1992 (1 = named MVP)	0.46	.30	2.30	1	.13	1.58	0.87	2.87
Math SGP Category [low growth]*			1.04	2	.60			
Typical growth	0.09	.22	0.16	1	.69	1.09	0.71	1.67
High growth	0.22	.21	1.03	1	.31	1.24	0.82	1.88
Constant	0.48	.15	10.13	1	.00	1.61		

Ethnicity = 4 White, not Hispanic

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
College graduate (1 = yes, graduate)	0.35	.06	38.64	1	.00	1.42	1.27	1.59
Named MVP in 1992 (1 = named MVP)	0.68	.09	53.51	1	.00	1.97	1.64	2.37
Math SGP Category [low growth]*			3.91	2	.14			
Typical growth	0.14	.07	3.85	1	.05	1.15	1.00	1.32
High growth	0.05	.07	0.55	1	.46	1.05	0.92	1.20
Constant	-0.14	.05	7.74	1	.01	0.87		

Ethnicity = 5 Hispanic or Latino

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.36	.16	4.75	1	.03	1.43	1.04	1.96
Named MVP in 1992 (1 = named MVP)	1.07	.26	16.96	1	.00	2.92	1.75	4.85
Math SGP Category [low growth]*			2.90	2	.24			
Typical growth	0.27	.17	2.42	1	.12	1.31	0.93	1.83
High growth	0.22	.16	1.87	1	.17	1.25	0.91	1.72
Constant	-0.12	.12	0.99	1	.32	0.88		

Logistic Regression by Ethnicity - Reading Performance Growth

Ethnicity = 2 Asian or Pacific Islander

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.10	.21	0.22	1	.64	1.10	0.73	1.67
Named MVP in 1992 (1 = named MVP)	-0.09	.36	0.06	1	.80	0.91	0.45	1.84
Reading SGP Category [low growth]*			2.16	2	.34			
Typical growth	0.33	.28	1.42	1	.23	1.39	0.81	2.40
High growth	0.01	.25	0.00	1	.97	1.01	0.61	1.65
Constant	-0.25	.24	1.08	1	.30	0.78		

Ethnicity = 3 Black, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.39	.21	3.58	1	.06	1.48	0.99	2.21
Named MVP in 1992 (1 = named MVP)	0.44	.31	2.12	1	.15	1.56	0.86	2.83
Reading SGP Category [low growth]*			1.26	2	.53			
Typical growth	-0.14	.21	0.45	1	.50	0.87	0.58	1.31
High growth	0.11	.23	0.26	1	.61	1.12	0.72	1.74
Constant	0.59	.15	14.84	1	.00	1.80		

Ethnicity = 4 White, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.36	.06	41.04	1	.00	1.43	1.28	1.60
Named MVP in 1992 (1 = named MVP)	0.69	.09	54.01	1	.00	1.98	1.65	2.38
Reading SGP Category [low growth]*			0.51	2	.78			
Typical growth	0.05	.07	0.50	1	.48	1.05	0.92	1.21
High growth	0.02	.07	0.09	1	.76	1.02	0.90	1.16
Constant	-0.11	.05	4.40	1	.04	0.90		

Ethnicity = 5 Hispanic or Latino

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.35	.16	4.61	1	.03	1.42	1.03	1.95
Named MVP in 1992 (1 = named MVP)	1.06	.26	16.58	1	.00	2.88	1.73	4.79
Reading SGP Category [low growth]*			0.89	2	.64			
Typical growth	-0.07	.17	0.18	1	.67	0.93	0.67	1.29
High growth	-0.16	.17	0.89	1	.35	0.86	0.62	1.18
Constant	0.11	.12	0.80	1	.37	1.12		

Logistic Regression by Gender - Math Performance Growth

Gender = 1 Male

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	0.38	.07	27.42	1	.00	1.46	1.26	1.68
Named MVP in 1992 (1 = named MVP)	0.60	.10	36.43	1	.00	1.81	1.50	2.20
Math SGP Category [low growth]*			4.99	2	.08			
Typical growth	0.18	.09	3.85	1	.05	1.19	1.00	1.42
High growth	0.01	.08	0.01	1	.90	1.01	0.86	1.18
Constant	0.01	.06	0.04	1	.84	1.01		

Gender = 2 Female

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	0.30	.07	20.78	1	.00	1.35	1.18	1.53
Named MVP in 1992 (1 = named MVP)	0.61	.13	21.12	1	.00	1.84	1.42	2.38
Math SGP Category [low growth]*			1.45	2	.49			
Typical growth	0.08	.08	1.11	1	.29	1.09	0.93	1.27
High growth	0.08	.08	1.00	1	.32	1.08	0.93	1.25
Constant	-0.20	.06	12.60	1	.00	0.82		

Logistic Regression by Gender- Reading Performance Growth

Gender = 1 Male

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	0.38	.07	28.78	1	.00	1.47	1.27	1.69
Named MVP in 1992 (1 = named MVP)	0.60	.10	37.24	1	.00	1.83	1.51	2.22
Reading SGP Category [low growth]*			0.03	2	.98			
Typical growth	0.01	.09	0.03	1	.87	1.01	0.86	1.20
High growth	0.00	.08	0.00	1	1.00	1.00	0.86	1.17
Constant	0.06	.06	0.84	1	.36	1.06		

Gender = 2 Female

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	0.31	.06	22.95	1	.00	1.36	1.20	1.55
Named MVP in 1992 (1 = named MVP)	0.59	.13	20.01	1	.00	1.81	1.40	2.35
Reading SGP Category [low growth]*			2.47	2	.29			
Typical growth	0.09	.08	1.25	1	.26	1.09	0.94	1.28
High growth	-0.03	.08	0.14	1	.71	0.97	0.84	1.13
Constant	-0.17	.06	8.03	1	.00	0.84		

SPSS Output for Informational Involvement

Logistic Regression - Math Performance Growth

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR	
							Lower	Upper
College graduate (1 = yes, graduate)	0.21	.05	17.80	1	.00	1.24	1.12	1.37
In Academic School in 2000	0.63	.06	109.30	1	.00	1.87	1.66	2.11
Ever in AP Program	0.19	.05	14.15	1	.00	1.21	1.10	1.33
Math SGP Category [low growth]*			0.04	2	.98			
Typical growth	0.00	.06	0.00	1	.98	1.00	0.89	1.13
High growth	-0.01	.06	0.02	1	.88	0.99	0.89	1.11
Constant	-0.28	.04	39.79	1	.00	0.75		

Logistic Regression - Reading Performance Growth

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.22	.05	18.37	1	.00	1.24	1.12	1.37
In Academic School in 2000	0.64	.06	112.67	1	.00	1.89	1.68	2.12
Ever in AP Program	0.19	.05	13.66	1	.00	1.21	1.09	1.33
Reading SGP Category [low growth]*			0.37	2	.83			
Typical growth	-0.02	.06	0.16	1	.69	0.98	0.87	1.10
High growth	-0.03	.06	0.35	1	.55	0.97	0.87	1.08
Constant	-0.27	.05	34.37	1	.00	0.77		

Logistic Regression by Ethnicity - Math Performance Growth

Ethnicity = 2 Asian or Pacific Islander

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	-0.30	.22	1.81	1	.18	0.74	0.48	1.15
In Academic School in 2000	0.74	.23	10.52	1	.00	2.09	1.34	3.27
Ever in AP Program	0.11	.23	0.25	1	.61	1.12	0.72	1.75
Math SGP Category [low growth]*			0.15	2	.93			
Typical growth	-0.11	.29	0.15	1	.70	0.89	0.51	1.58
High growth	-0.05	.25	0.04	1	.85	0.95	0.58	1.56
Constant	0.19	.25	0.56	1	.45	1.21		

Ethnicity = 3 Black, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.41	.20	4.31	1	.04	1.51	1.02	2.23
In Academic School in 2000	0.83	.22	13.97	1	.00	2.29	1.48	3.54
Ever in AP Program	0.17	.18	0.87	1	.35	1.19	0.83	1.70
Math SGP Category [low growth]*			7.35	2	.03			
Typical growth	-0.56	.21	6.89	1	.01	0.57	0.38	0.87
High growth	-0.37	.21	3.32	1	.07	0.69	0.46	1.03
Constant	0.13	.16	0.68	1	.41	1.14		

Ethnicity = 4 White, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.28	.06	21.91	1	.00	1.32	1.18	1.49
In Academic School in 2000	0.58	.07	62.53	1	.00	1.78	1.54	2.05
Ever in AP Program	0.17	.06	8.48	1	.00	1.19	1.06	1.34
Math SGP Category [low growth]*			1.12	2	.57			
Typical growth	0.05	.07	0.44	1	.51	1.05	0.91	1.20
High growth	-0.03	.07	0.15	1	.70	0.97	0.86	1.11
Constant	-0.35	.05	43.17	1	.00	0.71		

Ethnicity = 5 Hispanic or Latino

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.01	.17	0.00	1	.97	1.01	0.73	1.39
In Academic School in 2000	0.70	.17	17.89	1	.00	2.01	1.46	2.78
Ever in AP Program	0.18	.15	1.52	1	.22	1.20	0.90	1.59
Math SGP Category [low growth]*			3.10	2	.21			
Typical growth	0.03	.17	0.03	1	.86	1.03	0.74	1.44
High growth	0.26	.16	2.59	1	.11	1.30	0.94	1.79
Constant	-0.34	.13	6.57	1	.01	0.72		

Logistic Regression by Ethnicity - Reading Performance Growth

Ethnicity = 2 Asian or Pacific Islander

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	-0.30	.23	1.79	1	.18	0.74	0.48	1.15
In Academic School in 2000	0.73	.22	10.84	1	.00	2.08	1.35	3.23
Ever in AP Program	0.10	.22	0.21	1	.64	1.11	0.71	1.72
Reading SGP Category [low growth]*			0.34	2	.84			
Typical growth	0.16	.28	0.33	1	.56	1.18	0.68	2.04
High growth	0.07	.25	0.07	1	.79	1.07	0.65	1.75
Constant	0.07	.25	0.08	1	.78	1.07		

Ethnicity = 3 Black, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.38	.20	3.67	1	.06	1.46	0.99	2.15
In Academic School in 2000	0.84	.22	14.59	1	.00	2.33	1.51	3.59
Ever in AP Program	0.12	.18	0.43	1	.51	1.13	0.79	1.62
Reading SGP Category [low growth]*			2.67	2	.26			
Typical growth	-0.28	.20	1.96	1	.16	0.75	0.51	1.12
High growth	0.03	.21	0.02	1	.89	1.03	0.68	1.56
Constant	-0.04	.15	0.07	1	.80	0.96		

Ethnicity = 4 White, not Hispanic

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.28	.06	21.69	1	.00	1.32	1.17	1.48
In Academic School in 2000	0.58	.07	63.71	1	.00	1.79	1.55	2.07
Ever in AP Program	0.17	.06	8.11	1	.00	1.19	1.05	1.34
Reading SGP Category [low growth]*			0.38	2	.83			
Typical growth	0.03	.07	0.20	1	.66	1.03	0.90	1.18
High growth	0.04	.07	0.35	1	.55	1.04	0.91	1.18
Constant	-0.36	.05	45.77	1	.00	0.70		

Ethnicity = 5 Hispanic or Latino

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.01	.17	0.00	1	.96	1.01	0.73	1.40
In Academic School in 2000	0.73	.17	19.19	1	.00	2.07	1.49	2.87
Ever in AP Program	0.19	.15	1.65	1	.20	1.21	0.91	1.60
Reading SGP Category [low growth]*			1.85	2	.40			
Typical growth	-0.05	.17	0.08	1	.78	0.95	0.69	1.32
High growth	-0.22	.17	1.70	1	.19	0.81	0.58	1.11
Constant	-0.15	.13	1.25	1	.26	0.86		

Logistic Regression by Gender - Math Performance Growth

Gender = 1 Male

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.35	.07	21.54	1	.00	1.42	1.22	1.64
In Academic School in 2000	0.76	.09	72.91	1	.00	2.14	1.80	2.55
Ever in AP Program	0.31	.07	17.80	1	.00	1.37	1.18	1.58
Math SGP Category [low growth]*			1.47	2	.48			
Typical growth	0.11	.09	1.45	1	.23	1.11	0.93	1.33
High growth	0.04	.08	0.28	1	.60	1.04	0.89	1.23
Constant	-0.46	.07	44.36	1	.00	0.63		

Gender = 2 Female

	B	S.E.	Wald	df	Sig.	OR	95% C.I.for OR Lower	Upper
College graduate (1 = yes, graduate)	0.11	.07	2.42	1	.12	1.11	0.97	1.28

	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
In Academic School in 2000	0.52	.08	40.98	1	.00	1.68	1.44	1.98
Ever in AP Program	0.08	.07	1.49	1	.22	1.09	0.95	1.25
Math SGP Category [low growth]*			0.88	2	.64			
Typical growth	-0.07	.08	0.85	1	.36	0.93	0.79	1.09
High growth	-0.04	.08	0.33	1	.56	0.96	0.82	1.11
Constant	-0.14	.06	5.49	1	.02	0.87		

Logistic Regression by Gender - Reading Performance Growth

Gender = 1 Male

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
College graduate (1 = yes, graduate)	0.36	.07	22.91	1	.00	1.43	1.23	1.65
In Academic School in 2000	0.77	.09	75.35	1	.00	2.17	1.82	2.58
Ever in AP Program	0.30	.07	16.88	1	.00	1.35	1.17	1.57
Reading SGP Category [low growth]*			0.52	2	.77			
Typical growth	-0.05	.09	0.34	1	.56	0.95	0.80	1.13
High growth	-0.05	.08	0.41	1	.52	0.95	0.81	1.11
Constant	-0.37	.06	33.69	1	.00	0.69		

Gender = 2 Female

							95% C.I.for OR	
	B	S.E.	Wald	df	Sig.	OR	Lower	Upper
College graduate (1 = yes, graduate)	0.10	.07	2.30	1	.13	1.11	0.97	1.27
In Academic School in 2000	0.53	.08	41.94	1	.00	1.69	1.44	1.99
Ever in AP Program	0.09	.07	1.52	1	.22	1.09	0.95	1.25
Reading SGP Category [low growth]*			0.02	2	.99			
Typical growth	-0.01	.08	0.01	1	.93	0.99	0.85	1.16
High growth	-0.01	.08	0.02	1	.88	0.99	0.85	1.15
Constant	-0.17	.06	6.99	1	.01	0.84		