Spring 5-13-2022

The Information Content of High School Courses: How Early Exposure to STEM Impacts Future Decisions

Phillip E. Gustafson
Georgia State University

ABSTRACT

THE INFORMATION CONTENT OF HIGH SCHOOL COURSES: HOW EARLY EXPOSURE TO STEM IMPACTS FUTURE DECISIONS

By

PHILLIP EDWARD GUSTAFSON

May, 2022

Committee Chair: Dr. Thomas Mroz

Major Department: Economics

Increasing participation in STEM is a major policy goal, yet not much is known about how exposure to math and science in high school influences future STEM attainment. To examine this connection, I develop a dynamic structural model of high school curriculum and college major choice using data from the Educational Longitudinal Study of 2002. I examine the effectiveness of high school STEM exposure on resolving student uncertainty about academic aptitude and derive the impact of raising graduation requirements in math and science on educational outcomes and student welfare. I find that, while students are initially uncertain about their abilities at the beginning of high school, much of this is resolved upon completion. Therefore, counterfactual simulations that increase the amount of information contained in high school STEM courses do little to affect students' ultimate educational choices. Raising graduation requirements in math and science primarily boosts post-secondary attainment in college non-STEM for median individuals. I also find suggestive evidence that higher requirements in mathematics increase high school dropout rates for men. These effects may vary across the ability distribution, with effects on dropout concentrated among lower ability individuals. Because uncertainty is resolved quickly in high school, these effects are driven by changes in course-taking patterns rather than changes in beliefs. Ultimately, my findings do not rule out only modest negative welfare effects of these policies, with (imprecise) point estimates of the reduction in expected lifetime utility from increasing graduation requirements in mathematics from 2 to 4 years to be around 4.8% for men and 4.0% for women.

The Information Content of High School Courses: How Early Exposure to STEM Impacts Future Decisions

By

Phillip Edward Gustafson

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree
of
Doctor of Philosophy
in the
Andrew Young School of Policy Studies
of
Georgia State University

GEORGIA STATE UNIVERSITY

2022

ACCEPTANCE

This dissertation was prepared under the direction of the candidate's Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Andrew Young School of Policy Studies of Georgia State University.

Dissertation Chair:	Dr. Thomas Mroz
Committee:	Dr. Daniel Kreisman
	Dr. Melinda Pitts
	Dr. Jonathan Smith

Electronic Version Approved:

Dr. Sally Wallace, Dean
Andrew Young School of Policy Studies
Georgia State University
May, 2022

ACKNOWLEDGMENTS

I would like to express my extreme gratitude to everyone who has supported me and my work throughout graduate school. First I must thank my wife, Dr. Jennifer Mai, for being by my side throughout my entire post-secondary education. From when we were undergraduate students together, to when we were both completing our graduate studies $1,000$ miles away from each other, you stuck with me the entire way and I would not be where I am without your love and support.

I also would like to thank my advisor, Dr. Thomas Mroz. It has been an absolute privilege to complete my degree under your supervision. Listening to your perspectives on research and how to think about economic and econometric ideas has been immensely transformative for me, and I would not have known to approach my research in the same way without your encouragement and advice.

I would be remiss to not thank the members of my committee: Dr. Daniel Kreisman, Dr. Melinda Pitts, and Dr. Jonathan Smith. Your unique perspectives played a key role in guiding my research and challenging me to make it the best that it can be.

I must also acknowledge the generous financial support I received from the Provost's Dissertation Fellowship, which greatly assisted in the completion of my graduate studies.

And finally, to my parents, Erik and Stacey: I would not be who I am, or where I am, today without the unconditional support I receive from you, and for this I am truly grateful.

Table of Contents

List of Tables

List of Figures

Introduction

In 2006, the state of Michigan implemented one of the most comprehensive high school curriculum reforms in the United States, the 'Michigan Merit Curriculum', with the intent of giving "students the learning opportunity, knowledge and skills they need to succeed in college or the workplace," (*Michigan Merit Curriculum High School Graduation Requirements* (2008)). To achieve these ends, the policy mandated that students complete a higher number of courses across a variety of subjects in order to receive a high school diploma.

Michigan has not been alone. Since the 1980s, American policymakers have frequently raised high school graduation requirements, with an especially prominent focus on getting students to take more Science, Technology, Engineering, and Mathematics (STEM) (Goodman (2019)). In 1998, only 13 states required students to complete at least 3 years of math and science to graduate, but by 2018 the number of states had gone up to 37 (NCES (1999),NCES (2018)). As a consequence, students today must take more of these courses than those only a generation before.

However, it is not entirely clear what we should expect the effects of such curriculum requirements to be. While taking extra courses in math and science should, through human capital accumulation, improve student success in college and the labor market, the act of forcing students to take courses that they otherwise would not have taken could have unintended consequences. For instance, they may place an undue burden on students who are on the margin of completing high school and induce them to drop out prematurely. Or, for those students who choose to adhere to stricter academic standards, their course-taking in other areas may be affected. These students may choose to substitute away from taking more courses in, say, physical science, if they are expected to take more years of mathematics before graduating. Even if students' other decisions are unaffected, exposure to more math and science could alter students' beliefs about their own academic aptitude which, depending on the positive or negative nature of this exposure, could have significant impacts on whether a student chooses to go to college and in which field they decide to major.

In this paper, I analyze the impact of STEM course-taking in high school and how it affects future educational decisions. I focus on understanding the mechanisms through which exposure to more STEM in high school affects collegiate attainment, and determining the effects of math and science graduation requirements on student welfare and the probability that a student decides to complete a bachelors degree in a STEM field. Effects on collegiate STEM attainment are particularly relevant to policy-makers in the United States. STEM degrees yield significantly higher wages than their non-STEM counterparts, and progress in science and technology is believed to have high societal value (Altonji, Blom, and Meghir (2012),Altonji, Arcidiacono, and Maurel (2015)). Since U.S. policymakers already have considerable control over high school curricula, if simply increasing graduation requirements in STEM is effective at improving post-secondary STEM attainment, this may be a more easily implemented and lower cost alternative to other proposals aimed at achieving the same outcomes [1].

A line of literature has investigated the impacts of course-taking in high school on future educational and labor market outcomes, typically finding small but positive impacts. Altonji (1995) provides one of the first looks at this in the economics literature. Using data from the National Longitudinal Survey of the High School Class of 1972 and an instrumental variables approach, he finds small positive effects on future wages and the probability of going to college from taking additional credits in a variety of high school subjects. Levine and Zimmerman (1995) use a similar approach to Altonji and a different data set. They find positive effects of of math course-taking on women's wages and probability of majoring in a technical field in college. Following up on these studies, Rose and Betts (2004) document heterogeneity in the impact of math course-taking in high school, finding, in particular, that more advanced math courses have a stronger effect on wages than less advanced ones.

Joensen and Nielsen (2009) provide more solid evidence than prior literature on the causal effect of math course-taking in high school on wages and higher education attendance. They utilize a novel natural experiment within a Danish high school that induced some students to take advanced

[1] One such proposal is to lower college tuition for students majoring in STEM. Such a policy has been implemented in New York (*NYS Science, Technology, Engineering and Mathematics (STEM) Incentive Program* (n.d.))

mathematics as part of their overall high school curriculum. They find that, years later, these students had 30% higher earnings than their peers who did not take advanced math, and that part of this effect is due to the additional mathematics further inducing them into higher education.

Fewer studies attempt to determine the effects of stricter graduation requirements directly. Lillard and DeCicca (2001) use aggregate data from across the United States to look at the association between higher course-taking requirements and graduation rates, and find evidence that graduation requirements induce high school dropout. Similarly, Jacob et al. (2017) look at the effect of these policies in the context of the state of Michigan using state longitudinal data and find some evidence that the 'Michigan Merit Curriculum' may have caused lower achieving students to not finish high school.

Another line of work has attempted to determine why some students go to college and choose to major in STEM fields. A common theme in this literature has been a recognition of the importance of students' uncertainty about their own academic prowess when making college-going decisions (Arcidiacono (2004), Stinebrickner and Stinebrickner (2012), Stinebrickner and Stinebrickner (2014), Kinsler and Pavan (2015)) . Stange (2012) finds that the resolution of uncertainty represents a large portion of the option value of college attendance, while Arcidiacono et al. (2016) estimate that many more students would finish college, and have better post-secondary matches, if they had full certainty about their academic abilities upon exiting high school.

This paper extends the literature on high school STEM courses in several ways. First, I determine the extent to which students are able to resolve academic uncertainty, before making college-going decisions, through their high school course-taking. I use approaches developed in James (2011) and Arcidiacono et al. (2016) to estimate a structural learning process and a quasi-structural choice model, using detailed high school transcript data from a nationally representative longitudinal study in the U.S., that allows me to determine the amount of information on unobserved academic aptitude that is contained in high school math, physical science, and life science courses. Unlike many previous models in the educational choice literature that have estimated learning processes using strong distributional assumptions, I estimate the model using a non-parametric approximation to

the unobserved ability distributions. I then use this model to simulate raising and lowering the precision of ability signals in these courses to determine their effects on choices. I find that students resolve much of their uncertainty by the end of high school, and that relatively large alterations to the informational content of high school courses does little to change the median student's ultimate educational outcomes.

Second, I estimate the impact of raising graduation requirements in math and science, and compare these 'policy relevant treatment effects' to the 'average treatment effect' of simply raising the number of courses students take in these subject areas, while holding everything else constant[2]. Because I estimate a learning process along with a dynamic choice model I am able to control for students' beliefs about their abilities in the estimation of these effects and look at the effect of these policies across the ability distribution. For the median male and female student, increasing graduation requirements in math and science primarily induces students to graduate from college in non-STEM fields. In addition, for males, I find potentially important increases in the probability of dropping out of high school as a result of raising graduation requirements in math, with evidence suggesting that these increases are driven by students from the bottom of the ability distribution. Any effects on educational outcomes appear to be the result of changes to students' accumulation of different STEM courses and not from changes to students' beliefs.

Finally, I quantify the effects of graduation requirements on student welfare. I do this by estimating a full dynamic structural model of high school curriculum and college major choice and use it to determine the welfare effects of increasing graduation requirements in math and science at the beginning of high school. Point estimates suggest that the effects on student welfare are small. However, my estimates are imprecise, and I am unable to rule out very high negative effects of strict curriculum standards on median students' welfare.

Ultimately, my findings indicate that graduation requirements may have a positive impact for some students while potentially placing an undue burden on others. The prospect of stricter requirements in high school STEM significantly increasing collegiate STEM attainment for the

[2]See Heckman and Vytlacil (2001) for a discussion of policy relevant treatment effects and their relationship to other common treatment parameters

median student appears dim. However, the welfare analysis does not rule out very small welfare effects of graduation requirement increases, and treatment effect estimates do not rule out rather small increases in high school dropout. Therefore, the possibility remains that the social gains from collegiate attainment in non-STEM resulting from these policies are a net benefit to society.

This dissertation is a single chapter divided into 7 main sections. Section 1.1 lays out a general model of high school curriculum and college major choice while imposing as few functional form and distributional assumptions as possible. Section 1.2 provides a brief overview of the data used in this analysis, with a much more detailed description of its preparation provided in Appendix A. Section 1.3 takes the general model and describes the additional assumptions and structure needed to estimate the effects of information and graduation requirements in high school on future educational outcomes using a ‘quasi-structural’ approach. Section 1.4 discusses the results of this quasi-structural analysis. Section 1.5 provides an overview of the additional assumptions needed to estimate a full structural model of high school curriculum and college major choice that allows for the analysis of welfare effects. Section 1.6 presents the results of the structural analysis. Finally, Section 1.7 concludes.

Chapter 1

The Information Content of High School Courses: How Early Exposure to STEM Impacts Future Decisions

1.1 Model Overview

In this portion of the paper, I describe the model under which it is assumed that students are making their educational choices. In order to keep the discussion as general as possible, I forgo making any specific functional form or distributional assumptions until later sections which describe the implementation and estimation of the model using real data.

1.1.1 Initial Conditions

Students begin making decisions in their 9^{th} grade year. Each individual is enrolled in a high school that has observable characteristics S_i which can affect students' experiences at that school. Students are defined by a set of observable characteristics that influence their decisions. Some, X_i and τ_i, are fixed at the beginning of the decision process, while others, Z_{it}, can change over time based on people's choices. X_i and Z_{it} are assumed to be observable, with certainty, to both the student and the econometrician, while τ_i is assumed to only be observable to the student. Both types of individual level variation may influence the utility of different educational options and a student's performances in academic courses.

Before high school begins, students are also endowed with a vector of abilities, $\mathbf{A}_i$, which are unobservable to both the student and the econometrician. Instead, students begin high school with only an understanding of how these abilities are distributed in the population, and as they move through their schooling, they gather information about where they are likely to lie within the distribution. Where a student believes herself to be in the ability distribution at any given time affects her relative valuation of the educational options available to her.

1.1.2 The Decision Problem: Overview

From any action, d_{it} taken at time t, students receive a current period utility, U_{it}^{d}. The ultimate goal of each agent is to make a sequence of choices over the course of their lifespan, $\mathbf{d_i}$, that maximizes their total discounted expected lifetime utility:

$$\mathbf{d}_i = \text{argmax } E\left\{\sum_{t=9}^{T}\beta^{t-9}\sum_{d}[U_{it}^{d}]1\{d_{it}=d\}\right\} \tag{1.1}$$

where β is a discount factor and d is any choice that is available to the student at time t. This rather complicated expression can be re-written using Bellman's Principle of Optimality, which states that an equivalent problem is for the student to choose:

$$d_{it} = \text{argmax } U_{it}^{d} + \beta E\left\{V_{it+1}|d_{it}=d\right\} \tag{1.2}$$

for every period t, where d_{it} is the student's decision in time t, and $E\left\{V_{it+1}|d_{it}=d\right\}$ represents the student's expected maximum achievable lifetime utility onward from t given that they chose d in that period.

A student's entire life-time can be broken into three main stages: high school, college, and the labor market. High school is assumed to last four periods: 9^{th} grade, 10^{th} grade, 11^{th} grade, and 12^{th} grade. Upon completing high school a student can then decide whether to complete college and, if so, in which field to major. Once college has been completed, all students enter the labor force and get paid a wage based on their individual characteristics and previous educational experiences.

1.1.3 The Decision Problem: Labor Market

The choices students face, and the utility of the options available to them, differ across the three stages of their lifetime. While students who have completed college have no option but to enter the labor force, after 9^{th} grade any student can choose to quit their schooling and work instead. Such a decision is assumed to be permanent, meaning that once students have chosen to discontinue their

schooling, they may not re-enter. While this is clearly an unrealistic assumption, the data being used in this study do not contain enough detailed information to model adequately leaving and re-entering school.[1].

Each year, t, students place a value on labor market participation:

$$U_{it}^{L} = E\{U_{it}^{L}(W_{it}, \epsilon_{it}^{L})|I_{it}\} \tag{1.3}$$

where W_{it} is total wages (earnings) in that period, ϵ_{it}^{L} is an unobserved idiosyncratic shock to utility, and I_{it} is the student's information set at time t. This information set is influenced by the years of experience students have taking different STEM and non-STEM courses throughout high school, and determine the beliefs they have about where they are in the ability distribution at time t. Generally speaking, the wages that are earned in a particular period may be a function of some, or all, of the individual's observable and unobservable characteristics, along with a wage specific idiosyncratic shock, γ_{it}^{w}:

$$W_{it} = f(X_i, Z_{it}, \tau_i, \mathbf{A}_i, \gamma_{it}^{w}) \tag{1.4}$$

As a result of the assumption that students cannot re-enter schooling, the value of leaving school (according to Bellman's Principle of Optimality) is simply one's current period discounted expected value of life-time utility in the labor market[2]:

$$v_{it}^{L} = E\left\{\sum_{t=\bar{t}}^{T=65} \beta^{\bar{t}-t} U_{it}^{L}\right\} \tag{1.5}$$

where β is a discount factor and the expectation is taken over the distribution of one's beliefs.

[1]The implications of such an assumption are discussed in Stange (2012), who also estimates a dynamic model of education choice that imposes this restriction on labor market entry. For a model of major choice that *does* incorporate labor market entry and exit see Arcidiacono et al. (2016).

[2]I assume individuals live until 65 in this model. Students begin their schooling at age 14, graduate high school at age 18, and exit college at age 22

1.1.4 *The Decision Problem: College*

If the individual has not entered the labor market, she is either enrolled in college or high school. In college, students have the option to major in one of three different fields of study: non-STEM (NS), math and physical science (MPS), or life science (LS)[3]. Because the data I am using does not have very detailed information on the timing of college major choices beyond one's final major choice, college is modeled as lasting one period[4].

A student's current period utility from choosing to major in field m is:

$$U_{it}^{m} = E\{U_{it}^{m}(X_i, Z_{it}, \tau_i, \mathbf{A}_i, \epsilon_{it}^{m})|I_{it}\} \tag{1.6}$$

While equation (1.6) describes a student's current period utility from choosing one of the post-secondary options available, students also think dynamically about how that choice today will impact utility in the future. In the case of college, they consider how the field in which they choose to major will influence their future lifetime utility in the labor market:

$$\begin{aligned} v_{it}^{m} &= E\{u_{it}^{m}(X_i, Z_{it}, \tau_i, \mathbf{A}_i, \epsilon_{it}^{m})|I_{it}\} + \beta E\left\{V_{it+1}|X_i, I_{it+1}, Z_{it+1}, \tau_i, d_{it} = m\right\} \\ &= E\{u_{it}^{m}(X_i, Z_{it}, \tau_i, \mathbf{A}_i, \epsilon_{it}^{m})|I_{it}\} + \beta E\left\{\sum_{\bar{t}=t+1}^{T} \beta^{\bar{t}-t+1} U_{i\bar{t}}^{L}\right\} \end{aligned} \tag{1.7}$$

1.1.5 *The Decision Problem: High School*

In high school, students choose combinations of courses to take each year, of which they have a maximum of 9 options from which to choose. A student may take a year of non-STEM (NS), mathematics (M), physical science (PS), and/or life science (LS) in any given year of high school

[3]Splitting STEM major categories into math/physical science and life science allows me to more closely look at differences in STEM major choice across genders, as Kahn and Ginther (2017) show that the gender gap in STEM primarily exists between 'math intensive' and 'non-math intensive' majors. For another model of college major choice that distinguishes between math intensive STEM and life science, see Saltiel (2019)

[4]The purpose of this paper, and the model, is to describe and predict how choices made in high school impact ultimate college choices, not to model decision-making in college explicitly. While this assumption is a simplification of students' true decision-making environments, it should not impact my analysis of the relationship between high school choices and college outcomes too dramatically.

enrollment. I assume that enrollment in high school necessitates taking a year of non-STEM courses[5]. Having the option to dropout, take a year of non-STEM, and any possible combination of the three types of STEM courses, means that the largest set of options a student can face is:

$$\begin{pmatrix} \text{Dropout} \\ NS \\ NS, M \\ NS, PS \\ NS, LS \\ NS, M, PS \\ NS, M, LS \\ NS, PS, LS \\ NS, M, PS, LS \end{pmatrix} \tag{1.8}$$

Not every student will face this exhaustive choice set in every year. For one, students are assumed to be unable to dropout of high school until at least grade 10, meaning agents in the model are precluded from choosing the dropout option until then. More importantly, however, is that graduation requirements, which vary across individual high schools, also shape the feasible set of options that students from which students can choose. For instance, if a student is required to take 4 years of math to graduate high school, they will *never* face a choice set with the options $\{NS\}, \{NS, PS\}, \{NS, LS\}$ or $\{NS, PS, LS\}$ available to them. Students who finds themselves in high schools requiring 3 years of math, on the other hand, will face this same restricted choice set only if they reach a point where they *must* take mathematics in order to satisfy graduation requirements before the end of high school (i.e. if they reach their senior year having only taken 2 years of math previously)[6].

[5]This assumption is not very restrictive, as *very* few students report being enrolled in high school and not taking any courses that are classified as non-STEM. Appendix A gives an overview of how courses are classified and how many students in the sample report not taking non-STEM during a year of enrollment in high school

[6]When satisfying science graduation requirements, students can choose to take either physical science, life science, or both. I assume that taking both of these in the same year only reduces the number of required years left by 1

The current period utility students receive from taking a bundle of courses, b, from their feasible set at time t may depend on student level observable and unobservable characteristics, along with high school specific characteristics S_i:

$$U_{it}^b = E\left\{u_{it}^b(X_i, Z_{it}, \tau_i, S_i, \mathbf{A}_i, \epsilon_{it}^b)|I_{it}\right\} \tag{1.9}$$

As with college, students think dynamically about their choices, meaning the total value of choosing a particular bundle of courses is:

$$v_{it}^b = E\{u_{it}^b(X_i, Z_{it}, \tau_i, S_i, \mathbf{A}_i, \epsilon_{it}^b)|I_{it}\} + \beta E\left\{V_{it+1}|I_{it+1}, X_i, Z_{it+1}, \tau_i, S_i, d_{it} = b\right\} \tag{1.10}$$

However, unlike college, high school lasts 4 periods. Therefore, the future attainable utility from after making a decision, $E\left\{V_{it+1}|I_{it+1}, X_i, Z_{it+1}, \tau_i, S_i, d_{it} = b\right\}$, is not simply one's expected lifetime utility in the labor market from that point forward, but instead the *expected maximum* attainable utility of all possible choices in the next period:

$$E\left\{V_{it+1}|I_{it}, d_{it} = b\right\} = E\left\{\max\left(v_{it+1}^{\text{dropout}}, v_{it+1}^{NS}, v_{it+1}^{NS,M}, ..., v_{it+1}^{NS,M,PS,LS}\right)|I_{it}, d_{it} = b\right\} \tag{1.11}$$

The expectation in this expression is taken over all possible future idiosyncratic utility and information shocks.

1.1.6 The Learning Process

An integral part of the above choice model is that students make decisions under uncertainty, of which there are two types. The first type is a general uncertainty about the exact utility one will receive from making any choice d. This uncertainty is captured by the idiosyncratic utility shocks, ϵ_{it}^d, entering all of the choice equations in the model. The value of this shock is revealed to the student before making their current period choices, but they do not know what its values will be in the periods that come after. Because these shocks are idiosyncratic, their values are consistently

unpredictable over time. Students cannot gain any additional insight into what they will be, and only know the distribution from which the shock is drawn. They do, however, factor this distributional information into their valuation of the different options available to them.

The second type of uncertainty in this model is uncertainty about one's underlying abilities, $\mathbf{A}_i$. Unlike the utility shocks, students *can* gain insight into what these values are as they move forward through their decision process. This information is gathered through the grades or GPAs they receive when taking courses.

Assume that abilities are continuous, and let the population distribution of abilities be denoted $\psi(\mathbf{A}_i)$, from which every individual has an ability vector drawn before they enter high school. In any given period, student i has a belief about where they are in this distribution, which can itself be represented by the probability distribution function $\tilde{\psi}_{it}(\mathbf{A}_i)$.

When a student takes a year of a subject, j, the student receives a subject specific GPA for that year. The exact value of this GPA, G^j_{it}, can be a function of both observable and unobservable characteristics, along with an idiosyncratic shock γ^j_{it} drawn from some probability distribution:

$$G^j_{it} = G^j(X_i, Z_{it}, \tau_i, \mathbf{A}_i, \gamma^j_{it}) \tag{1.12}$$

Students know their values of X_i, Z_{it}, and τ_i and how they will affect the grades that they receive. If these were the only factors influencing grades, then students would be able to perfectly predict their GPA every period. However, both unobserved ability and the idiosyncratic grade shock affect grades as well, which causes any GPA *actually* received to deviate in some way from where the student would expect it to be. Because two random elements affect GPAs simultaneously, students cannot perfectly determine how much of this deviation is due to their unobserved abilities and how much can be attributed to idiosyncratic elements. However students are assumed to be Bayesians, and given the information available to them they can update their beliefs in a systematic manner at the end of every period to get successively better approximations of where they are likely to lie in the population ability distribution.

Let $\mathbf{G}_{it}$ be a vector of grades that the student received at the end of time t. Her beliefs at the beginning of the next period, $t + 1$, can be written using Bayes rule as:

$$\tilde{\psi}_{it+1}(\mathbf{A_i}|\mathbf{G}_{it}) = \frac{\tilde{\psi}_{it}(\mathbf{A}_i)P(\mathbf{G}_{it}|\mathbf{A}_i)}{\int P(\mathbf{G}_{it}|\mathbf{A}_i)\tilde{\psi}_{it}(\mathbf{A}_i)d\mathbf{A}_i} \tag{1.13}$$

where $P(\mathbf{G}_{it}|\mathbf{A}_i)$ is the probability distribution function of grades conditional on unobserved ability (and other known factors which are suppressed for clarity). This posterior distribution has a variance that shrinks as the number of grades received increases. If a student received an *infinite* number of signals on their abilities, then this variance would be zero and the student would know their abilities *exactly*. However, given the finite number of signals a student can realistically achieve in their lifetime, some uncertainty will always remain about their true underlying position in the population distribution.

1.2 Data

The data being used in this study is the National Center for Education Statistics' (NCES) Educational Longitudinal Study of 2002 (ELS) (NCES (2002)). This is a panel data set following a nationally representative sample of United States high school students, who were in 10^{th} grade in 2002, through their secondary and post-secondary educational years and into their early labor market experiences.

Initial sampling of students was conducted in two stages. First, a random sample of high schools, including public and private, was selected. From this group, a random sample of students from within each school was collected. Student participants and their parents, along with students' teachers and their school's administrators, completed extensive questionnaires touching on topics related to student academic performance, home life, teachers' qualifications and perceptions, school characteristics and requirements, and other detailed educational information. In addition, students were administered standardized tests in both reading and mathematics in order to compare their academic preparation with other students around the nation.

An additional 3 surveys were conducted over the course of the next several years. The first was completed 2 years later, in 2004, when most students in the sample were completing their final year of high school. Participants were surveyed on similar topics as the base-year survey, and were also questioned about their post-secondary expectations and plans. They were also administered another standardized test in mathematics, both to compare their aptitude to students in other schools and to measure any gains in quantitative prowess since the first wave of data collection.

A third survey was conducted in 2006, 2 years into many of the 10^{th} grade cohort's post-secondary education, while a final survey was collected 6 years later in 2012. As most college-going sample members had completed their undergraduate studies by this point, this final survey collected information on final degree obtained and early labor market data on jobs and wages.

Two components of these data are vital to the completion of this study. The first is the availability of information on individual high schools' graduation requirements in math and science collected from participant school administrators. While academic standards specifying the number of math and science classes that students must complete are often set at the state level, it is also common for local school districts to set their own standards. A state may require all of its high school students to complete 3 years of math to graduate, but a high school within that state can require all of its students to complete 4 years. Because this study models individual choice behavior, and requirements at the local level are likely more relevant to what the student *actually* must complete in order to graduate, utilizing this information is vital to the estimation of my model.

Secondly, in the second and fourth wave of survey collection the NCES also collected students' transcripts from their high schools and all colleges attended respectively. This gives me access to information on every course that a student completed in every year of high school and college. I am able to classify courses in high school as being non-STEM, math, physical science, or life science using a taxonomy developed by the NCES to categorize courses (NCES (2000b)). I can also calculate overall GPAs in college for students who major in different fields.

I use the data available on individual, family, and high school level characteristics, along with the information from high school and college transcripts, to generate a panel of course-taking and

college-going decisions from age 14 and onwards. A detailed overview of how the data is cleaned and constructed is given in Appendix A.

1.3 Quasi-Structural Approach

The model described in Section 1.1 is very general. It is presented without functional forms for the grade equations or the utility functions, and without imposing distributional assumptions on the random variables. In this section, I add more structure onto the model to transform it into one that is conducive for estimation. This allows me to answer two of the three main research questions presented in the introduction.

1.3.1 Adding More Structure: Learning Process

Let us assume that unobserved ability has two components:

$$\mathbf{A}_i = [R_i, Q_i] \sim \psi(R_i, Q_i) \tag{1.14}$$

where R_i is a verbal ability and Q_i is a quantitative ability[7]. The distribution of these unobserved abilities is left unrestricted; the two factors may be correlated and the shape of the distribution may take any form. However, let us further assume that equation (1.12) takes the following parametric form for every subject j[8]:

$$G_{it}^{j} = \alpha_0^j + \alpha_1^j X_i + \alpha_2^j Z_{it} + \alpha_{3t}^j R_i + \alpha_{4t}^j Q_i + \gamma_{it}^j, \;\; \gamma_{it}^j \sim N(0, \sigma_{jt}^2) \tag{1.15}$$

where j can be any combination of school type (high school or college) and subject matter (non-STEM, math/physical science, life science). Let us also specify equation (1.4) as:

$$\ln(W_{it}) = \alpha_0^w + \alpha_1^w X_i + \alpha_2^w Z_{it} + \alpha_3^w R_i + \alpha_4^w Q_i + \gamma_{it}^w, \;\; \gamma_{it}^w \sim N(0, \sigma_{wt}^2) \tag{1.16}$$

[7] See Kinsler and Pavan (2015) who also estimate a structural model of education choice with a quantitative (math) and verbal ability

[8] Arcidiacono et al. (2016) specify grade and wage equations in a similar manner

Here a few additional assumptions are made beyond the general grade and wage functions presented in Section 1.1. First, it is assumed that grades and *log wages* are related to observable and unobservable characteristics in a linear fashion. Second, the idiosyncratic components of grades are assumed to be additively separable from both observables and abilities, and are also normally distributed. Finally, unobserved variation in the student population,τ_i, that is known to the individuals but not the econometrician is assumed to not enter grade and wage equations. This means that the only unobserved heterogeneity affecting these outcomes is academic abilities.

In addition to grades and wages, performances on three standardized cognitive tests that were administered to students by the ELS are also modeled: a verbal test and a quantitative test given in grade 10, and another quantitative test given in grade 12. Scores on these exams take the following form:

$$
\begin{aligned}
T_i^{R,10} &= \alpha_0^{R,10} + \alpha_1^{R,10} X_i + \alpha_2^{R,10} R_i + \gamma_i^{R,10} \quad \gamma_{it}^{R,10} \sim N(0, \sigma^2_{R,10t}) \\
T_i^{Q,10} &= \alpha_0^{Q,10} + \alpha_1^{Q,10} X_i + \alpha_2^{Q,10} Q_i + \gamma_i^{Q,10} \quad \gamma_{it}^{Q,10} \sim N(0, \sigma^2_{Q,10t}) \\
T_i^{Q,12} &= \alpha_0^{Q,12} + \alpha_1^{Q,12} X_i + \alpha_2^{Q,12} Q_i + \gamma_i^{Q,12} \quad \gamma_{it}^{Q,12} \sim N(0, \sigma^2_{Q,12t})
\end{aligned}
\tag{1.17}
$$

It is assumed that students do not actually observe their scores on these exams,[9] and therefore information on how they performed is not used by them when making decisions. However, jointly estimating test score equations alongside the grade and wage equations aids in identification of the ability distributions.

By assuming a linear form on the grade equations, learning in this model can be thought of in an intuitive manner. Every student enters a period of schooling with knowledge about themselves, X_i and Z_{it}, and how someone with these characteristics will, on average, perform in subject j. They then form an expectation about what grade they will receive if they choose to take a course

[9]While I was unable to find ELS documentation that stated whether students are shown the results of their standardized tests, it appears to be common practice in other NCES surveys to not reveal this information. Kinsler and Pavan (2021), which uses data from the Early Childhood Longitudinal Study-Kindergarten Class of 1999 (ECLS-K), rests heavily on this feature of the data. Another NCES survey, the National Assessment of Educational Progress (NAEP), states on its website that participants are not shown their scores (NCES (n.d.)).

in subject j that period:

$$E\left\{G_{it}^j|X_i, Z_{it}\right\} = \alpha_0^j + \alpha_1^j X_i + \alpha_2^j Z_{it} \tag{1.18}$$

When they actually receive this grade, however, it may be higher or lower than they expected. The difference between their realized grades and their expected grades provides a signal on their unobserved ability:

$$\lambda_{it} = G_{it}^j - E\left\{G_{it}^j|X_i, Z_{it}\right\} = \alpha_{3t}^j R_i + \alpha_{4t}^j Q_i + \gamma_{it}^j \tag{1.19}$$

The precision of this signal (i.e. how informative it is of their unobserved abilities) depends on the variance of γ_{it}^j. If it is close to 0, one's deviation from expected grades is due almost entirely to unobserved ability, while if it is very high, there is a strong likelihood that deviations are due to completely random factors. Since students know the distribution of idiosyncratic shocks, they also know how informative each course is on their unobserved abilities. Therfore, when making decisions they can consider, as part of their valuation of taking a particular bundle of courses, how informative that bundle will be about their abilities.

1.3.2 Estimation: Learning Parameters

Fully estimating this learning process requires estimating α_k^j for all j along with the population distribution $\psi(R_i, Q_i)$. If these parameters are known, then it is possible to determine every agents' beliefs at any given time.

Every individual in the sample has a set of data from the period in which they enter the sample ($t = 1$) to the period in which they are last observed ($t = T_i$). These data consist of:

- Observable Characteristics: $\mathbf{X}_i = [X_i, Z_{i1}, ..., Z_{iT_i}, S_i]$
- Grades: $\mathbf{G}_i$
- Cognitive Test Scores: $\mathbf{T}_i$
- Wages at age 26: W_{i26}

With these, the grade, test score, and wage parameters, θ_G, can be estimated by maximizing the log-likelihood function:

$$\Lambda_G = \ln\left[\int \prod_i L_G(\mathbf{G}_i, W_{i26}, \mathbf{T}_i | \mathbf{X}_i, \theta_G, R_i, Q_i)\psi(R_i, Q_i) dR_i dQ_i\right] \tag{1.20}$$

where R_i and Q_i must be integrated out because they are unobserved. However, this makes estimation very difficult. While taking the natural log of a likelihood function typically eases estimation by making terms in the likelihood expression additively separable, this is untrue with the integral appearing inside of the logarithm. However, using recent insights from James (2011) and Arcidiacono et al. (2016), additive separability can be restored and estimation can be feasible by utilizing an extension of the Expectation Maximization (EM) algorithm[10]

The log-likelihood function above can be interpreted as the 'log of the expected likelihood', with the expectation being taken over the joint distribution of R_i and Q_i. Instead, define the 'expected log-likelihood' function as:

$$\begin{aligned} \tilde{\Lambda}_G &= \int \ln\left[\prod_i L_G(\mathbf{G}_i, W_{i26}, \mathbf{T}_i | \mathbf{X}_i, \theta_G, R_i, Q_i)\right] \psi(R_i, Q_i) dR_i dQ_i \\ &= \int \sum_i \ln\left[L_G(\mathbf{G}_i, W_{i26}, \mathbf{T}_i | \mathbf{X}_i, \theta_G, R_i, Q_i)\right] \psi(R_i, Q_i) dR_i dQ_i \end{aligned} \tag{1.21}$$

Each term of this 'auxiliary' function is now additively separable, and with the idiosyncratic errors assumed to be normal, the expected log likelihood for any grade (or wage) j at time t can be written as:

$$\tilde{\Lambda}_G^j = -\frac{1}{2}(2\pi\sigma_{jt}^2) - \int \frac{1}{2\sigma_{jt}^2}(G_{it}^j - E\{G_{it}^j | X_i, Z_{it}\} - \alpha_{3t}^j R_i - \alpha_{4t}^j Q_i)^2 \tilde{\psi}(R_i, Q_i) dR_i dQ_i \tag{1.22}$$

[10]Another key insight from these two papers is that the learning process can be separately estimated from the choice process. This is because choices only depend on *beliefs* about abilities in every period, not *actual* abilities.

which, using the basic algebraic properties of expected values and variances, can be rewritten as:

$$\begin{aligned}\tilde{\Lambda}^j_G = -\frac{1}{2}(2\pi\sigma^2_{jt}) - \frac{1}{2\sigma^2_{jt}}[(\alpha^j_{3t})^2 Var(R_i) + (\alpha^j_{4t})^2 Var(Q_i) + \\ 2\alpha^j_{3t}\alpha^j_{4t} Cov(R_i, Q_i)(G^j_{it} - E\{G^j_{it}|X_i, Z_{it}\} - \alpha^j_{3t}E\{R_i\} - \alpha^j_{4t}E\{Q_i\})^2]\end{aligned} \tag{1.23}$$

The EM algorithm maximizes the likelihood function in equation (1.20) indirectly by repeatedly maximizing this auxiliary function in a procedure described shortly. However, a particular distribution of R_i and Q_i has not been assumed, and in order to maximize equation (1.23) it must be possible to calculate elements of the expression that depend on this distribution, such as $E\{R_i\}$,$E\{Q_i\}$, and $Cov(R_i, Q_i)$.

One way to approach this would be to assume that R_i and Q_i are drawn from some family of distributions. In the literature utilizing learning models such an assumption is commonly imposed, typically an assumption of normality[11]. Such an assumption has some desirable properties in a model where agents update their beliefs in a Bayesian manner. For instance, if abilities are normally distributed (and so are students' priors) then after receiving new information, agents' posterior beliefs will still follow a normal distribution. In addition, the exact mean and variance of this posterior distribution can be calculated quickly using known closed-form expressions. This facilitates computationally feasible estimation, with the trade-off being that the assumption of normality is essentially arbitrary.

In the approach used in this paper, a method of non-parametrically estimating unobserved distributions in choice models, suggested in Train (2008), is extended to estimate the learning process without imposing normality, and avoiding the use of unwarranted distributional assumptions with only slight increases in computational cost.

[11] See Erdem and Keane (1996), Arcidiacono (2004), James (2011), and Arcidiacono et al. (2016) for examples of learning models where unobserved factors are assumed to be normally distributed

This method works by assuming that the marginal distributions of R_i and Q_i have a finite number of support points, κ:

$$\begin{aligned} R_i &\in \{R_1, R_2, ..., R_{\kappa-1}, R_\kappa\} \\ Q_i &\in \{Q_1, Q_2, ..., Q_{\kappa-1}, Q_\kappa\} \end{aligned} \tag{1.24}$$

with the exact values taken on by these support points defined before estimation begins. The marginal distributions of the ability factors, $\psi(R_i)$ and $\psi(Q_i)$, are then defined by a set of probability weights associated with each point:

$$\begin{aligned} \psi(R_i) &= \left\{P_1^R, P_2^R, ..., P_r^R, P_{r+1}^R, ..., P_{\kappa-1}^R, P_\kappa^R\right\}, \quad P_r^R = Pr\left[R_i = r\right] \\ \psi(Q_i) &= \left\{P_1^Q, P_2^Q, ..., P_q^Q, P_{q+1}^Q, ..., P_{\kappa-1}^Q, P_\kappa^Q\right\}, \quad P_q^Q = Pr\left[Q_i = q\right] \end{aligned} \tag{1.25}$$

and their joint distribution, $\psi(R_i, Q_i)$ is the set of all possible combinations of R_i and Q_i grid points along with the probability weights associated with each combination:

$$\psi(R_i, Q_i) = \left\{P_{11}^{RQ}, P_{12}^{RQ}, ..., P_{1,\kappa}^{RQ}, P_{21}^{RQ}, ..., P_{\kappa\kappa-1}^{RQ}, P_{\kappa\kappa}^{RQ}\right\}, \quad P_{rq}^{RQ} = Pr[R_i = r, Q_i = q] \tag{1.26}$$

If one were to assume that R_i and Q_i were independent, estimating this joint distribution would only require estimating 2κ parameters: a set of κ probability weights for R_i and a set of κ probability weights for Q_i. However, in order to be as general as possible an independence restriction is not imposed, meaning that κ^2 parameters need to be estimated.

Train (2008) finds, through Monte-Carlo experiments, that approximating latent distributions using a grid approach such as this performs well as long as the number of grid points, κ, is sufficiently large and the range of the pre-specified values of the grid are wide enough. As the computational complexity of this procedure grows exponentially in κ, selecting a large enough number of grid points for a precise approximation has to be carefully weighed against its feasibility.

Ultimately, $\kappa = 41$ grid points, with values equally spaced from -2.5 to 2.5 were chosen for the marginal distributions (implying a total of $1,681$ grid points for the joint distribution)[12].

By using this grid approach, the unobserved ability distribution, $\psi(R_i, Q_i)$, is now a discrete probability mass function defined by a set of $1,681$ probability weights. Students' beliefs, $\tilde{\psi}_t(R_i, Q_i)$, are also discrete probability mass functions characterized by a set of individual and time specific probability weights $\mathbf{P^{RQ}}_{it}$. The estimation procedure begins with an initial guess of the grade and wage parameters, $(\theta_G^{(1)})$, and joint ability distribution, $(\psi^{(1)}(R_i, Q_i) = \mathbf{P^{RQ}}^{(1)})$. The following steps are then performed:

1. For each individual in the sample, the parameter guesses, observed grades, test scores, and wage outcomes and used with Bayes' rule to get their posterior beliefs for the last period in which they appear in the sample[13]:

$$\begin{aligned} \tilde{\psi}_{iT_i}(R_i, Q_i) = \mathbf{P^{RQ}}_{iT_i}^{(1)} &= Pr[R_i = r, Q_i = q | \mathbf{G}_i, \mathbf{T}_i, W_{i26}] \;\; \forall r, q \\ &= \frac{f(\mathbf{G}_i, \mathbf{T}_i, W_{i26} | R_i = r, Q_i = q) P_{rq}^{RQ}}{\sum_{r'q'} f(\mathbf{G}_i, \mathbf{T}_i, W_{i26} | R_i = r', Q_i = q') P_{r'q'}^{RQ}} \;\; \forall r, q \end{aligned} \tag{1.27}$$

2. Using $\tilde{\psi}_{iT_i}(R_i, Q_i)$, maximize the expected log-likelihood in equation (1.23) to obtain new grade and wage parameter estimates $\theta_G^{(2)}$.

3. Use the set of individual posterior beliefs, $\mathbf{P^{RQ}}_{iT_i}^{(1)}$ to get a new guess of the population joint ability distribution as:

$$\mathbf{P^{RQ}}^{(2)} = \frac{\sum_i \mathbf{P^{RQ}}_{iT_i}^{(1)}}{N} \tag{1.28}$$

   where N is the number of individuals in the sample.

[12]Ensuring a sufficient range of grid points is not too much of a concern in this application. This is because the scale of the latent abilities is measured in high school GPA points (see subsection (1.3.4)), which places a natural restriction on how large or small the values of the latent abilities can be (as GPAs only range from 0.0 to 4.0). Nevertheless, I confirm that the range was likely chosen to be wide enough by observing that extreme values of the grids are assigned very low probability weights by the estimation procedure

[13]To estimate the grade parameters, I act as if students can see their cognitive test scores and update the students' final posteriors with this information incorporated. However, when calculating beliefs for use in the choice equations later, I omit the test scores in the calculation of posteriors

4. Repeat step 1 with $\theta_G^{(2)}$ and $\mathbf{P}^{\mathbf{RQ}(2)}$ as initial guesses

This set of steps is repeated until the parameter estimates (or the value of the log likelihood function) converge.

1.3.3 Estimation: Choices

With the grade parameters, wage parameters, and the unobserved ability distribution estimated, it is possible to assign beliefs to every individual in the sample at any observed period of decision making. Section 1.1.2 gives an overview of how students are assumed to be making decisions based, in part, around these beliefs. It is a model of utility maximization and dynamic optimization, where students have information on observed and unobserved factors about themselves and face persistent and idiosyncratic shocks to their valuations of different educational pathways over time. To actually estimate the preference parameters, θ_u, governing the decision process requires more assumptions and structure on the choice model than have been imposed so far. However, useful analysis can still be conducted without fully estimating the dynamic discrete choice model. Taking beliefs as given, reduced-form analyses that relate observable and unobservable characteristics to outcomes can be conducted.

This is the approach taken in this section. For each period of schooling, the conditional choice probabilities (CCP) of any given decision, conditional on students' observable characteristics, beliefs, and a set of fixed unobservable characteristics, are estimated. Once the parameters of the CCPs have been estimated, they are used to run simulations that change the parameters of the learning process, or alter high school curriculum policy, with the intent of determining the resulting effects on students' final educational outcomes. However, given that this is a reduced-form approach, any simulations must be conducted within the empirical support of the data. To generate out of sample predictions, or conduct welfare analyses, the full structural model would need to be estimated, as is done in Section 1.5. Approaching the problem in this manner is similar to an approach used by Bernal and Keane (2010), who lay out a dynamic choice model of female employment and child-care decision-making and then approximate the decision rules of female

agents in what they deem a 'quasi-structural' approach. More similar to this paper, Thomas (2019) uses such an approach, combined with the estimation methods developed in James (2011) and Arcidiacono et al. (2016), to approximate decision rules and estimate a quasi-structural model of college course-taking decisions and learning about unobserved abilities.

Let $Pr[d_{it} = b | X_i, Z_{it}, S_i, \tilde{R}_{it}, \tilde{Q}_{it}]$ denote the CCP of making decision b at time t, conditional on observable characteristics, X_i and Z_{it}, observable high school characteristics, S_i, and beliefs about verbal and quantitative ability at time t, $\tilde{R}_{it}$ and $\tilde{Q}_{it}$. While, in theory, such a choice probability could be estimated from the data non-parametrically, the curse of dimensionality resulting from the large number of characteristics being controlled for makes this infeasible. Instead, we can approximate this using a flexible multinomial logit:

$$\begin{aligned} Pr[d_{it} = b | X_i, Z_{it}, S_i, \tilde{R}_{it}, \tilde{Q}_{it}] &\approx \tilde{P}^b_{it}(X_i, Z_{it}, S_i, \tilde{R}_{it}, \tilde{Q}_{it}) \\ &\approx \frac{\exp(f(X_i, Z_{it}, S_i, \tilde{R}_{it}, \tilde{Q}_{it}; \beta^{CCP,b}))}{\sum_{b'} \exp(f(X_i, Z_{it}, S_i, \tilde{R}_{it}, \tilde{Q}_{it}; \beta^{CCP,b'}))} \end{aligned} \tag{1.29}$$

where $\beta^{CCP,b}$ is a vector of choice specific parameters and $f(X_i, Z_{it}, S_i, \tilde{R}_{it}, \tilde{Q}_{it})$ is a polynomial index function. If equation (1.29) is sufficiently flexibly specified, and students' decisions only depend on the observable characteristics and beliefs for which have been controlled, then any predicted variation in choice probabilities resulting from variation in these variables can be interpreted causally.

However, it is likely that there are still characteristics driving differences in students' choice patterns that are not accounted for by this set of observables and beliefs. In order to control for potential remaining unobserved heterogeneity, a mixture model is estimated instead of a standard multinomial logit. It is assumed that all students have a fixed 'type', τ_i, which is drawn from a distribution with 5 points of support[14]. Types are potentially correlated with a set of fixed observable individual level characteristics $\tilde{x}_i$. Denote the population probability that any individual with characteristics $\tilde{x}_i$ is of 'type' τ as $\pi^{\tau}_{\tilde{x}}$. The likelihood function for this flexible mixed multinomial

[14]Modeling unobserved heterogeneity in this manner is a common feature of dynamic models of choice. See Heckman and Singer (1984) for the classic example of such an approach.

logit can then be written as:

$$
\begin{aligned}
L^{CCP} &= \prod_i \sum_\tau \pi_{\tilde{x}}^\tau \prod_t \prod_b \left[\tilde{P}_{it}^b(\tilde{x}_i, X_i, Z_{it}, S_i, \tilde{R}_{it}, \tilde{Q}_{it}, \tau_i = \tau)\right]^{\mathbb{1}\{d_{it}=b\}} \\
\rightarrow \ln[L^{CCP}] &= \sum_i \ln\left[\sum_\tau \pi_{\tilde{x}}^\tau \prod_t \prod_b \left[\tilde{P}_{it}^b(\tilde{x}_i, X_i, Z_{it}, S_i, \tilde{R}_{it}, \tilde{Q}_{it}, \tau_i = \tau)\right]^{\mathbb{1}\{d_{it}=b\}}\right]
\end{aligned} \tag{1.30}
$$

Maximizing the parameters of this likelihood function allows for the calculation of conditional choice probabilities for any combination of observables and unobservables (within the empirical support of the data). However, maximization of this function runs into the same issues encountered when attempting to maximize the grade and wage likelihood function described in section 1.3.2, namely that the log-likelihood is not additively separable. Applying the EM algorithm to this problem can ease maximization in a similar manner as before.

For this problem, each iteration, ζ, of the EM algorithm begins with a guess of the population type probabilities, $\pi_{\tilde{x}}^{\tau,\zeta}$, and a guess of the CCP parameters, $\theta^{CCP,\zeta}$. Treating these guesses as fixed, observed choices over all periods in the sample are used to calculate the posterior probability that individual i is type τ:

$$
\begin{aligned}
&\Pr[\tau_i = \tau | \tilde{x}_i, X_i, Z_{it}, \tilde{R}_{it}, \tilde{Q}_{it}, d_{i1}, ..., d_{iT_i;\theta^{CCP,\zeta}}] = q_i^{\tau,\zeta+1} \\
&= \frac{\pi_{\tilde{x}}^{\tau,\zeta} \prod_t \prod_b \left[\tilde{P}_{it}^b(\tilde{x}_i, X_i, Z_{it}, S_i, \tilde{R}_{it}, \tilde{Q}_{it}, \tau_i = \tau)\right]^{\mathbb{1}\{d_{it}=b\}}}{\sum_{\tau'} \pi_{\tilde{x}}^{\tau',\zeta} \prod_t \prod_b \left[\tilde{P}_{it}^b(\tilde{x}_i, X_i, Z_{it}, S_i, \tilde{R}_{it}, \tilde{Q}_{it}, \tau_i = \tau')\right]^{\mathbb{1}\{d_{it}=b\}}}
\end{aligned} \tag{1.31}
$$

Then, using these *conditional* type probabilities, guesses of the CCP parameters are updated by maximizing the expected log-likelihood function:

$$
\sum_i \sum_t \sum_\tau q_i^{\tau,\zeta+1} \ln\left[\prod_b \left[\tilde{P}_{it}^b(\tilde{x}_i, X_i, Z_{it}, S_i, \tilde{R}_{it}, \tilde{Q}_{it}, \tau_i = \tau)\right]^{\mathbb{1}\{d_{it}=b\}}\right] \tag{1.32}
$$

Finally, population type probabilities are updated using the following formula:

$$
\pi_{\tilde{x}}^{\tau,\zeta+1} = \frac{\sum_i (q_i^{\tau,\zeta+1})(\mathbb{1}\{\tilde{x}_i = \tilde{x}_i\})}{\sum_i (\mathbb{1}\{\tilde{x}_i = \tilde{x}_i\})} \tag{1.33}
$$

These steps are repeated until the difference in parameter values (or change in the likelihood function) between each iteration has converged to approximately 0.

The index function specified in this finite mixture multinomial logit, $f(\tilde{x}_i, X_i, Z_{it}, S_i, \tilde{R}_{it}, \tilde{Q}_{it}, \tau_i)$ is a flexible functional form that includes time fixed effects, 'type' specific intercepts for all high school choices, and interactions and higher order polynomial terms for many of the observable characteristics and students' beliefs. In addition, high school graduation requirements are explicitly taken into account by restricting the CCP choice set whenever remaining math and science requirements would preclude a student from taking a particular combination of courses in time t. Two variables that measures, in each year, how many years of math and science a student has left to take are also included in the index function. To see why this is important, suppose there are two identical students in grade 9. One must take 3 years of math to graduate, while the other only has to take 2 years. In grade 9, neither of their choice sets will be restricted, as both could technically refuse to take math that year and still have enough time left in high school to complete their requirements. If these variables measuring how many years of math and science a student has left to take were not included, then the CCP for both of these students would be the same for every possible choice. This is likely not the case, however, because the student required to take 3 years of math will have their choice set restricted in grade 10 if they do not take math today, which still may affect their choices made today relative to the other student. Including these extra variables takes this dynamic consideration into account in the CCP estimation.

The set of observable characteristics that is allowed to be correlated with a student's unobserved 'type', $\tilde{x}_i$, consists of gender and high school graduation requirements. This is done, primarily, to control for the fact that students in high schools with stricter graduation requirements in math and science may be systematically different from students in schools with lesser requirements. Instead of conditioning 'type' on all possible combinations of math and science graduation requirements (of which some combinations have *very* little support in the data), the *maximum* number of years of math and science required at one's high school in grade 9 is used. If a student attends a high school requiring 3 years of math and 4 years of science, they will be categorized under the same

conditioning set as someone who attends a high school requiring 4 years of math and 4 years of science, *or* 4 years of math and 3 years of science. Because there are very few students in the data who are required to take fewer than 2 years of math or science, these students are categorized with anyone attending a high school where the maximum number of required years is 2.

1.3.4 Identification

Identification of the learning parameters requires a set of exclusion restrictions and normalizations. In order to separately identify verbal and quantitative abilities, at least one of the grade, test score, or wage equations must admit only verbal ability while another only admits quantitative ability. To meet this requirements, it is assumed that only verbal ability affects high school non-STEM grades, and only quantitative ability affects high school math/physical science grades. Scores on the standardized test administered by the ELS are also subject to exclusion restrictions, with verbal ability assumed to only affect verbal test scores and quantitative ability assumed to only affect quantitative test scores. Non-STEM and math/physical science grades are allowed to be impacted by both ability factors in college.

The mean of the ability factors is not separately identified from the intercept terms in the grade, test score, and wage equations. Because of this, the means of the verbal and quantitative ability distributions must be normalized to 0. Additionally, assumptions are needed to identify the coefficients on ability factors in these equations. Abilities have no inherent scale, so one of the verbal ability coefficients and one of the quantitative ability coefficients must be normalized to 1. The choice of equation in which to do this is arbitrary and does not affect estimation results, however it *does* affect the interpretation of the values of the ability factors. For verbal ability, its coefficient in high school non-STEM is chosen to be normalized to 1, and for quantitative ability this is done for the coefficient in high school math/physical science. Taken together, these scale and expected value normalizations mean that the values of these ability factors should be interpreted in terms of deviations, in high school non-STEM (or math/physical science) GPA points, from expected GPA.

Identifying the distribution of unobserved heterogeneity relies heavily on the dynamic nature of the model, as an individual's unobserved heterogeneity 'type' is identified primarily off of one's choices over time compared one's observationally equivalent peers (Arcidiacono et al. (2016)). For instance, suppose one unobserved heterogeneity 'type' in the population greatly enjoys mathematics. If we observe two individuals with identical characteristics, but with one consistently choosing to take more mathematics throughout high school, the estimation procedure would put a strong weight on that individual having the unobserved 'type' that enjoys mathematics.

The learning process that affects students' choices over time also provides exogenous variation that can help identify choice parameters. Imagine, again, that there are two individuals who are completely identical. They have the same observable characteristics, the same unobserved heterogeneity 'type', and, in reality, lie in the same part of both the verbal and quantitative ability distributions. These students will begin their journey through high school with the same beliefs about their abilities, and may choose to take the same courses in their first year. Although they will, on average, receive the same grades, at any given time the grades that they *actually* receive will likely differ because of the idiosyncratic shocks to grades. As a result of these differences, one may believe that they have higher abilities than they actually do, while the other may believe they have lower abilities. Because these differences in beliefs were driven by the idiosyncratic shocks, which are completely exogenous, these provide exogenous variation in choices during the next period of the model.

In addition, graduation requirements in math and science can help identify the model. As long as the probability of any two students being placed in high schools with different graduation requirements is independent (conditional on the set of observables and unobservables being controlled for), then difference in curriculum will cause two identical students to make different choices over time for conditionally exogenous reasons. Arguing that conditionally exogenous shocks in one time period can transmit into future periods to provide identification is similar to an argument used in Mroz and Savage (2006).

Finally, variation in course offerings across high schools is used to serve as exclusion restrictions between the choice equations and the grade, test score, and wage equations to help achieve identification. Schools in the ELS data set provide information on the number of math, physical science, and life science courses that they offer at the school which I assume affects the value of choosing to take a year of one of these courses, but not the actual grade one receives.

1.4 Quasi-Structural Approach: Results

This section presents the results from the quasi-structural estimation.

1.4.1 Sample Characteristics

Table 1.1 summarizes the independent variables used in estimation from the final estimation sample. The final sample consists of approximately $8,000$ individuals. In all of the grade and choice equations, controls for basic demographic characteristics such as sex and race are included. Also included are indicators for whether at least one of a students' parents holds a bachelors degree or higher, and whether at least of their parents reports working in a STEM field.

While cognitive factors have been shown to be important determinants of success in schooling and the labor market, it is becoming increasingly clear that non-cognitive skills are, perhaps, equally important (Kautz et al. (2014)). While the ELS does not include direct measures of non-cognitive abilities, such as standardized personality tests, it does include information on teachers' perceptions of students behavior, such as how often they complete their homework and how well they relate to other students. Following an approach in Lleras (2008), these perceptions are used to proxy for non-cognitive abilities and are included in all grade and choice equations.

Finally, as discussed in Section 1.3.4, information from high school course catalogs is utilized to control for the number of math, physical science, and life science courses offered at high schools in the ELS sample. These act as exclusion restrictions, only entering choice equations but not grade or wage equations.

Table 1.1: Summary Statistics

	Variable
Male	.498
	(.5)
Black	.081
	(.273)
Hispanic	.11
	(.313)
Has Parent who Graduated College	.461
	(.499)
Has Parent who Works in STEM	.268
	(.443)
Student Often Completes Homework	.776
	(.417)
Student is Often Tardy	.106
	(.308)
Student is Often Disruptive in Class	.132
	(.339)
Student is Exceptionally Passive	.113
	(.316)
Student Relates Well with Others	.901
	(.298)
Math Courses Offered	14.064
	(6.123)
Physical Science Courses Offered	14.28
	(7.984)
Life Science Courses Offered	9.093
	(6.341)
Observations*	7920

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
*All sample sizes rounded to nearest 10

Table 1.2 reports the fraction of students in the sample who attend high schools with different levels of graduation requirements in mathematics and science. No high school in the sample reports having 0 graduation requirements in math and science, and very few report only requiring their students to take 1 year. The vast majority of students in the sample are required to take 2 or 3 years of math and science if they wish to obtain a high school diploma. However, a somewhat significant number of students *are* subject to taking 4 year in these subjects.

Table 1.2: Graduation Requirements in Math & Science

Years Required	Math	Science
0	0	0
1	.005	.017
2	.248	.307
3	.604	.612
4	.142	.064
Observations*	7920	

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
*All sample sizes rounded to nearest 10

1.4.2 Model Fit

Before conducting simulations from and exploring the estimates of the quasi-structural model, it is important to see how well it fits the moments of the observed data. To do this, the choices of everyone in the estimation sample who does not attrit in grade 9 is forward simulated using the estimated grade and quasi-structural choice parameters 100 times. Table 1.3 displays the model's ability to fit accumulated years of math, physical science, and life science. The columns displaying moments from the estimation data reveal some key features of high school STEM course taking. Although about 25% of the estimation sample attend high schools only requiring two years of math, very few students only take 2 years. The vast majority actually complete a full 4 years of mathematics, meaning that requirements are not binding for many high school students. Men appear to take slightly more math than women, though the difference is not very meaningful.

A larger gap between men and women appears in the physical science and life science categories. Over 50% of men take 3 or 4 years of physical science, while only about 38% of women do. The opposite holds true for life science, where women are more likely to take 2, 3, or 4 years of this subject than men. Overall, however, the total accumulation of physical science courses appears to be greater than life science. 44% of men and 37% of women finish high school having only taken one

Table 1.3: Accumulated Years of High School STEM: Actual vs. Simulated

Math	Actual (Male)	Sim. (Male)	Actual (Female)	Sim. (Female)
0 Years	.000	.000	.000	.000
1 Year	.000	.000	.000	.000
2 Years	.023	.022	.022	.027
3 Years	.267	.277	.285	.297
4 Years	.711	.702	.693	.676
Physical Science				
0 Years	.016	.012	.017	.015
1 Year	.117	.122	.155	.166
2 Years	.357	.368	.444	.444
3 Years	.390	.380	.323	.318
4 Years	.121	.118	.061	.056
Life Science				
0 Years	.013	.016	.009	.007
1 Year	.441	.429	.365	.359
2 Years	.392	.405	.436	.439
3 Years	.129	.122	.152	.156
4 Years	.025	.027	.038	.038

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)

year of life science, whereas 12% of men and 16% of women finish with only one year of physical science courses. Simulated course accumulations match these features well, with differences of only a percentage point or two at most between simulated moments and the actual moments.

Also evaluated is the model's ability to predict the final outcomes of students in the sample, with these comparisons being presented in Table 1.4. Looking at the actual moments from the estimation data shows that men are slightly more likely to drop out of high school than women are, and that men complete college at a slightly lower rate. Among students who *do* go on to complete college, drastic differences appear between men and women. While women are more likely to complete non-STEM majors than men, their take-up of college STEM is very similar, with 11% of men completing STEM majors compared to about 10% of women. However, a clear gap appears when comparing the *types* of STEM majors that they complete. Only 2% of women go on to complete a math/physical science degree, while about 7% of men do. Instead, women are more likely to complete a degree in life science if they decide to major in STEM, with about 8% of women in the sample completing a bachelors degree in this field. Men are less likely to attain a life science degree

than women, but this difference is not as stark as the difference in math/physical science completion between the sexes. The model fits these moments well, though their are a couple of discrepencies. Notably, the model underpredicts the number of high school dropouts and, likely as a result of this, overpredicts the number of students who go on to complete high school but do not finish college.

Table 1.4: Post-H.S. Choices: Actual vs. Simulated

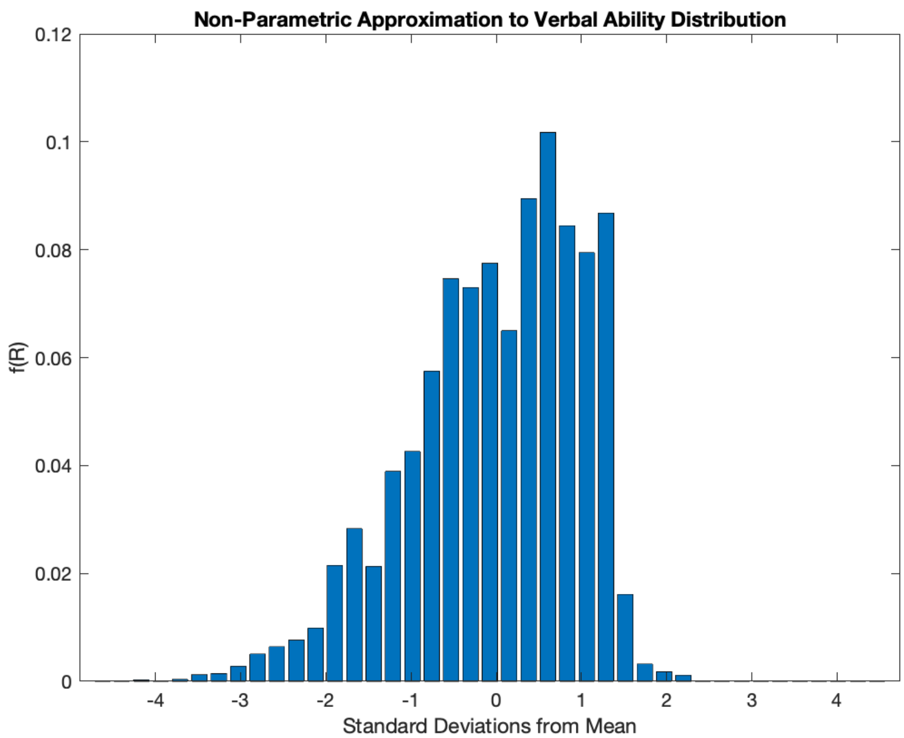
Final Choice	Actual (Male)	Sim. (Male)	Actual (Female)	Sim. (Female)
High School Dropout	.096	.080	.073	.065
High School Diploma Only	.500	.526	.475	.504
Non-STEM Major	.292	.290	.355	.342
Math/Physical Science Major	.071	.065	.017	.016
Life Science Major	.041	.039	.080	.072

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)

1.4.3 Unobserved Ability and Learning

Figure 1.1 presents the non-parametric approximations to the unobserved ability marginal distributions. The scale of the distributions have been normalized by their respective standard deviations. Not too much information can be gained by simply looking at the graphs, though they do not appear to be normally distributed. The distribution of verbal ability has more mass in the left tail than in the right, and the distribution of quantitative ability has a lot of mass centered in the middle and very thin tails.

The estimated standard deviations of the verbal and quantitative distributions are 0.551 and 0.668 respectively, indicating that quantitative ability is slightly more variable in the population than verbal ability. Recall, from subsection (1.3.4) that the ability factors are measured in terms of high school non-STEM GPAs for verbal ability, and math/physical science GPAs for quantitative ability. These estimated standard deviations mean that, a student who lies 1 standard deviation higher in the unobserved verbal or quantitative ability distribution than another, completely identical looking student, will, on average, score about half a GPA point higher in their high school non-STEM or high school math/physical science courses than an observationally equivalent peer.



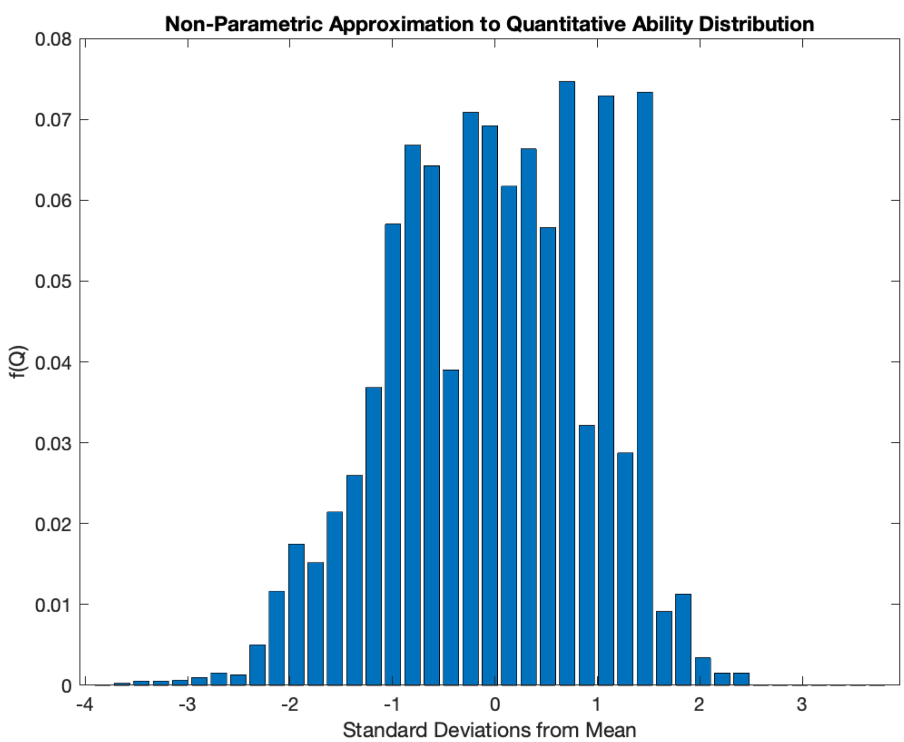


Figure 1.1: Estimated Unobserved Ability Distributions

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)

Estimates of the joint distribution imply that the two ability factors are *not* independent, as their estimated correlation is 0.893. This has important implications for the way in which students learn about their underlying multidimensional abilities, as this implies that a student who performs poorly in a course that completely uses verbal ability will believe herself to be lower in both the verbal *and* quantitative ability distributions. This contrasts with a hypothetical situation in which verbal and quantitative abilities are independent, in which case the student would believe herself to be lower in the verbal ability distribution, but have no additional information about her quantitative ability. This means that a student may discover that she is not well suited for highly quantitative courses and majors after only taking non-STEM courses.

Tables 1.5 through 1.7 display estimated grade and wage parameters, while Table 1.8 contains parameters estimates from the standardized test score equations which are used to help secure identification. Grade and wage equations include the individual level characteristics listed in Table 1.1, with intercepts and effects allowed to vary between high school and college. The interpretation of these parameters is intuitive, they are the effect of different observable student characteristics on the subject specific GPA one would expect to receive if one's unobserved abilities were exactly at the mean of their distributions. As discussed in subsection (1.3.4) certain exclusion restrictions and scaling normalizations must be imposed in order to identify the effects and distributions of verbal and quantitative ability. Verbal ability is excluded from the high school math/physical science GPA equations, and quantitative ability is excluded from the high school non-STEM equation. Life science is allowed to be impacted by both verbal and quantitative ability, and both abilities are allowed to affect all three grade equations in college. Although these are the only exclusion restrictions imposed in grade estimation, these values *are* restricted to be greater than or equal to zero in estimation. This allows the procedure to include extra exclusion restrictions on abilities if the data does not support their inclusion.

By looking at the factor loadings on verbal and quantitative ability across equations, one can get a sense of how important each ability is in performance across different subjects. The loading on verbal ability is much higher than the loading on quantitative ability in both high school and

Table 1.5: Grade Parameters

	Math/Physical Science	Life Science	non-STEM
Constant	2.000	2.242	2.540
	(.028)	(.033)	(.02)
In Third or Fourth Year of Course	−.089	.168	.033
	(.009)	(.018)	(.005)
Male	−.045	−.101	−.140
	(.011)	(.018)	(.01)
Black	−.394	−.310	−.287
	(.024)	(.023)	(.016)
Hispanic	−.222	−.232	−.180
	(.022)	(.03)	(.018)
Parent Has College Degree	.226	.219	.183
	(.014)	(.015)	(.01)
Parent Works in STEM	.017	.020	–
	(.011)	(.014)	
Student Often Completes Homework	.654	.546	.459
	(.018)	(.018)	(.012)
Student Is Often Disruptive	−.234	−.281	−.174
	(.024)	(.028)	(.016)
Student is Often Tardy	−.240	−.28	−.223
	(.023)	(.022)	(.015)
Student is Passive	−.209	−.234	−.158
	(.027)	(.035)	(.02)
Student Relates well to Others	.110	.096	.155
	(.027)	(.032)	(.015)
In College	1.121	.422	.264
	(.045)	(.033)	(.023)
(In College)×Male	−.025	.014	−.003
	(.014)	(.018)	(.011)
(In College)×Black	.131	−.007	.004
	(.023)	(.024)	(.019)
(In College)×Hispanic	−.032	.094	.043
	(.022)	(.025)	(.018)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
*All sample sizes rounded to nearest 10
Standard errors obtained from 50 bootstrap re-samples

college life science, meaning unobserved variation in life science grades is driven more strongly by verbal skills than quantitative. Verbal and quantitative skills appear to be similarly important in college non-STEM academic performance, with verbal ability having slightly higher factor loading. For college math/physical science, verbal ability is given a loading of 0, even though it was not

Table 1.6: Grade Parameters (Cont.)

	Math/Physical Science	Life Science	non-STEM
(In College)×Parent Has College Degree	$-.064$	$-.045$	.001
	(.016)	(.016)	(.011)
(In College)×Parent Works in STEM	.019	.014	–
	(.014)	(.017)	
(In College)×Student Often Completes Homework	$-.562$	$-.319$	$-.215$
	(.017)	(.021)	(.014)
(In College)×Student Is Often Disruptive	.252	.081	.059
	(.021)	(.029)	(.016)
(In College)×Student is Often Tardy	.633	.193	.144
	(.023)	(.021)	(.017)
(In College)×Student is Passive	.241	.037	.050
	(.024)	(.033)	(.014)
(In College)×Student Relates well to Others	$-.155$	.079	$-.170$
	(.025)	(.03)	(.02)
Verbal Ability (H.S.)	–	.987	1.000
		(.029)	
Quantitative Ability (H.S.)	1.000	.201	–
		(.062)	
Verbal Ability (College)	.000	.464	.302
	(.018)	(.024)	(.026)
Quantitative Ability (College)	.553	.062	.236
	(.015)	(.027)	(.02)
σ^2 (In First or Second Year)	.358	.466	.132
	(.004)	(.007)	(.002)
σ^2 (In Third or Fourth Year)	.466	.577	.156
	(.028)	(.07)	(.022)
σ^2 (College)	.162	.140	.132
	(.013)	(.009)	(.025)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
*All sample sizes rounded to nearest 10
Standard errors obtained from 50 bootstrap re-samples

restricted from the equation in estimation, and its coefficient on quantitative ability is the highest out of all three subjects. As with life science, the wage estimation implies that wages are driven more by verbal ability than quantitative, though the standard errors on the loadings are high.

The variances of the idiosyncratic error terms are also estimated. The magnitude of the idiosyncratic variances determines how informative a particular course is on unobserved ability. Variances are allowed to differ between high school and college, and also between lower level and upper level courses. A course is assigned to be upper level if the student has already taken two previous years of that subject, and lower level if two years have not yet been taken. Lower level high school courses are estimated to have a smaller idiosyncratic variance than upper level ones, which is somewhat

Table 1.7: Log Wage Parameters

	ln(Wage)
Constant (H.S. Dropout)	-2.228
	(.072)
Constant (Terminal H.S. Graduate)	-2.148
	(.079)
Constant (non-STEM Major)	-1.988
	(.089)
Constant (Math/Physical Science Major)	-1.785
	(.114)
Constant (Life Science Major)	-1.907
	(.094)
Male	.230
	(.024)
Black	$-.347$
	(.047)
Hispanic	$-.131$
	(.038)
Parent Has College Degree	.014
	(.019)
Parent Works in STEM	$-.050$
	(.028)
Student Often Completes Homework	.087
	(.034)
Student Is Often Disruptive	.098
	(.04)
Student is Often Tardy	$-.004$
	(.041)
Student is Passive	$-.072$
	(.036)
Student Relates well to Others	.283
	(.041)
Verbal Ability	.136
	(.047)
Quantitative Ability	.050
	(.04)
σ^2	.958
	(.146)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
*All sample sizes rounded to nearest 10
Standard errors obtained from 50 bootstrap re-samples

Table 1.8: Cognitive Test Score Parameters

	Grade 10 Verbal	Grade 10 Quant.	Grade 12 Quant.
Constant	.049	−.113	−.269
	(.035)	(.037)	(.043)
Male	−.036	.244	.289
	(.021)	(.018)	(.016)
Black	−.497	−.599	−.549
	(.029)	(.031)	(.03)
Hispanic	−.356	−.367	−.29
	(.026)	(.029)	(.028)
Parent Has College Degree	.387	.370	.439
	(.019)	(.015)	(.02)
Parent Works in STEM	–	.101	.079
		(.018)	(.018)
Student Often Completes Homework	.180	.241	.309
	(.029)	(.022)	(.021)
Student Is Often Disruptive	−.251	−.251	−.259
	(.033)	(.03)	(.03)
Student is Often Tardy	−.079	−.085	−.149
	(.035)	(.027)	(.029)
Student is Passive	−.102	−.192	−.222
	(.034)	(.034)	(.038)
Student Relates well to Others	.072	.072	.031
	(.03)	(.032)	(.04)
σ^2	.584	.428	.378
	(.009)	(.007)	(.007)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
*All sample sizes rounded to nearest 10
Standard errors obtained from 50 bootstrap re-samples

surprising as it suggests that more advanced courses are slightly less informative of latent ability. However, variances for the college GPA error terms are much smaller than in high school implying that performance in college is very informative.

1.4.4 The Information Content of High School Courses

A natural question to ask is how useful high school courses are at resolving the uncertainty that students face about their academic aptitude. Table 1.9 shows of the average amount of uncertainty left over at the end of each high school grade level in the data. Uncertainty is calculated in each year by dividing the posterior variance of students' beliefs by the total variance of the ability factor

in the population (which also acts as the students' prior beliefs in grade 9). After the first period of schooling, only about 20% of the uncertainty about verbal ability and 26% of the uncertainty about quantitative ability remains. Reductions continue, at a diminishing rate, after each subsequent year of high school, with about 7% verbal uncertainty and 12% quantitative uncertainty remaining at the end of high school. These results imply that students learn quite a large amount about their underlying academic aptitude while in high school.

Table 1.9: Average Uncertainty At End of Grade (Percentage of Initial Uncertainty)

Grade	Verbal - Male	Verbal - Female	Quantitative - Male	Quantitative - Female
9^{th}	20.788%	20.707%	26.651%	26.573%
	(6.817)	(6.778)	(5.371)	(5.36)
10^{th}	11.672%	11.640%	16.946%	16.910%
	(3.591)	(3.576)	(3.671)	(3.662)
11^{th}	8.774%	8.751%	13.619%	13.605%
	(2.65)	(2.637)	(3.174)	(3.184)
12^{th}	7.051%	7.041%	11.633%	11.687%
	(2.093)	(2.078)	(2.956)	(3.013)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
Standard errors obtained from 50 bootstrap re-samples

To determine how important these uncertainty reductions in high school are in determining students' ultimate final educational choices, counterfactual simulations are run that increase and decrease the amount information contained in high school STEM courses. The median male and female student in the sample is generated along with every possible combination of high school graduation requirements in math and science ranging from 2 to 4 years required. Each of these individuals' choices are then forward simulated $1,000$ times in three different counterfactual environments: one in which the precision on high school STEM signals is the same as in reality (the baseline), one in which these signals are 75% *more* precise, and one in which signals are 75% *less* precise. For each of the $1,000$ simulations, a different ability vector and unobserved 'type' are drawn from their estimated unconditional distributions. For reference, Table 1.10 displays average final educational choices for this group of median simulated individuals and Table 1.11 displays average years of accumulated high school STEM for this group.

Table 1.12 presents the effects of these counterfactual simulations on the probability that a median individual ends up in different final educational outcomes. The point estimates are not very large, though the standard errors do not rule out that the true effects may be larger than what is estimated. The most notable change is an approximately 1 percentage point increase in the probability of the median male student choosing to major in a math/physical science field when STEM signals are more precise, which is statistically different from 0 at the 5% level. Overall, these

Table 1.10: Post-H.S. Choices: Median Student

Final Choice	Male	Female
High School Dropout	.037	.037
High School Diploma Only	.566	.551
Non-STEM Major	.281	.326
Math/Physical Science Major	.080	.018
Life Science Major	.036	.067

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
Standard errors obtained from 50 bootstrap re-samples

Table 1.11: Accumulated Years of High School STEM: Median Students

Math	Male	Female
0 Years	0	0
1 Year	0	0
2 Years	.032	.041
3 Years	.218	.261
4 Years	.750	.698
Physical Science		
0 Years	.006	.013
1 Year	.093	.143
2 Years	.334	.417
3 Years	.405	.352
4 Years	.162	.075
Life Science		
0 Years	.017	.006
1 Year	.387	.307
2 Years	.408	.436
3 Years	.150	.197
4 Years	.038	.054

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
Standard errors obtained from 50 bootstrap re-samples

results suggest that policy interventions intended to moderately increase the amount of information contained in high school STEM may not be very effective at altering students' ultimate decisions.

Table 1.12: Counterfactuals: Precision of Signal in STEM Courses - Effect Relative to Baseline

	75% Less Precise		**75% More Precise**	
Final Choice	*Male*	*Female*	*Male*	*Female*
High School Dropout	0	$-.001$	.001	.002
	(.001)	(.001)	(.001)	(.001)
High School Diploma Only	.003	.005	$-.002$	0
	(.002)	(.002)	(.003)	(.004)
Non-STEM Major	0	$-.004$	$-.01$	$-.003$
	(.002)	(.002)	(.004)	(.003)
Math/Physical Science Major	$-.003$	0	.011	.002
	(.002)	(.001)	(.003)	(.001)
Life Science Major	0	0	0	0
	(.001)	(.002)	(.002)	(.002)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
Standard errors obtained from 50 bootstrap re-samples

1.4.5 Raising Graduation Requirements: Policy Relevant Treatment Effects

Next, the effect of increasing graduation requirements in mathematics and science on students' future educational outcomes is considered, which is conceptually different from the effect of increasing a students' completion in math and science by some amount, while holding everything else constant. This is because a student attending a high school with higher graduation requirements faces a different set of incentives throughout their *entire* secondary educational experience, affecting other variables over time such as high school completion, coursetaking in other subjects, and beliefs about academic aptitude. As such, the effects being estimated in this section can be considered Policy Relevant Treatment Effects (PRTE) as opposed to the Average Treatment Effect (ATE).

To estimate the PRTE of changing high school graduation requirements, the quasi-structural choice model is used to simulate the median male and female student $1,000$ times in every possible graduation requirement policy environment observed in the data[15]. As with the information

[15] Even though I do observe a very small number students in high schools requiring 1 year of math or science, I do not consider these in the simulations as identification is likely to be very weak in this part of the empirical support.

counterfactual experiments, for each simulated student a different vector of unobserved ability and unobserved heterogeneity type is drawn from their estimated unconditional distributions for each simulated individual.

Table 1.13 displays the PRTE of increasing graduation requirements in math and science on the probability of ending up in different final educational categories. The top of Table 1.13 displays the effects after raising math requirements, and the bottom displays the effects after raising science requirements. The most notable effect of such policies appears to be on the probability of attaining a bachelors degree in a non-STEM field. Raising graduation requirements in mathematics from 2 to 4 years is estimated to increase the probability of the median male student obtaining a non-STEM degree by 4.7 percentage points. The same change is estimated to increase this same probability by 6.7 percentage points for the median female student. There is suggestive evidence that the effect of graduation requirements in mathematics on college STEM attainment is low, with the largest point estimate for this category being a noisy 0.09 percentage point increase for the median male student. A similar story holds for raising science requirements, with most of the effect being on non-STEM collegiate attainment, albeit with smaller point estimates.

A small increase in high school dropout, a potential concern for policy makers when considering raising graduation requirements, is found for both men and women when increasing requirements in mathematics, but not when raising them in science. The largest effect appears to be for the median male, for whom increasing the number of required math courses from 2 to 4 years is predicted to increase the probability of dropping out of high school by 1.2 percentage points. A lack of very strong effects for median students may not be very surprising, especially with regards to high school dropout. Dropping out of high school is a relatively rare event, and by restricting the analysis to the median student, we are necessarily looking at a group of students who, based only on these observables, would have a low probability of dropping out of high school.

However, there does exist heterogeneity on unobservables in this ‘median’ analysis, namely across the unobserved ability distribution which is independent of observable characteristics. To see how effects vary across unobservables, the same PRTEs are considered across the quantitative ability

Table 1.13: PRTE: Raising Grad. Reqs. - Effects on Final Outcomes*

	3 Years Required		**4 Years Required**	
	Male	*Female*	*Male*	*Female*
PRTE: Math Reqs.				
High School Dropout	.006	.006	.012	.005
	(.004)	(.003)	(.008)	(.007)
High School Diploma Only	−.026	−.028	−.07	−.08
	(.009)	(.008)	(.015)	(.014)
Non-STEM Major	.021	.025	.047	.067
	(.007)	(.007)	(.012)	(.012)
Math/Physical Science Major	.001	−.001	.009	.001
	(.005)	(.002)	(.008)	(.003)
Life Science Major	−.001	−.002	.001	.007
	(.002)	(.004)	(.004)	(.006)
PRTE: Science Reqs.				
High School Dropout	−.003	−.003	−.008	−.004
	(.002)	(.004)	(.004)	(.006)
High School Diploma Only	−.005	−.005	−.015	−.014
	(.002)	(.004)	(.004)	(.006)
Non-STEM Major	.005	.004	.013	.01
	(.002)	(.004)	(.004)	(.006)
Math/Physical Science Major	0	.002	.003	.002
	(.002)	(.004)	(.004)	(.006)
Life Science Major	.003	.003	.007	.007
	(.002)	(.004)	(.004)	(.006)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
* Effects relative to 2 years required
Standard errors obtained from 50 bootstrap re-samples

distribution. Effects are compared for median students with drawn quantitative ability that is 'low' (less than 1 standard deviation from the mean), 'medium' (between 1 standard deviation below the mean and 1 standard deviation above the mean), and 'high' (greater than 1 standard deviation above the mean), to determine whether there is heterogeneity in treatment effects across these unobservables. Indeed, Table 1.14 provides evidence that raising math requirements in high school from 2 to 4 years is more detrimental to lower ability students' high school graduation rates than higher ability students. A low ability male student, with median observable characteristics, is predicted to have a 3.4 percentage point higher probability of dropping out of high school if they are required to take 4 years of math rather than 2. For women, the effect on dropout is not as strong, though the estimated effects for both sexes are rather noisy and may be stronger or weaker than point estimates suggest.

Table 1.14: PRTE: Raising Grad. Reqs. from 2 to 4 Years - Final Outcomes Across Ability Distribution

	Men			**Women**		
	Low Quant.	*Med. Quant.*	*High Quant.*	*Low Quant.*	*Med. Quant.*	*High Quant.*
PRTE: Math Reqs.						
High School Dropout	.034	.011	0	.006	.007	−.002
	(.026)	(.007)	(.003)	(.024)	(.007)	(.005)
High School Diploma Only	−.09	−.073	−.044	−.077	−.077	−.091
	(.028)	(.017)	(.019)	(.025)	(.016)	(.019)
Non-STEM Major	.056	.055	.016	.071	.064	.073
	(.015)	(.013)	(.02)	(.013)	(.015)	(.019)
Math/Physical Science Major	0	.008	.019	0	.002	−.002
	(.004)	(.008)	(.02)	(.003)	(.003)	(.007)
Life Science Major	0	−.001	.009	0	.004	.021
	(.004)	(.004)	(.01)	(.005)	(.007)	(.013)
PRTE: Science Reqs.						
High School Dropout	−.028	−.005	0	−.013	−.003	−.003
	(.027)	(.005)	(.004)	(.027)	(.006)	(.006)
High School Diploma Only	.037	−.026	−.017	0	−.014	−.026
	(.029)	(.016)	(.018)	(.028)	(.014)	(.019)
Non-STEM Major	−.009	.02	.007	.013	.013	−.003
	(.012)	(.012)	(.02)	(.011)	(.013)	(.006)
Math/Physical Science Major	0	.007	−.009	0	.001	.005
	(.004)	(.007)	(.015)	(.002)	(.003)	(.006)
Life Science Major	0	.005	.019	0	.003	.026
	(.003)	(.004)	(.011)	(.004)	(.005)	(.016)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
Standard errors obtained from 50 bootstrap re-samples

Interestingly, raising graduation requirements in science appears to have the opposite effect on dropout rates for low ability students, with probability of dropout estimated to be 2.8 percentage points lower for men, and 1.3 percentage points lower for women when required to take 4 years of science rather than 2. A less pronounced effect on dropout for science requirements is not unexpected, as these requirements are less restrictive for students than with mathematics. This is because a student who is required to take 4 years of science has two subjects they can choose from, physical science and life science, when deciding how to satisfy them. However, it should again be noted that the estimated effects on dropout are rather noisy and could, in actuality, be positive.

The effects of raising graduation requirements on collegiate attainment also varies across the ability distribution and by gender. For men, when raising graduation requirements in math, most of the predicted increase in the probability of achieving a non-STEM bachelors degree is driven by students with ‘low’ and ‘medium’ quantitative abilities, while for women these effects are similar across all three quantitative ability categories. In addition, high ability men drive most of the

predicted increase in collegiate math/physical science attainment and high ability women make up the bulk of the increase in collegiate life science attainment.

Besides knowing how graduation requirements affect the probability of students ending up with different educational outcomes, it is important for policy makers to know the *types* of students who end up in these categories. For instance, a policy that influences a large number of low ability students to obtain math/physical science bachelors degrees may have a different social value than one that gets a small number of high ability students to obtain these degrees. Looking at how these effects vary over the ability distribution can give an idea of how ability sorting is affected by graduation requirements. Table 1.15 looks at this directly by displaying the effect of raising math and science graduation requirements on the average verbal and quantitative ability of students across educational outcomes (as measured in standard deviations from the population mean). While effects are imprecise, point estimates suggest that such policies mostly induce lower ability students to attain higher levels of education, resulting in the average verbal and quantitative abilities in these education groups to decrease. The exception is with life science, where estimates suggest moderate increases in the average ability of students who obtain a life science major. These results are consistent with estimates of the PRTE across the quantitative ability distribution, which indicate that graduation requirements simultaneously induce some lower ability students to attain higher education and others to drop out of high school.

1.4.6 Raising Graduation Requirements: Mechanisms

While the previous subsection explored the *total* effect of graduation requirements for median individuals, both in aggregate and across the unobserved ability distribution, the results do not reveal the mechanisms through which effects are operating. There are three primary ways that altering graduation requirements can affect the future outcomes of students who graduate high school:

1. By directly increasing the number of math or science classes with which a student graduates.

2. By indirectly increasing/decreasing the number of courses in *other* subjects a student chooses to take.

Table 1.15: PRTE: Raising Grad. Reqs. from 2 to 4 Years - Ability Sorting

	Men		**Women**	
	Verbal	*Quant.*	*Verbal*	*Quant.*
PRTE: Math Reqs.				
High School Dropout	$-.041$	$-.006$	$-.058$	$-.019$
	(.11)	(.088)	(.117)	(.102)
High School Diploma Only	$-.001$	.004	$-.052$	$-.047$
	(.023)	(.019)	(.02)	(.018)
Non-STEM Major	$-.081$	$-.103$	$-.047$	$-.058$
	(.022)	(.022)	(.019)	(.021)
Math/Physical Science Major	$-.031$	$-.042$	$-.002$	$-.03$
	(.046)	(.057)	(.118)	(.12)
Life Science Major	.011	.013	.03	.02
	(.055)	(.069)	(.036)	(.041)
PRTE: Science Reqs.				
High School Dropout	.055	.022	$-.034$	.01
	(.121)	(.094)	(.13)	(.11)
High School Diploma Only	$-.036$	$-.035$	$-.029$	$-.03$
	(.021)	(.019)	(.018)	(.018)
Non-STEM Major	$-.013$	$-.005$	$-.012$	$-.017$
	(.021)	(.025)	(.019)	(.019)
Math/Physical Science Major	$-.043$	$-.053$	.046	$-.005$
	(.035)	(.044)	(.102)	(.09)
Life Science Major	.018	.024	.06	.079
	(.049)	(.067)	(.038)	(.044)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
Standard errors obtained from 50 bootstrap re-samples

3. By altering the beliefs students have about their underlying abilities at the time of high school exit.

Tables 1.16 and 1.17 display the estimated PRTE, on course accumulation and beliefs, of increasing math or science requirements from 2 to 4 years for the median male and female student. Table 1.16 displays average estimates while Table 1.17 displays estimates across the quantitative ability distributions. Effects on beliefs are measured in terms of standard deviations from the population mean.

Increasing math requirements increases the number of years of mathematics taken for students across the ability distribution, though this effect diminishes for higher ability students. This reflects

the fact that these requirements are more binding for lower ability students who are less likely to take many years of math without being coerced. When more years of math are required, students are also likely to react by altering the number of physical science and life science courses they take. For physical science, point estimates suggest that effects differ by gender, with men taking more physical science courses when they are required to take more math, whereas women do not. The exception is with high ability women who do appear to take a bit more physical science when coerced into taking more mathematics. These effects are imprecisely estimated, however, so one cannot rule out a negative effect on physical science accumulation.

The effect of math requirements on life science course accumulation is stronger and more precise. Both men and women appear to substitute away from life science when required to take more math,

Table 1.16: PRTE: Raising Grad. Reqs. from 2 to 4 Years - State Variables

	Men	**Women**
PRTE: Math Reqs.		
Years of Mathematics	.491	.589
	(.058)	(.05)
Years of Physical Science	.044	.012
	(.052)	(.057)
Years of Life Science	−.269	−.229
	(.059)	(.058)
Expected Verbal Ability	.017	.007
	(.011)	(.009)
Expected Quant. Ability	.014	.007
	(.009)	(.007)
PRTE: Science Reqs.		
Years of Mathematics	.003	.005
	(.056)	(.05)
Years of Physical Science	.596	.560
	(.057)	(.06)
Years of Life Science	.35	.427
	(.059)	(.057)
Expected Verbal Ability	−.015	−.006
	(.01)	(.009)
Expected Quant. Ability	−.013	−.006
	(.008)	(.008)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
Standard errors obtained from 50 bootstrap re-samples

Table 1.17: PRTE: Raising Grad. Reqs. from 2 to 4 Years - State Variables - Across Ability Distribution

	Men			**Women**		
	Low Quant.	*Med. Quant.*	*High Quant.*	*Low Quant.*	*Med. Quant.*	*High Quant.*
PRTE: Math Reqs.						
Years of Mathematics	.682	.48	.388	.702	.589	.506
	(.079)	(.064)	(.054)	(.086)	(.054)	(.049)
Years of Physical Science	.042	.033	.074	.01	.006	.033
	(.08)	(.052)	(.056)	(.078)	(.06)	(.053)
Years of Life Science	$-.285$	$-.257$	$-.296$	$-.177$	$-.232$	$-.256$
	(.077)	(.06)	(.066)	(.077)	(.059)	(.063)
$E\{R_i\}$	.021	.009	.002	$-.011$	.008	.005
	(.027)	(.005)	(.002)	(.023)	(.005)	(.003)
$E\{Q_i\}$	.007	.007	.003	$-.007$	.007	.006
	(.015)	(.004)	(.003)	(.017)	(.004)	(.004)
PRTE: Science Reqs.						
Years of Mathematics	$-.022$	.005	.019	.044	$-.004$	.008
	(.066)	(.059)	(.053)	(.06)	(.051)	(.054)
Years of Physical Science	.767	.587	.511	.706	.548	.494
	(.085)	(.06)	(.064)	(.094)	(.06)	(.073)
Years of Life Science	.526	.333	.279	.568	.422	.342
	(.08)	(.06)	(.063)	(.073)	(.057)	(.066)
$E\{R_i\}$	$-.035$	$-.007$	.006	$-.017$	$-.003$	.004
	(.027)	(.004)	(.002)	(.023)	(.003)	(.003)
$E\{Q_i\}$	$-.02$	$-.009$	.008	$-.011$	$-.005$	.006
	(.015)	(.004)	(.003)	(.014)	(.003)	(.004)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
Standard errors obtained from 50 bootstrap re-samples

meaning that graduation requirements in mathematics ‘crowd out’ life science course-taking to an extent. For men, these effects do not differ very heavily across the ability distribution, whereas for women higher ability students substitute more heavily away from life science when they need to take more math. While there is enough statistical evidence to reject a hypothesis that higher math requirements induce students to take *more* life science, standard errors preclude us from saying, with a high level of certainty, that differences across the ability distribution truly exist.

For both men and women, increasing science requirements from 2 to 4 years induces greater accumulation of both physical science and life science courses. In addition, large effects on mathematics course-taking are not seen as a result of changes in science requirements. Interestingly, both genders appear to primarily increase their science coursetaking in physical science rather than life science, though women take a slightly more life science than men do.

Point estimates on graduation requirements' effects on beliefs are small, and we can rule out very large effects on beliefs of more than a few hundredths of a standard deviation even though estimated effects are not statistically different from 0. The evidence suggests that higher ability students see very little changes to their beliefs as a result of increasing STEM coursetaking requirements. This is likely because higher ability students tend to take more math and science regardless of graduation requirements, so any additional courses they are forced to take when requirements are raised do little to alter their information sets. Since lower ability students end up taking relatively *more* math and science when requirements go up, they react more to the additional information than their higher ability counterparts.

Overall, graduation requirements appear to have minor effects on beliefs, but sizeable effects on accumulation of math, physical science, and life science courses. This suggests that, unless small changes in beliefs have extraordinarily powerful effects on choices, effects of graduation requirements are driven primarily by changes to course-taking and not changes to beliefs. To better disentangle these mechanisms, average treatment effects (ATE) of course accumulation and belief alteration on post-secondary educational choices are estimated. Each simulated median individual, across all high school graduation requirements, is taken at the end of their secondary education. Then, holding everything else constant, their accumulated number of courses or their beliefs are altered to determine the resulting effect on final educational outcomes.

Table 1.18 presents the average treatment effect of increasing math, physical science, or life science course accumulation by 1 for median male and female students[16]. Holding everything else constant, increasing years of math by 1 has a positive effect on the probability of going to college and completing a non-STEM major, but little effect on anything else. Increasing years of physical science at the end of high school results in larger effects on non-STEM attainment than when increasing years of math, and also a sizeable effect on the probability of completing a math/physical science degree for both men and women. Raising life science course accumulation

[16]Students who have already taken 4 years of one of these courses do not have their accumulations altered

has a positive effect on the probability of completing a life science major, but also a negative effect on the probability of completing college or majoring in a non-STEM or math/physical science field.

Table 1.18: ATE: Increase Course by 1 at End of H.S.

	Men	**Women**
Years of Mathematics		
High School Dropout	0	0
	(0)	(0)
High School Diploma Only	$-.026$	$-.029$
	(.017)	(.016)
Non-STEM Major	.017	.022
	(.012)	(.014)
Math/Physical Science Major	.006	$-.002$
	(.009)	(.004)
Life Science Major	.004	.008
	(.007)	(.011)
Years of Physical Science		
High School Dropout	0	0
	(0)	(0)
High School Diploma Only	$-.065$	$-.055$
	(.017)	(.016)
Non-STEM Major	.046	.05
	(.012)	(.014)
Math/Physical Science Major	.024	.011
	(.009)	(.004)
Life Science Major	$-.005$	$-.007$
	(.007)	(.011)
Years of Life Science		
High School Dropout	0	0
	(0)	(0)
High School Diploma Only	.061	.059
	(.017)	(.016)
Non-STEM Major	$-.059$	$-.083$
	(.012)	(.014)
Math/Physical Science Major	$-.024$	$-.003$
	(.009)	(.004)
Life Science Major	.021	.027
	(.007)	(.011)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
Standard errors obtained from 50 bootstrap re-samples

It is important to clarify how these effects should be interpretted. This is not akin to estimating the effect of taking every median individual and increasing their course accumulation in a particular subject from x years to $x + 1$. Instead, we are estimating the effect of increasing a students' course accumulation to be 1 year higher than they would have chosen otherwise, holding *everything else constant*. In this light, some of these effects make more sense. For instance, the average student chooses to take 3 or 4 years of math even if they are not required to, while the same cannot be said about physical science and life science. If returns to coursetaking are diminishing, we would expect to see smaller effects of additional math coursetaking on collegiate STEM attainment than with physical science and life science.

Perhaps the most perplexing result from this analysis comes from the estimated effects of additional life science coursetaking, which indicates that, while more exposure to life science increases the likelihood of completing a life science degree if one goes on to complete college, it also decreases the overall probability of attaining a college degree. This may be explained by the types of courses that are categorized as life science, which can be found in Appendix A. Among these are some that might be considered to be 'career and technical education' courses and/or have a high value in the labor market, such as those in the 'allied health courses' designation. These types of classes may increase the utility of directly entering the labor force instead of going to college. In addition, within the 'high school diploma only' educational category, are students who never received any college education, those who *did* go to college but did not finish, and those who went to college but only completed an associates degree. If life science courses are valuable in combination with the types of careers that associates degrees complement, we would expect to see this type of pattern in the average treatment effect.

Finally, Table 1.19 presents the average treatment effect of increasing beliefs about verbal or quantitative ability by 0.25 standard deviations at the end of high school[17]. Beliefs appear to be very important determinants of post-secondary choices, with each ability factor affects educational

[17] Students for whom increasing beliefs by 0.25 standard deviations would bring their belief values out of sample do not have their beliefs altered in this analysis

outcomes differently. Both have positive effects on overall collegiate attainment, with verbal ability increasing the probability of students completing non-STEM majors and quantitative ability increasing the probability of students completing math/physical science majors. In addition, women see an uptick in the probability of completing life science majors when they believe themselves to be better quantitatively.

Table 1.19: ATE: Increase Beliefs by 0.25 SDs at End of H.S.

	Men	**Women**
Expected Verbal Ability		
High School Dropout	0	0
	(0)	(0)
High School Diploma Only	$-.063$	$-.05$
	(.017)	(.016)
Non-STEM Major	.096	.066
	(.012)	(.014)
Math/Physical Science Major	$-.043$	$-.004$
	(.009)	(.004)
Life Science Major	.009	$-.012$
	(.007)	(.011)
Expected Quant. Ability		
High School Dropout	0	0
	(0)	(0)
High School Diploma Only	$-.015$	$-.01$
	(.017)	(.016)
Non-STEM Major	$-.079$	$-.027$
	(.012)	(.014)
Math/Physical Science Major	.09	.012
	(.009)	(.004)
Life Science Major	.004	.025
	(.007)	(.011)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
Standard errors obtained from 50 bootstrap re-samples

1.5 Structural Approach

This section describes the set up and estimation of the full structural model. Estimating the full model allows for the analysis of the welfare effects of high school curriculum policies. However,

in order for estimation to be achievable in a computationally feasible manner, additional structure must be imposed on the general model presented in Section (1.1.2).

1.5.1 Adding More Structure

The model laid out in Section (1.1.2) did not assume any functional forms for the flow utilities, however we need to make assumptions about these in order to estimate the model. It is assumed that all flow utilities are additively separable between observable characteristics and unobserved shocks. The utility of being in the labor market is specified as only depending on the expected log-wage received:

$$\begin{aligned} E\{U^L_{it}(W_{it}, \epsilon^L_{it})|I_{it}\} &= E\{u^L_{it}\} + \epsilon^L_{it} \\ &= \delta_W E\{\ln[W_{it}|I_{it}]\} + \epsilon^L_{it} \\ \epsilon^L_{it} &\sim \text{Type 1 Extreme Value} \end{aligned} \tag{1.34}$$

while the flow utility of choosing to go to college and major in field m, is the difference between one's psychic benefit and expected psychic cost of attending college and majoring in field m[18]:

$$\begin{aligned} E\{U^m_{it}(X_i, Z_{it}, \tau_i, \mathbf{A}_i, \epsilon^m_{it})|I_{it}\} &= u^m_{it}(X_i, \tau_i) - E\{c^m_{it}(\mathbf{A}_i, Z_{it})|I_{it}\} + \epsilon^m_{it} \\ &= \delta^m_0 + \delta^m_1 X_i + \delta^m_2 \tau_i - \delta^m_3 \tilde{R}_{it} - \delta^m_4 \tilde{Q}_{it} - \delta^m_5 Z_{it} + \epsilon^m_{it} \\ \epsilon^m_{it} &\sim \text{Type 1 Extreme Value} \end{aligned} \tag{1.35}$$

Here, utility is specifed to be linear in parameters. The psychic benefit of a particular major is a function only of fixed characteristics known to the student before beginning high school. This includes both observable (i.e. demographics, parental background, non-cognitive proxies) and unobservable (i.e. unobserved heterogeneity 'type' τ_i) fixed characteristics. The expected psychic cost of majoring in field m in time t, is a function of one's beliefs about one's unobserved verbal and

[18]Framing utility in terms of psychic costs and psychic benefits is similar to the utility interpretation in Stange (2012)

quantitative ability, along with other time varying observables that depend on previous educational choices (i.e. accumulated years of high school math, physical science, and life science).

Flow utility in high school has a similar specification to college. Every course, c, that can be taken in a particular year has a utility associated with it that, again, depends on the psychic benefits and expected psychic costs of taking that course:

$$\begin{aligned} E\{U_{it}^c(X_i, Z_{it}, \tau_i, \mathbf{A}_i)|I_{it}\} &= u_{it}^c(X_i, \tau_i) - E\{c_{it}^c(\mathbf{A}_i, Z_{it})|I_{it}\} \\ &= \delta_0^c + \delta_1^c X_i + \delta_2^c \tau_i - \delta_3^c \tilde{R}_{it} - \delta_4^c \tilde{Q}_{it} - \delta_5^c Z_{it} \end{aligned} \tag{1.36}$$

However, a student's decision in high school is not over which course to take, but which *bundle* of courses, b, to take. The utility of a bundle is assumed to simply be the sum of the utilities of the individual courses that it includes, plus an idiosyncratic error term:

$$\begin{gathered} E\{U_{it}^b(X_i, Z_{it}, \tau_i, \mathbf{A}_i)|I_{it}\} = \sum_{c \in b} [u_{it}^c(X_i, \tau_i) - E\{c_{it}^c(\mathbf{A}_i, Z_{it})|I_{it}\}] + \epsilon_{it}^b \\ \epsilon_{it}^b \sim \text{Type 1 Extreme Value} \end{gathered} \tag{1.37}$$

1.5.2 Estimation

The goal of full structural estimation is to estimate the parameters of the utility functions specified above. The total value of any decision d made in time t can be written, using Bellman's equation, as:

$$v_{it}^d = U_{it}^d(X_i, Z_{it}, \tau_i, \tilde{R}_{it}, \tilde{Q}_{it}) + \beta E\{V(X_i, Z_{it+1}, \tau_i, \tilde{R}_{it+1}, \tilde{Q}_{it+1})_{it+1}|d_{it} = d\} + \epsilon_{it}^d \tag{1.38}$$

The discount factor is set to 0.9 per period throughout estimation[19]. The assumption that idiosyncratic errors are distributed Type 1 Extreme Value means that this discrete choice problem admits a

[19]A value for this parameter is usually set by the researcher prior to estimation. Arcidiacono et al. (2016) also chooses a discount factor of 0.9.

standard multinomial logit form, where the probability of making any decision in time t is:

$$\begin{aligned} P_{it}^{d}(X_i, Z_{it}, \tau_i, \tilde{R}_{it}, \tilde{Q}_{it}) &= \frac{\exp(v_{it}^{d})}{\sum_{d'} \exp(v_{it}^{d'})} \\ &= \frac{\exp(U_{it}^{d}(X_i, Z_{it}, \tau_i, \tilde{R}_{it}, \tilde{Q}_{it}) + \beta E\{V(X_i, Z_{it+1}, \tau_i, \tilde{R}_{it+1}, \tilde{Q}_{it+1})_{it+1} | d_{it} = d\})}{\sum_{d'} \exp(U_{it}^{d'}(X_i, Z_{it}, \tau_i, \tilde{R}_{it}, \tilde{Q}_{it}) + \beta E\{V(X_i, Z_{it+1}, \tau_i, \tilde{R}_{it+1}, \tilde{Q}_{it+1})_{it+1} | d_{it} = d'\})} \end{aligned} \tag{1.39}$$

and the structural utility parameters can be estimated by maximizing the following likelihood function:

$$L^{u} = \prod_{i} \sum_{\tau} \pi_{\tilde{x}}^{\tau} \prod_{t} \prod_{d} [P_{it}^{d}(X_i, Z_{it}, \tau_i, \tilde{R}_{it}, \tilde{Q}_{it})]^{\mathbb{1}\{d_{it}=d\}} \tag{1.40}$$

There is nothing conceptually different between estimating the parameters of this dynamic multinomial logit and estimating the the quasi-structural approximations to the decision rules described in Section 1.3.3. One could apply the EM algorithm in this scenario to restores additive separability and ease estimation, as is done in the grade and CCP estimation. However, unlike those scenarios, computational issues arise when trying to maximize the expected log likelihood function. This is because the index function of the dynamic logit contains the future value term $E\{V_{it+1} | d_{it} = d\}$, which is notoriously difficult to evaluate and which must be calculated potentially hundreds of times in the estimation procedure for every individual in every time period observed in the sample. With the number of choices and state variables in the model, a full information maximum likelihood approach to estimating this model is infeasible.

However, an important simplification of the dynamic decision problem can be applied given three key assumptions (Arcidiacono and Ellickson (2011)):

1. Utility errors are distributed Type 1 Extreme Value.
2. Utility errors are additively separable from the observable portion of utility.
3. Labor market entrance is an absorbing state.

The first two assumptions allow expectation of $V_{it+1}|d_{it}=d$ over future ϵ_{it}^{d} shocks to be re-written as a closed-form expression[20]:

$$E\{V_{it+1}|d_{it}=d\} = E\left\{\ln\left[\sum_{d_{it+1}}\exp(v_{it+1}^{d_{it+1}})\right]\right\}+\gamma \tag{1.41}$$

where the remaining expectation is taken over future shocks to the student's information set, and γ is the Euler-Mascheroni constant.

Now consider the decision $d_{it+1}=L$, which is the choice to enter the labor market. This is available to agents at any point in the decision process (except the first period, grade 9). The inside of the natural logarithm in equation 1.41 can be multiplied by $\frac{\exp(v_{it+1}^{L})}{\exp(v_{it+1}^{L})}$ and re-arranged to get the following expression:

$$\begin{aligned} E\left\{\ln\left[\sum_{d_{it+1}}\exp(v_{it+1}^{d_{it+1}})\right]\right\}+\gamma &= E\left\{\ln\left[\exp(v_{it+1}^{L})\left(\frac{\exp(v_{it+1}^{L})}{\sum_{d_{it+1}}\exp(v_{it}^{d_{it+1}})}\right)^{-1}\right]\right\}+\gamma \\ &= E\left\{-\ln\left[\left(\frac{\exp(v_{it+1}^{L})}{\sum_{d_{it+1}}\exp(v_{it}^{d_{it+1}})}\right)\right]+v_{it}^{L}\right\}+\gamma \end{aligned} \tag{1.42}$$

Because a Type 1 Extreme Value error distribution has been assumed, the expression in the natural logarithm is equivalent to the probability that any individual chooses to enter the labor market in $t+1$, conditional on all observable and unobservable characteristics that are known to the individual at that time. This means Bellman's equation can be rewritten as:

$$v_{it}^{d} = U_{it}^{d}+E\left\{-\ln\left[\Pr\left(d_{it+1}=L|X_i,Z_{it+1},\tau_i,\tilde{R}_{it+1},\tilde{Q}_{it+1},d_{it}=d\right)\right]+v_{it+1}^{L})\right\}+\gamma+\epsilon_{it}^{d} \tag{1.43}$$

[20]The arguments of the V_{it+1} function are suppressed here for notational clarity

Suppose we think about the value of making any decision $d_{it} = \alpha$ relative to the value of dropping out of school $d_{it} = L$. Using the expression that was just derived, this can be written as:

$$
\begin{aligned}
&v^*_{it}(\alpha) + \epsilon^*_{it}(\alpha) = v^\alpha_{it} - v^L_{it} + (\epsilon^\alpha_{it} - \epsilon^L_{it}) \\
&= U^\alpha_{it} + \delta_W \left(\beta E \left\{v^L_{it+1} | d_{it} = \alpha\right\} - v^L_{it}\right) \\
&+ \beta E \left\{- \ln \left[\Pr \left(d_{it+1} = \alpha | X_i, Z_{it+1}, \tau_i, \tilde{R}_{it+1}, \tilde{Q}_{it+1}, d_{it} = \alpha\right)\right]\right\} + \beta\gamma + (\epsilon^\alpha_{it} - \epsilon^L_{it})
\end{aligned}
\tag{1.44}
$$

where v^L_{it} is the discounted expected lifetime utility of entering the labor market in period t, which can be easily calculated from equation (1.34). This final equation implies a particular interpretation. The relative value of continuing schooling, and choosing option α in particular, is equal to the flow utility of that schooling option (U^α_{it}) along with the expected gain in discounted life-time utility one could receive by delaying dropout until next period ($E \left\{v^L_{it+1}\right\} - v^L_{it}$). In addition, one has to consider the dynamic optimality of making the dropout decision at time t. To do this, they can look at the fraction of other people that choose to dropout with the same observable characteristics and beliefs, which is captured by the $E \left\{- \ln \left[\Pr \left(d_{it+1} = \alpha | X_i, Z_{it+1}, \tau_i, \tilde{R}_{it+1}, \tilde{Q}_{it+1}, d_{it} = \alpha\right)\right]\right\} + \gamma$ term. This goes to γ as probability of dropping out goes to 1, and to negative infinity as the probability of dropping out goes to 0.

This way of writing the dynamic discrete choice problem was developed in Hotz and Miller (1993) and expanded upon in Arcidiacono and Miller (2011) to incoporate unobserved heterogeneity[21]. Estimation of the structural parameters proceeds by first estimating the quasi-structural model as in Section 1.3.3 which provides approximations of the CCPs and estimates of the unobserved 'type' distribution. The future value of *any* decision can then be calculated by using the wage parameters and estimated CCP approximations and plugging them into the index function of the dynamic logit likelihood in equation (1.40). Estimation of the structural utility parameters is then accomplished by treating these future values as fixed in a standard conditional multinomial logit estimation, using the individual level unobserved 'type' probabilities from the quasi-structural esti-

[21] See Arcidiacono and Ellickson (2011) for an easily accessible overview of applying CCP methods in the estimation of dynamic discrete choice models with unobserved heterogeneity

mation as probability weights. Utilizing this approach greatly reduces the computational complexity of the problem and makes estimation exponentially easier to implement.

1.6 Structural Approach: Results

This section presents the results from the full structural estimation procedure described in the previous section.

1.6.1 Utility Parameter Estimates

Tables 1.20 and 1.21 display the estimated structural utility parameters in high school and college respectively. 5 unobserved 'types' of individuals are identified who, even though they may look the same on observable characteristics, have different psychic benefits of educational options. The distribution of this unobserved heterogeneity, including how it varies between men and women and across high school graduation requirements, can be found in Table 1.22.

The marginal utility of log wages, δ_W in the description of the explicit structural model, is estimated to be 1.787. As the estimates stand, their scale has no obvious meaning. If the value of entering the labor market was estimated in terms of level wages, one could normalized the scale of the utility parameters to be in terms of U.S. dollars ($100,000$s of 2020 U.S. dollars in this paper). However, it is common to model labor market entry in terms of log wages rather than level[22]. In addition, attempts at modeling labor market entry in terms of level wages yielded odd estimates of the marginal utility of income[23]. This may be because most of the model takes place in high school, where expected wages do not change very much from year to year even conditional on different choices. By using log wages instead, a non-linearity is introduced which changes the year to year valuation of labor market entry. The estimated marginal utility of log wages can be used to place utility estimates in the log wage scale, however seeing as this does not really add to their

[22] Arcidiacono (2004) and Arcidiacono et al. (2016) are two examples of dynamic structural models of educational choice using log of earnings in the utility function of work

[23] Point estimates of the marginal utility of level income were 0 in the full sample, with the parameter restricted to be non-negative in estimation. Relative estimates of the other utility parameters did not appear to change much between a model with level wages and a model with log wages

Table 1.20: H.S. Flow Utility Parameters

	Math	Physical Science	Life Science	non-STEM
Constant (Type 1)	.783	−.282	−1.245	−6.193
	(.672)	(.322)	(.255)	(1.442)
Constant (Type 2)	.3	−1.006	−.663	−6.595
	(.644)	(.326)	(.24)	(1.435)
Constant (Type 3)	.444	−1.026	−.437	−6.049
	(.677)	(.309)	(.249)	(1.476)
Constant (Type 4)	.859	−1.067	−.071	−5.747
	(.702)	(.313)	(.254)	(1.519)
Constant (Type 5)	1.039	−.361	−.667	−6.609
	(.684)	(.329)	(.236)	(1.457)
Male	−.324	.288	−.007	.618
	(.368)	(.141)	(.156)	(.331)
Black	.177	−.023	−.106	−.87
	(.095)	(.042)	(.041)	(.325)
Hispanic	.207	−.094	−.159	−.507
	(.066)	(.035)	(.04)	(.145)
Parent Has College Degree	.354	.125	−.035	.006
	(.042)	(.032)	(.029)	(.109)
Parent Works in STEM	.081	.055	−.031	–
	(.042)	(.03)	(.03)	
Student Often Completes Homework	.145	.153	−.107	.651
	(.053)	(.033)	(.04)	(.12)
Student is Often Disruptive	−.128	−.174	.059	.091
	(.064)	(.038)	(.03)	(.129)
Student is Often Tardy	−.019	−.089	.059	−.488
	(.058)	(.041)	(.037)	(.133)
Student is Exceptionally Passive or Withdrawn	.055	−.054	−.023	−.23
	(.071)	(.04)	(.039)	(.148)
Student Relates well to Others	.086	−.009	−.094	.816
	(.065)	(.042)	(.046)	(.212)
Number of Courses Offered at HS**	.434	.206	.493	–
	(.089)	(.034)	(.045)	
Number of Courses Offered at HS Sq.**	−.075	−.04	−.085	–
	(.021)	(.008)	(.012)	
Expected Verbal Ability (Men)	–	–	−.237	1.431
			(.144)	(.196)
Expected Verbal Ability (Women)	–	–	−.487	1.833
			(.146)	(.174)
Expected Quantitative Ability (Men)	.43	.365	.087	–
	(.124)	(.033)	(.12)	
Expected Quantitative Ability (Women)	.524	.394	.225	–
	(.098)	(.037)	(.124)	

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
*All sample sizes rounded to nearest 10
**Per 10 courses offered
Standard errors obtained from 50 bootstrap re-samples

Table 1.21: College Flow Utility Parameters

	Math/Physical Science	Life Science	non-STEM
Constant (Type 1)	−23.332	−18.075	−13.42
	(3.966)	(3.403)	(2.612)
Constant (Type 2)	−23.335	−17.537	−11.643
	(4.167)	(3.425)	(2.564)
Constant (Type 3)	−19.954	−16.501	−9.847
	(3.929)	(3.368)	(2.58)
Constant (Type 4)	−18.111	−16.189	−10.431
	(4.004)	(3.378)	(2.626)
Constant (Type 5)	−21.719	−15.074	−11.693
	(4.076)	(3.328)	(2.619)
Male	1.27	−.151	.065
	(.337)	(.255)	(.17)
Black	−1.826	−.831	−1.166
	(.479)	(.344)	(.292)
Hispanic	−1.231	−1.19	−1.189
	(.288)	(.268)	(.159)
Parent Has College Degree	1.649	1.382	1.493
	(.153)	(.165)	(.074)
Parent Works in STEM	.174	.495	–
	(.164)	(.151)	
Student Often Completes Homework	2.132	2.109	1.291
	(.321)	(.25)	(.132)
Student is Often Disruptive	−.312	−.546	−.169
	(.309)	(.275)	(.12)
Student is Often Tardy	−1.111	−1.086	−.548
	(.382)	(.287)	(.14)
Student is Exceptionally Passive or Withdrawn	.17	−.963	−.462
	(.334)	(.318)	(.165)
Student Relates well to Others	1.022	.735	1.04
	(.383)	(.375)	(.249)
Years of Math Taken in H.S.	1.201	.925	.74
	(.278)	(.202)	(.1)
Years of Phys. Sci. Taken in H.S.	.817	.235	.515
	(.136)	(.101)	(.075)
Years of Life Sci. Taken in H.S.	−.934	.3	−.569
	(.151)	(.1)	(.076)
Expected Verbal Ability (Men)	–	1.269	2.886
		(.552)	(.319)
Expected Verbal Ability (Women)	–	2.268	2.719
		(.757)	(.314)
Expected Quantitative Ability (Men)	2.546	1.204	−.549
	(.374)	(.426)	(.243)
Expected Quantitative Ability (Women)	3.429	.683	−.384
	(.212)	(.537)	(.221)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
*All sample sizes rounded to nearest 10
**Per 10 courses offered
Standard errors obtained from 50 bootstrap re-samples

Table 1.22: Unobserved Heterogeneity

	Type 1	Type 2	Type 3	Type 4	Type 5
Males					
H.S. Req. 2 or Fewer	.371	.164	.282	.124	.059
	(.035)	(.021)	(.027)	(.022)	(.016)
H.S. Req. 3	.324	.111	.327	.123	.115
	(.015)	(.013)	(.016)	(.012)	(.012)
H.S. Req. 4	.35	.069	.373	.147	.061
	(.036)	(.024)	(.043)	(.035)	(.019)
Females					
H.S. Req. 2 or Fewer	.356	.131	.340	.075	.099
	(.029)	(.021)	(.026)	(.023)	(.025)
H.S. Req. 3	.294	.137	.311	.093	.165
	(.019)	(.016)	(.016)	(.012)	(.014)
H.S. Req. 4	.240	.232	.215	.183	.129
	(.035)	(.033)	(.034)	(.036)	(.027)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
"H.S. Req. n" is defined as being in a high school where the maximum math or science requirement is n.
Standard errors obtained from 50 bootstrap re-samples

interpretability, parameter estimates are kept in their unnormalized scale. Differences in magnitudes of the preference parameters can still be used to compare the importance of different factors in students' utility valuations of educational options.

The relative value of utility estimates elucidate gender differences in the valuation of different educational pathways, even conditional on expected ability and other important individual characteristics. In high school, men value taking physical science and non-STEM more than women, whereas women value taking math and life science more (though only the effect of sex on physical science utility is statistically different from 0). In college, these gender differences are slightly different. Holding everything else constant, men place a higher value on majoring in math or physical science, whereas women value majoring in life science more than men. Men are estimated to prefer non-STEM majors slightly more than women, but this difference is not statistically significant.

Parental background is also shown to increase students' valuations of taking math and physical science courses in high school. Having a parent who has a college degree or who works in a STEM field has a positive effect on the utility of taking math or physical science. However, both of these factors slighly lower the utility of taking life science, and has a noisy small effect on the utility of

taking non-STEM. In college, both of these parental characteristics positively influence the utility of going to college (though the effect of parental STEM employment is not statistically different from 0). Parental collegiate attainment is most potent on the utility of majoring in a math/physical science field, whereas a parent's employment status in STEM is most important in life science.

The variables used to proxy for students' non-cognitive abilities have stronger effects on the utility of college-going decisions than the utility of course choices in high school. The most important non-cognitive proxies in high school are whether a student often completes their homework and whether a student is disruptive in class. A student's tardiness, passiveness, and relatability are also estimated to be strongly important in the decision of whether to take non-STEM. In college, tendency to complete homework assignments and one's relatability have the strongest positive effects on the utility majoring in any field, whereas one's tendency to be tardy has the strongest negative effect. Passiveness has negative effects on the utility of majoring in life science or non-STEM, but a slight positive effect on the utility of majoring in math/physical science.

Accumulated years of courses taken in high school are allowed to affect the utility of different majors. Every additional year of math or physical science has a large effect on the value of going to college and majoring in any subject, with math is being most important in math/physical science and life science, and physical science being most important in math/physical science and non-STEM. Years of life science have an unanticipated effect on the utility of majoring in either math/physical science or non-STEM, where the estimated coefficient is negative and statistically different from 0 at the 5% level. This, again, may be the result of career and technical education courses being included in the high school life science categorization as discussed previously. A more detailed understanding of these seemingly paradoxical relationships between life science coursetaking in high school and some future educational outcomes should be considered in future research.

In the flow utility specifications, verbal and quantitative ability beliefs are allowed to enter the same equations in which they enter in the grade estimation. The coefficients on expected verbal and quantitative ability represent the marginal utility one receives from a student believing themselves to be 1 unit higher in the ability distribution. Quantitative ability is most important in math and

physical science. There is evidence that men value quantitative ability less strongly in life science than women, with the point estimate on quantitative ability beliefs for women being 258% larger for women. In addition, verbal ability beliefs have a strong positive influence on the utility of taking non-STEM in high school, but a negative effect on taking life science. This negative effect on life science is not statistically different from 0 for men, but is significant for women.

In college, beliefs appear to have a stronger impact on choices than in high school. A one unit increase in quantitative or verbal ability has a positive effect on the utility of majoring in math/physical science or life science. Verbal ability beliefs have a strong positive impact on the utility of majoring in non-STEM, however quantitative beliefs have small negative effects on the utility of choosing such a major. These negative effects are noisy, and values very close to 0 cannot be ruled out.

1.6.2 The Welfare Effects of Graduation Requirements

An important question to ask is "what is the effect on student welfare of stricter high school graduation requirements?" Because these academic standards restrict students' choice sets, they necessarily must have a negative impact on total student welfare. However, if these are small, then the social benefits of the additional educational attainment resulting from these requirements could outweigh the harm to individual welfare.

To calculate willingness to pay, the median male and female student are taken and simulated $1,000$ times in the same fashion as described in Section 1.3. For each simulated individual, in both grades 9 and 10, the CCPs and structural flow utility parameter estimates are used to calculate the value of every possible decision that can be made. Total expected lifetime utility, before any idiosyncratic shocks are realized, is then calculated using the standard formula[24]:

$$V_{it} = E\{\max(v_{it}^1, v_{it}^2, ..., v_{it}^9)\} = \ln\left[\sum_{d=1}^{9} exp(v_{it}^d)\right] + \gamma \tag{1.45}$$

[24] See Train (2009) for an overview the properties of discrete choice models

where v_{it}^d is the value of making choice d at time t and γ is the Euler-Mascheroni constant. The welfare effects of increasing graduation requirements at time t is the percentage change in V_{it} for individual i when 2 years are required to graduate versus when 3 or 4 years are required.

Table 1.23 displays average percentage change in expected lifetime utility when increasing graduation requirements from 2 years of math/science to 3 and 4 years of math/science. In grade 9, when no information has been revealed through grades and all students have the same beliefs about their abilities, the median male experiences a reduction in expected lifetime utility of 3.1% and 4.8% when raising graduation requirements in mathematics from 2 to 3 and 2 to 4 years respectively. Women experience similar reductions, albeit at a smaller estimated magnitude. In grade 10, after students' have received their first round of high school grades (and according to the results from Section 1.4.4, around 80% of uncertainty about abilities has been resolved), effects of graduation requirements on expected lifetime utility slightly decrease.

Table 1.23: Percentage Change in Expected Lifetime Utility from Graduation Requirements - Median Student

	3 Years Required		**4 Years Required**	
	Male	*Female*	*Male*	*Female*
Math Requirements				
Grade 9	$-.031$	$-.022$	$-.048$	$-.04$
	(4.743)	(1.724)	(7.04)	(3.346)
Grade 10	$-.024$	$-.019$	$-.03$	$-.026$
	(.608)	(.872)	(.865)	(1.417)
Science Requirements				
Grade 9	.01	.006	.028	.021
	(1.281)	(1.617)	(2.614)	(4.455)
Grade 10	.013	.009	.032	.022
	(.341)	(1.064)	(.692)	(1.376)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
Standard errors obtained from 50 bootstrap re-samples

Welfare effects of science requirements are a little different. First, estimates suggest that students experience an *increase* in expected lifetime utility when science requirements are raised. Such a result is actually contradictory to the assumptions of the model and is likely an artifact of the fact that the CCPs that determine the future value of any decision are approximate instead of exact.

Because these estimated welfare effects are so small, even tiny errors in the approximation of CCPs can lead to the estimates being, on average, greater than or less than 0.

Table 1.24 displays estimates of changes in welfare for students who are of low quantitative ability (1 standard deviation below the mean), medium quantitative ability (between 1 standard deviation below the mean and 1 standard deviation above the mean), and high quantitative ability (1 standard deviation above the mean). Suprisingly, these welfare measures do not vary heavily across the distribution of abilities after some uncertainty has been resolved. The estimates suggest that increased math requirements *may* be more burdensome on high ability students after an information shock. This could be because these students are, on average, likely to believe themselves to be higher in the ability distribution at this point than their lower ability peers. As such, they may not necessarily need to accumulate more math courses in high school in order to have the same expected cost of attending college, and the choice set restrictions only serve to force them into courses they do not need.

Table 1.24: Average Percentage Change in Expected Lifetime Utility in Grade 10 by Ability - Median Student

	3 Years Required		**4 Years Required**	
	Male	*Female*	*Male*	*Female*
Math Requirements				
Low Quant.	$-.017$	$-.016$	$-.023$	$-.022$
	(.718)	(.614)	(.883)	(.696)
Med. Quant.	$-.024$	$-.019$	$-.031$	$-.027$
	(.785)	(1.239)	(1.143)	(2.054)
High Quant.	$-.029$	$-.022$	$-.032$	$-.027$
	(.608)	(.772)	(.865)	(1.354)
Science Requirements				
Low Quant.	.007	.005	.017	.012
	(1.206)	(.359)	(1.016)	(.618)
Med. Quant.	.01	.006	.028	.02
	(.377)	(.564)	(.85)	(1.197)
High Quant.	.019	.013	.05	.032
	(.404)	(5.454)	(.986)	(5.32)

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)
Standard errors obtained from 50 bootstrap re-samples

1.7 Conclusion

When making their college-going decisions, students place a high value on both the amount of human capital they have accumulated in different STEM subjects and the beliefs they have about their academic aptitudes. Even without being coerced, most students choose to complete 3 or 4 years of mathematics in pursuit of their high school diplomas, and at least 1 to 2 years of physical or life science. In addition, by the end of high school students have a relatively good sense of where they lie in the ability distribution. This is likely because abilities utilized in different courses are highly correlated, allowing students to use signals obtained from their non-STEM courses to gather precise information about their skills in STEM subjects. This means that policies aimed at making the information contained in STEM courses more precise are likely to have a small impact on student outcomes.

Raising graduation requirements in mathematics appears to have positive effects on the probability that a median male or female student goes to college and majors in a non-STEM field. Increasing these requirements from 2 to 4 years is estimated to increase the probability of the median male student completing one of these degrees by 4.7 percentage points, and the median female by 6.7 percentage points. Changing science requirements in the same way also has a positive effect on non-STEM major completion, albeit at a smaller magnitude. Both of these policy changes are predicted to have little impact STEM major completion.

It is the alterations that such policies make to students' high school course-taking portfolios that change their final decisions, rather than alterations to their beliefs about their own abilities. Raising graduation requirements in mathematics increases the amount of math that the median student takes and reduces their accumulation of life science courses, while stricter curriculum requirements in science increase both physical science and life science course-taking without any large effects on years taken in mathematics.

However, there is suggestive evidence that these effects vary across the ability distribution. Point estimates suggest that increasing graduation requirements in mathematics from 2 to 4 years for

low ability median male students increases their probability of dropping out by (an imprecise) 3.4 percentage points. For high ability students in this same category, effects on high school dropout are precisely estimated to be about 0, with potentially relatively large positive effects on their probability of attaining a bachelors degree in STEM.

Ultimately, the effects of graduation requirements may be burdensome for some students while also improving collegiate outcomes for others. Standard errors on treatment effect estimates allow us to conclude that graduation requirements in mathematics have a positive impact on non-STEM major completion, but are unlikely to increase STEM major completion by a very large amounts. Even so, their welfare effects *may* not be overtly large, with estimates of the impact of graduation requirements on high school dropout and student welfare not ruling out modest negative effects. It is therefore within the realm of possibility that the social benefits of the additional college degrees obtained by imposing these policies could outweigh negative effects on students.

Appendices

Appendix A. Data Cleaning Details

This appendix describes, in detail, the procedures and assumptions I use to clean the ELS data and get it ready for estimation.

Demographic Characteristics

I begin with the full Educational Longitudinal Study of 2002 data consisting of $16,200$ individuals[25]. I remove anyone in the data who was not in 10^{th} grade in 2002 which drops the sample size to $16,020$.

From here, I collect the demographic covariates that I am interested in controlling for in the analysis:

- Whether the individual is male
- Whether the individual is black
- Whether the individual is hispanic

No one in the data set is missing any values for these characteristics, so the sample size remains the same.

Parental Education and Employment

There are two characteristics of students' parents I control for:

- Whether at least one of the student's parents has a college education
- Whether at least one of the student's parents works in a STEM field

I categorize a student as having a parent with a college degree if, in the first wave of the survey, either the mother or the father of the student reports having a 4-year college degree or higher. To

[25]All sample sizes are rounded to the nearest 10 in order to comply with NCES security standards

categorize the student as having a parent working in a STEM field, I utilize a variable that lists whether the mother or the father of the student is working in a STEM occupation as of the second wave of the ELS survey. Removing students who are missing these values drops the sample size to $11,520$.

Non-Cognitive Skill Proxies

An important factor in educational choices and outcomes is not only students' cognitive abilities, but also their non-cognitive skills. As the importance of these skills continue to be emphasized in the literature, more surveys are beginning to administer tests intended to measure these comprehensively. Unfortunately, the ELS did not administer any evaluation to its sample participants intended to measure these non-cognitive abilities. However, because of their likely importance in explaining students choices and outcomes, I choose to proxy for these using a approach similar to Lleras (2008) who uses the National Educational Longitudinal Study of 1988 (NELS) to explore the role of non-cognitive abilities on future outcomes.

The basic idea is to use information gathered from students' teachers on their perceptions of a student's behavior and sociability in the classroom as a proxy for different non-cognitive factors. In this study I utilize 5 questions that were asked to students' 10^{th} grade English and mathematics teachers:

1. "How often does this student complete homework assignments for your class?"
    - Possible responses: 'Never', 'Rarely', 'Some of the time', 'Most of the time', 'All of the time'
2. "How often is this student tardy in your class?"
    - Possible responses: 'Never', 'Rarely', 'Some of the time', 'Most of the time', 'All of the time'
3. "How often is this student disruptive in your class?"

    - Possible responses: ‘Never’, ‘Rarely’, ‘Some of the time’, ‘Most of the time’, ‘All of the time’

4. “Is this student exceptionally passive or withdrawn in your class?”

    - Possible responses: ‘Yes’, ‘No’

5. “Does this student seem to relate well to other students in your class?”

    - Possible responses: ‘Yes’, ‘No’

I recoded each of these responses into a binary variable, with the top two responses equal to 1 and the bottom three equal to 0 (except for the two questions that were already dichotomized). Responses from the student’s math teacher were used, unless that teacher’s response is missing (or the teacher responded ‘I don’t know’), in which case the response from the English teacher was used if available. Students with neither response available are dropped from the sample, bringing the sample size to $9,530$.

Cognitive Test Scores

Cognitive tests in reading and mathematics were administered in the first wave of sample collection, while a follow-up mathematics test was administered in the second wave of sample collection. I use these tests to help identify the distribution of unobserved verbal and quantitative ability in the student population. Very few remaining sample members are missing their first wave test scores, while quite a few more are missing their second wave mathematics test. Regardless, I do not exclude any students from the sample who are missing test scores, as in my model these scores are not observed by the student and only serve to help identify the ability distributions.

Wages

Because the ELS only collected data until most sample members were 26 years of age, there is not too much information on labor market outcomes such as wages. In the final survey wave,

respondents were asked about their earnings in 2011 (one year before the survey). I adjust this number from 2011 dollars to 2020 dollars, put it in terms of $\$100,000$s of dollars, and take its natural log. Anyone who is missing wage information is kept in the sample.

Final Outcomes

In my model, a student can end up, at age 26, being a high school dropout, a terminal high school graduate, a non-STEM major, a math/physical science major, or a life science major. Here I describe how I classify respondents into one of these 5 categories.

Recall that, in the model being estimated, students may drop out of high school in any year after 9^{th} grade and may not return. In order to determine which students to classify as high school dropouts, I make use of a variable constructed by the NCES that identifies 'likely dropouts'. These are either people within the sample who explicitly dropped out of high school at some point, or whose educational trajectories make the existence of at least one dropout spell during their high school experience very high. Any student that is classified as a likely dropout by the NCES is considered a dropout in my analysis.

Next, I determine who is to be considered a terminal high school graduate. I only consider individuals who report completing a bachelors degree or higher to be college graduates. Therefore, anyone who has only earned a 2 year degree or below, and was not a high school dropout, is classified as a terminal high school graduate.

Finally, I classify students by field of study. There are three major categories I consider: non-STEM, math/physical science, and life science. I assign majors to students based on the ELS determination of the student's field of study for their first known bachelors degree using the 2 digit Classification of Instructional Programs (CIP) codes[26] NCES (2000a):

- Life Science: Agriculture/operations/related sciences (01), natural resources and conservation (03), biological and biomedical sciences (26), health/related clinical sciences (51)

[26]The choice of which majors to include in each category are inspired by Arcidiacono et al. (2016) and Saltiel (2019)

- Math/Physical Science: Computer/information science/support (11), engineering (14), engineering technologies/technicians (15), mathematics and statistics (27), physical sciences (40)
- Non-STEM: All other majors

Given these criterion, I am unable to classify every students' final outcome at age 26. Some students attrited from the sample, meaning the NCES was not able to determine whether they completed college, while others may have completed college but the NCES was unable to determine their field of study. These individuals are left in the sample and I utilize as much of their information as possible in the estimation procedures. At this point in the data cleaning process, the sample size has been reduced to $9,450$ individuals, of which I know the final outcomes of $8,060$.

High School Characteristics

The ELS data has much information on the characteristics of high schools that students attend. Of this information, I primarily utilize two components: graduation requirements in mathematics and science, and information on the number of math, physical science, and life science courses offered at the school.

Schools were asked, in the second round of the survey, the number of years of math and science students are required to take to graduate from their high school. Answers were given in intervals, not exact values, so the answers needed to be recoded. Tables A1 provides this recoding.

Schools also provided a course catalog that contains a list of all courses offered at the high school. Along with their names, a numeric code is provided that classifies each course according to the Classification of Secondary School Courses (CSSC) taxonomy. I further classified these courses, using the CSSC codes, into non-STEM, math, physical science, and life science categories. The mapping between codes and my determination of the subject to which they belong is provided in the next section.

Any student attending a school that is missing graduation requirements or courses offerings will be omitted from the sample.

Table A1: Determination of Math Requirements from ELS Survey

Math/Science Requirements Reported	Value for this Study
At least 1 year but less than 2	1 Year
At least 2 years but less than 3	2 Years
At least 3 years but less than 4	3 Years
4 Years	4 Years

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)

High School Choices and Grades

In this section I describe how I take the high school transcript data collected by the ELS and turn it into a panel of high school coursetaking decisions and grade outcomes.

The first determination is how to classify courses on transcripts as being non-STEM, mathematics, physical science, or life science. Each course has associated with it a CSSC code, and each CSSC code corresponds to one of many specific subject categories. I take these codes and further categorize courses into the 4 subjects that I am considering in the following way:

- Mathematics Courses:
  - Mathematics courses:
    - CSSC code $\in [270000, 280000)$
  - Other misc. mathematics courses:
    - CSSC code $\in \{010151, 070171, 070172, 070221, 110121, 110122, 170651\}$
- Physical Science Courses:
  - Physical science courses:
    - CSSC code $\in [400000, 410000)$
  - Computer science courses:
    - CSSC code $\in [110000, 120000)$

- Engineering courses:
    - * CSSC code $\in [140000, 150000)$
- Engineering and related technologies courses:
    - * CSSC code $\in [150000, 160000)$
- Science technologies courses:
    - * CSSC code $\in [410000, 420000)$
- Other misc. physical science courses:
    - * CSSC code $\in \{300311, 300300, 300611\}$

• Life Science Courses:

- Life science courses:
    - * CSSC code $\in [260000, 270000)$
- Agricultural science courses:
    - * CSSC code $\in [020000, 030000)$
- Natural resources courses:
    - * CSSC code $\in [030000, 040000)$
- Allied health courses:
    - * CSSC code $\in [170000, 180000)$
- Health science courses:
    - * CSSC code $\in [180000, 190000)$
- Other misc. life science courses:
    - * CSSC code $\in \{300111, 300112, 300121, 300100, 300621, 300623, 300600\}$

• Non-STEM Courses:

– All other courses

After classification is complete, grades for each course are converted from letter grades to a 4.0 scale, with grades of A+ and A being 4.0 and a grade of F being 0.0 (I use standardized grades provided by the NCES that are comparable across high schools for this calculation). Some courses have non-letter grades associated with them: 'pass', 'unsatisfactory', 'withdrew', 'incomplete', 'non-graded', and 'blank'. I recode 'withdrew' to have a grade point of 0.0. For the other non-letter grades (except for non-graded) I impute their values using a flexible subject specific ordinary least squares interpolation.

Each course can also count for some number of credits. While high schools may have their own credit systems, as is done with grades, the NCES standardizes reported credits so that they are comparable across schools. For each year, I calculate a credit-weighted GPA for non-STEM, math/physical science, and life science courses which are used as GPA outcomes in the estimated model.

Recall that students can potentially choose 9 possible combinations of courses in any given year (provided that their high school graduation requirements are not limiting their choice set). A student is considered to have taken year of a particular subject if the number of reported credits for that year in that subject is greater than 0. To determine when a student drops out of high school, I first look at two variables in the ELS that contain the grade in which the student reported dropping out of high school or completing a GED. If a student who is determined to be a dropout is missing this variable, I assign their year of high school dropout to be the first grade in which the student is missing high school transcript data or has some other data problem with their transcript.

To determine if a student is meeting their STEM graduation requirements, I recalculate the number of years left in math or science after every year of high school. If a student takes a year of mathematics, then the required number of years left in math reduces by 1. If a student takes a year of physical science and/or life science, their required number of years left in science reduces by 1.

Of the students left in the sample after omitting those who are missing important individual level covariates, some have other data problems that cause me to be unable to create for them an

entire panel of high school choices. In order to help alleviate attrition problems that may arise by only keeping students with no data problems, I include as many years of these individual's data as possible. There are 5 main problems that arise in the creation of the high school grade and choice panel that preclude me from observing a student's entire high school career. The first issue is with non-STEM courses. Recall that in the model I assume students *must* take non-STEM when enrolled in high school. For most students this holds true, but there are a select few who report taking no non-STEM credits during one of their years of high school enrollment. When this happens, I attrit that person from the sample from that year forward.

Other reasons that cause me to need to attrit sample members include reporting credits for a subject but having no calculatable grades, being unable to identify the student's high school, attending a high school that did not report its graduation requirements or its course catalog, or not reporting any high school transcripts for a particular year. If none of these problems arise in a year for a student, but the student makes a decision that conflicts with their school's stated graduation requirements, I also attrit that student from the sample. Table (A2) reports the fraction of attriting students whose data has these issues.

Table A2: Reasons for Attrition from Sample in H.S.

Reason	% of Students Attriting in H.S.
Reps 0 non-STEM Credits	0.03%
No Calculatable Grade	1.26%
No Reported Grad. Reqs./Course Catalog	40.00%
Missing Transcript Data	39.65%
Choice Conflicts with Grad. Reqs.	18.87%

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)

College Choices and Wages

I utilize the information I collected on students' final educational choices (as described in subsection 1.7) to determine a student's post-secondary decision. Every individual who did not dropout or attrit in high school is faced with the decision of either not completing college or going on to obtain a bachelors degree in a non-STEM, math/physical science, or life science field. Grades in college are

simply a student's overall college GPA as calculated by the NCES in the ELS survey. Students who did not attrit in high school, but do not have their final outcomes or their college GPAs reported are counted as attritting in college.

Overall, approximately 40% of students who are not missing basic demographic characteristics attrit, at some point, from my final sample. Again, in order to reduce potential issues related to attrition I use each individual in the estimation procedure until they attrit. In this way, I am able to incorporate their information into parameter estimates and hopefully reduce bias or inconsistency. When running simulations of my model, I include all individuals who attrit grade 10 and onward. Table A3 displays the fraction of attriting students who are omitted in different periods of the model.

Table A3: Reasons for Attrition from Sample in H.S.

Period	Fraction of Attriting Students
9^{th} Grade	42.00%
10^{th} Grade	1.94%
11^{th} Grade	11.94%
12^{th} Grade	17.68%
Post H.S.	26.45%

SOURCE: U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002)

Finally, the ELS data did not follow students long enough to collect very much detailed information on labor market outcomes. Because of this, I am only able to use total wage data at age 26 to estimate wage regressions, and only for those students who did not attrit from the sample. I re-scale wages to be in terms of $\$100,000$, adjust for inflation by putting their value in 2020 dollars, and take their natural log[27]. Approximately 87.12% of non-attriting students have positive values for reported wages.

[27] Inflation calculated using *U.S. Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: All Items in U.S. City Average [CPIAUCSL], retrieved from FRED, Federal Reserve Bank of St. Louis* (n.d.)

Appendix B. Parametric Bootstrap Procedure

The standard errors on all parameter estimates are obtained using the parametric bootstrap procedure utilized and described in Arcidiacono et al. (2016):

1. Sample, with replacement, every individual in my dataset at grade 9.
2. Draw an unobserved type and ability for each student using the estimated distributions.
3. Forward simulate each individual's decision process from grade 9. At every point in time, I draw each individual's schooling choice using the estimated CCPs. Given their choice of course/major I then draw their grade realization using estimated grade parameters. Finally, I use these to update each student's information set at the end of the period.
4. Upon completion of the forward simulation, I use the simulated dataset to re-run the estimation procedure described in Section 1.3.2.

Once all 50 parameter estimates have been calculated, I use these to calculate the standard errors.

References

Altonji, Joseph (1995). “The Effect of High School Curriculum on Education and Labor Market Outcomes”. In: *Journal of Human Resources* 30.3, pp. 409–438.

Altonji, Joseph, Peter Arcidiacono, and Arnaud Maurel (2015). “The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects”. NBER Working Paper.

Altonji, Joseph, Erica Blom, and Costas Meghir (2012). “Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers”. NBER Working Paper.

Arcidiacono, Peter (2004). “Ability Sorting and the Returns to College Major”. In: *Journal of Econometrics* 121.1-2, pp. 343–375.

Arcidiacono, Peter and Paul Ellickson (2011). “Practical Methods for Estimation of Dynamic Discrete Choice Models”. In: *Annual Review of Economics* 3, pp. 363–394.

Arcidiacono, Peter and Robert Miller (2011). “Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity”. In: *Econometrica* 76.6, pp. 1823–1867.

Arcidiacono, Peter et al. (2016). “College Attrition and the Dynamics of Information Revelation”. NBER Working Paper.

Bernal, Raquel and Michael Keane (2010). “Quasi-Structural Estimation of a Model of Childcare Choices and Child Cognitive Ability Production”. In: *Journal of Econometrics* 156.1, pp. 164–189.

Erdem, Tülin and Michael Keane (1996). “Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets”. In: *Marketing Science* 15.1, pp. 1–20.

Goodman, Joshua (2019). “The Labor of Division: Returns to Compulsory High School Math Coursework”. In: *Journal of Labor Economics* 37.4.

Heckman, J. and B. Singer (1984). “A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data”. In: *Econometrica* 52.2, pp. 271–320.

Heckman, James and Edward Vytlacil (2001). “Policy Relevant Treatment Effects”. In: *The American Economic Review* 91.2, pp. 107–111.

Hotz, Joseph and Robert Miller (1993). “Conditional Choice Probabilities and the Estimation of Dynamic Models”. In: *The Review of Economic Studies* 60.3, pp. 497–529.

Jacob, Brian et al. (2017). “Are Expectations Alone Enough? Estimating the Effect of a Mandatory College-Prep Curriculum in Michigan”. In: *Educational Evaluation and Policy Analysis* 39.2, pp. 333–360.

James, Jonathan (2011). “Ability Matching and Occupational Choice”. Federal Reserve Bank of Cleveland, Working Paper.

Joensen, Juanna Schrøter and Helena Skyt Nielsen (2009). “Is There a Causal Effect of High School Math on Labor Market Outcomes?” In: *The Journal of Human Resources* 44.1, pp. 171–198.

Kahn, Shulamit and Donna Ginther (2017). “Women and STEM”. NBER Working Paper.

Kautz, Tim et al. (2014). “Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success”. NBER Working Paper.

Kinsler, Josh and Ronni Pavan (2015). “The Specificity of General Human Capital: Evidence from College Major Choice”. In: *Journal of Labor Economics* 33.4, pp. 933–972.

— (2021). “Local Distortions in Parental Beliefs over Child Skill”. In: *Journal of Political Economy* 129.1.

Levine, Phillip and David Zimmerman (1995). “The Benefit of Additional High-School Math and Science Classes for Young Men and Women”. In: *Journal of Business and Economic Statistics* 13.2, pp. 137–149.

Lillard, Dean and Philip DeCicca (2001). “Higher Standards, More Dropouts? Evidence Within and Across Time”. In: *Economics of Education Review* 20.5, pp. 459–473.

Lleras, Christy (2008). “Do Skills and Behaviors in High School Matter? The Contribution of Noncognitive Factors in Explaining Differences in Educational Attainment and Earnings”. In: *Social Science Research* 37.3, pp. 888–902.

Michigan Merit Curriculum High School Graduation Requirements (2008). `respectabilitylawcenter.com/docs/highschool-grad-requirements.pdf`. Accessed: 02-15-2022.

Mroz, Thomas and Timothy Savage (2006). “The Effect of High School Courses on Earnings”. In: *Journal of Human Resources* 41.2, pp. 259–293.

NCES (n.d.). *NAEP: The Nation’s Report Card FAQ*. `nces.ed.gov/nationsreportcard/about/faqs.aspx`. Accessed: 02-15-2022.

— (1999). *State requirements for high school graduation, in Carnegie units: 1993, 1996, and 1998*. `https://nces.ed.gov/programs/digest/d99/d99t157.asp`. Accessed: 02-15-2022.

NCES (2000a). *Classification of Instructional Programs*. nces.ed.gov/pubs2002/cip2000.

— (2000b). *CSSC Courses/Course Codes*. nces.ed.gov/surveys/hst/courses.asp. Accessed: 03-01-2022.

— (2002). *U.S. Department of Education, National Center for Education Statistics. Education Longitudinal Study of 2002 (ELS:2002).*

— (2018). *State course credit requirements for high school graduation, by state: 2018*. https://nces.ed.gov/programs/statereform/tab2_13.asp. Accessed: 02-15-2022.

NYS Science, Technology, Engineering and Mathematics (STEM) Incentive Program (n.d.). www.hesc.ny.gov/pay-for-college/financial-aid/types-of-financial-aid/nys-grants-scholarships-awards/nys-science-technology-engineering-and-mathematics-stem-incentive-program.html. Accessed: 02-15-2022.

Rose, Heather and Julian Betts (2004). "The Effect of High School Courses on Earnings". In: *The Review of Economics and Statistics* 86.2, pp. 497–513.

Saltiel, Fernando (2019). "What's Math Got to Do With It? Multidimensional Ability and the Gender Gap in STEM". Society for Economic Dynamics 2019 Meeting Paper.

Stange, Kevin (2012). "An Empirical Investigation of the Option Value of College Enrollment". In: *American Economic Journal: Applied Economics* 4.1, pp. 49–84.

Stinebrickner, Ralph and Todd Stinebrickner (2014). "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout". In: *The Review of Economic Studies* 81.1, pp. 426–472.

Stinebrickner, Todd and Ralph Stinebrickner (2012). "Learning about Academic Ability and the College Dropout Decision". In: *Journal of Labor Economics* 30.4, pp. 707–748.

Thomas, James (2019). "The Signal Quality of Grades Across Academic Fields". In: *Journal of Applied Econometrics* 34.4, pp. 566–587.

Train, Kenneth (2008). "EM Algorithms for Nonparametric Estimation of Mixing Distributions". In: *Journal of Choice Modelling* 1.1, pp. 40–69.

Train, Kenneth E. (2009). *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge University Press.

U.S. Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: All Items in U.S. City Average [CPIAUCSL], retrieved from FRED, Federal Reserve Bank of St. Louis (n.d.). `https://fred.stlouisfed.org/series/CPIAUCSL`.

Vita

Phillip Edward Gustafson is an economist with interests in labor economics, the economics of education, and applied econometrics. He obtained his PhD in economics from Georgia State University in the spring of 2022. As a doctoral student, he was a recipient of both the Second Century Initiative Doctoral Fellowship and the Provost's Dissertation Fellowship. He also served as a graduate research assistant for both Dr. Thomas Mroz and the Georgia Policy Labs, where he worked on projects in health economics and the economics of education respectively. Before attending Georgia State, he completed his undergraduate studies at the University of Massachusetts - Boston, where he received a bachelors degree in economics with minors in mathematics and political science.