

An Examination of the Relationship between Student Expectations and College Attainment

M. Gray Hunter

University of Kentucky

October 19, 2016

Abstract

This paper examines the role that students' educational expectations play in the attainment of post-secondary schooling. Data from the National Longitudinal Study of 1997 is used to document and analyze how student expectations influence the probability of attending college and completing a four-year degree. The results indicate the following. First, although overconfident, individuals as young as fifteen are willing and able to answer subjective probabilistic questions concerning education in a cohesive fashion. Second, although heterogeneous across race, gender, current schooling completion and parent education, a student's self-reported probability of obtaining a four-year college degree is influential in predicting whether or not they attend and ultimately complete a degree. Also, as students age their expectations play a larger roll in predicting completion. While the magnitude of the effect diminishes when including the standard economic controls, expectations remain significant. Parent expectations appear to play a role when students are young, but the significance diminishes as the student ages and gathers information related to the costs and benefits of a college degree. These indicate that students possess some form of private information that is not being completely captured by the standard variables used by econometricians to predict college attendance and completion.

Introduction

The recent increase in the cost of attending college has led to a number of discussions, including how to reduce these costs and whether a degree is still a worthwhile investment. However, the matriculation decision process and elements which contribute to both students' attendance and success are seldom discussed. I present new estimates of the factors that predict a student's decision to attend college and whether or not they complete a four-year degree that econometricians have not previously found. I focus on the predictive power of students' self-reported probability of obtaining a four-year college degree. Later I consider how the predictive power of these expectations vary as the student ages and progresses through different levels of schooling. We anticipate that as students gain more information about the costs and benefits of the degree as well as their ability to complete it, their expectations should play a larger role in predicting their outcomes.

Manski (1993) outlined the importance of understanding and identifying how a student's expectations influence their decision to attend college or not. He notes that the estimated returns to education could be flawed when economists rely on the assumption of rational expectations instead of attempting to understand the process by which students use information to form expectations about the costs and benefits of a college degree. Dominitz and Manski (1996) and Manski (2004) both examine these issues, but are limited by small sample sizes and cross-sectional data. Jacob and Wilder (2011) are able to observe students over time and study the role of expectations in degree attainment, but do not examine whether these also influence the individual's decision to attend college. They are also limited by the nature of their data as their measure of expectations is imprecise.

A unique set of questions from the National Longitudinal Study of 1997 (NLSY97) allows for an examination of how a student's subjective probability of completing a college degree changes as they progress through schooling and what role it plays in their outcomes by age 30. The survey asks a subset of students to report their probability of obtaining a four-year college degree by the time they turn 30, once in high school and again when some

have entered either college or the workforce. To date no other studies have used these data to examine the role of expectations in educational outcomes. The probability scale is an improvement on the previous data, as most other studies use data where students respond on a Likert scale regarding their educational expectations, or are simply asked the highest level of education they expect to complete without any notion of how confident they are in realizing that outcome.

The results indicate that, even after controlling for the factors previously identified in the economics and education literature, a student's self-reported probability of completing a four-year college degree plays a role in predicting whether they attempt to complete a college degree and if they are successful in obtaining a four-year degree. As students age and progress through more schooling, their expectations play a larger role in predicting completion. Also, parent expectations are estimated to be influential when students are young, but the effect diminishes as the student ages and gathers information related to the costs and benefits of a college degree. The estimates signal that students' expectations contain a form of private information that is helpful in determining post-secondary schooling attendance and success that isn't being captured by the previously identified controls, and suggests that youth and econometricians may possess and utilize different data when making schooling decisions. The use of interpretable subjective data on expectations, combined with choice data, can also help solve the identification problem discussed by Manski (1993).

The remainder of the paper is organized as follows. In section 2, I provide background information, discuss the previous literature, and present a conceptual framework for how expectations influence attainment. Section 3 includes a description and brief analysis of the data. In section 4, I analyze the role of expectations in regards to college matriculation and completion. First, I explore the data which indicates a positive relationship between expectations, other factors associated with post-secondary success, and attainment. I then test the statistical significance using regression analysis. Sensitivity analysis is included in section 5. Here, several variations in which expectations might alternatively influence both

outcomes are investigated. I also briefly examine parents' expectations and how they are related to their child's college attendance and completion. Section 6 discusses the implications of these findings and concludes the paper.

Background on Expectations

Since Manski's (1993) discussion of students as adolescent econometricians, economists have recognized the importance of identifying the information that students use regarding their decision to pursue a post-secondary education. Students face similar issues to econometricians in that they may possess different data, have different knowledge of the economy, and may process information in different ways. He focuses on the universal assumption that students form expectations homogeneously.¹ Using a model of school choice where students either do or do not condition based upon their ability, he shows how the assumption of a homogeneous expectations formation process for all students significantly alters the predictions from the model and leads to two identification problems.² This paper does not specifically focus on the identification issue, but reviews the literature on the previous factors that have been used to predict school choice and tests whether student expectations play a role econometricians have not previously found.

A large portion of the economics of education literature examines how the cost of attendance influences the attendance decision of students. Hilmer (1998) presents a theoretical model based on school pricing and ability that outlines the decision that students make when deciding to attend college or not, and if so, whether they should attend a two or four-year degree granting institution initially. As expected, the model predicts that when the net cost of attendance decreases more students should attempt to obtain a college degree overall as well as the opposite. Dynarski (1999) follows this and uses social security benefits to estimate

¹Economic studies of schooling behavior use this both implicitly and explicitly in the majority of previous research but rely on different information processing rules and hypothesized conditioning variables.

²The first is that, not knowing how youth perceive the returns to schooling, researchers cannot infer their decision processes based on their schooling choices. Second, is that not knowing a youth's decision processes, the objective returns to schooling cannot be estimated from data on realized outcomes.

that a \$1,000 increase in student aid results in a 3.6 percentage point increase in attendance and students complete an additional 1/10th a year of post-secondary schooling. Manski and Wise (1983), Leslie and Brinkman (1988), and Dynarski (2000) find similar results using a variety of different semi-natural experiments and program implementations.

Others have examined the role of gender, race, socioeconomic status, and information in college attendance and completion. Perna (2000) discusses how during the 1990s the percentage of undergraduates that were either African American or Hispanic increased significantly, yet they remained underrepresented in both the undergraduate and bachelor degree recipient populations compared to their relative representation in the traditional college-aged population. Her findings suggest that financial aid in the form of loans reduces the probability of enrolling for African Americans compared to other races. The probability of enrolling also increases for Whites and African Americans as the local unemployment rate increases, but Hispanics are uninfluenced. Lastly, she identifies several other non-cost related factors that have heterogeneous effects on college attendance based on race.

Bailey and Dynarski (2011) later demonstrate that overall college completion rates have increased, but the gaps in college entry, persistence, and graduation for children from high and low income families have increased. They note this is especially true for women, driven by an increase in the education of daughters of high-income parents. Hoxby and Avery (2012) and Hoxby et al. (2013) also find that low-income high achieving students either did not attend college or attended two-year institutions at a much higher rate than their high-income high achieving counterparts due to a lack of information about the “net” cost of college. Their experiment supplied students with information regarding available scholarships and estimates of the out-of-pocket cost of several surrounding colleges for income and achievement similar students. After supplying students with this information, they discovered that their attendance and completion rates increased such that they were comparable to those of similar ability. While information such as this is difficult to measure, it is possible that it could be contained in a students’ expectations of completing college in the future.

Research examining educational expectations and student outcomes dates back to the 1960s, with Willam H. Sewell, Otis D. Duncan, and their colleagues being credited with the seminal work demonstrating a positive correlation between expectations and both educational and occupational attainment (Jacob and Wilder, 2011). A hurdle that the expectations literature was forced to overcome was the apprehension that economists had regarding the accuracy and usefulness of self-reported subjective data. To combat this, Dominitz and Manski (1996) designed and applied an interactive computer-administered personal interview (CAPI) survey that elicited expectations of high school students and undergraduates regarding their income if they were to complete different levels of schooling. They find that respondents, even as young as high school juniors, are willing and able to respond meaningfully to questions eliciting their earnings expectations in probabilistic form. Dominitz and Manski (1997) followed by collecting data on the one-year-ahead income expectations of members of American households in a survey of economic expectations and found again that self-reported expectations were reliable and useful when making predictions.

The following papers are closely related to this research and echo the findings of Dominitz and Manksi (1996,1997) that expectations are meaningful. Bernheim (1990) used the Social Security Administration's Retirement History Survey to examine the elderly's expectations of Social Security benefits during a preretirement period and how they respond to new information. He finds that expectations of benefits are relatively noisy, but they can be used to make accurate predictions about their benefits in the future. He also notes sizable differences in accuracy conditional on race, gender, and marital status. Smith et al. (2001) used four waves of the Health and Retirement Survey (HRS) to test whether longevity expectations were related to the observed mortality for individuals. They discover that subjective beliefs about longevity are consistent with individuals' observed survival patterns, and that deaths are signaled through lower longevity expectations reported in previous periods. Lochner (2007) empirically examines belief updating of the perceived probability of arrest and its criminal deterrence effects. His estimates suggest that a higher perceived probability of

arrest reduces criminal participation. In each case, the authors find that expectations are reliable and have predictive power. This study adds to this literature by focusing on the expectations of teens and young adults and how they relate to educational outcomes.

Fischhoff et. al. (2000) focuses on the teens' responses in the first round of the NLSY97 to the eighteen questions regarding significant future life events. While the objective of the study was to examine the risk perceptions of teens, the authors comment extensively on the responses and how these related to public health figures concerning the mentioned event. The most relevant for this study are those related to the educational expectations of students. They examine the percent chance a student believes they: are in school next year, obtain a high school diploma by age 20, and earn a four-year college degree by age 30.³ Their findings indicate that as the time from anticipated completion increases so does student error. When predicting whether they will be a student next year, the mean is nearly identical to public health statistics for children of a similar age at the time. As students look further into the future, they are overly optimistic about their chances of graduating both high school and college, reporting means of nearly 93% and 73% respectively, when in reality the graduation rates are approximately 84% and 25% respectively.

Other findings from Fischhoff et al. (2000) include differences in reporting patterns between those still in school and those no longer enrolled. A strong correlation between expectations of graduating both high school and college is found with reported percent of peers that plan to go to college. Overall, they find that for the majority of categories both student and parent expectations are relatively accurate, at least on average, when compared to national public health data. This indicates that both adults and teens as young as fourteen are willing and able to answer probabilistic questions and are relatively accurate. While the findings are interesting, they are only able to compare the student's expectations to the outcomes of similar students during that time. This study isn't limited by this as both

³The high school expectations variable is not examined in this paper as graduating high school is a more common outcome. Later work will use the variable to look at how student expectations are formed and updated.

the students' expectations and their outcomes as of age 30 are observed. This allows for a compare of their expectations to their eventual outcomes and examine the predictive power they may have.

To this point, only two studies have inspected the role of student expectations. Zafar (2013) uses a unique panel data set of 161 Northwestern University sophomores that contains the students' subjective expectations regarding several of their future academic outcomes.⁴ He uses these expectations to estimate a choice model where selection of a college major is made under uncertainty in an attempt to explain the markedly different choices by males and females. He finds that students are generally overconfident in their future academic performance, but they revise their expectations in an expected way. However, using students from Northwestern results in sample selection issues. Northwestern is one of the top universities in the United States, it only admits approximately 13% of applicants, and has a ACT composite range of 31-34 (Forbes, 2016).⁵ It is possible that students who attend and complete their first year at Northwestern differ in both the information they receive and how they process it than a more representative sample of students across different colleges. My research is able to combat this as it uses a representative sample from across the US and observes students prior to attending college.

Jacob and Wilder (2011) utilize a combination of data from Monitoring the Future, High School and Beyond, the National Educational Longitudinal Survey of 1988, and the Educational Longitudinal Survey of 2002 to answer a number of questions related to educational expectations. They show that high school seniors expecting to complete at least some college rose from the mid 1970s up until the late 1990s where they peaked around 93%. However, while expectations rose during this period actual attainment did not follow suit, thus the predictive power of expectations fell. They found that over sixty percent of students update their educational expectations at least once between 8th grade and eight years post high-

⁴These include the percent chance that: they graduate with a GPA of 3.5 or better, enjoy courses, hours spent on coursework per week, find a job, enjoy work, work flexible hours, etc....

⁵Forbes lists Northwestern University as a top 15 college overall. Its ranked as the #14 in private colleges, #9 in research universities, and #2 overall in the Midwest. The SAT composite range is 1390-1560.

school, and that it is primarily based upon the acquisition of new information regarding their ability through signals like GPA, test scores, or graduating. They also note that students who do not update as significantly, or who they consider are more aligned at a young age, generally have a higher likelihood of reaching their expectations.

The literature above expresses the importance of incorporating subjective probability data into models of school choice. They show that self-reported expectations have predictive power in terms of a number of different outcomes. If student expectations can also be used to predict college attendance and completion it would reaffirm Manski's notion that students face similar issues to econometricians in that they may possess different data, have different knowledge of the economy, and may process information in different ways when it comes to school choice. Several relevant factors have been identified as influencing a student's decision to attend a post-secondary institution and their completion. I base the controls used in the models below on these findings prior to including expectations. None of the studies above are able to examine the role of expectations in this process using subjective probabilistic data gathered directly from the student as Manski suggests. Jacob and Wilder (2011) attempt this, but they are limited by students reporting their expectations of college attendance and completion on a Likert-like scale. This forces them to examine large, discontinuous changes in expectations, with students reporting their highest expected level of attainment without a measure of confidence attached to it. I further contribute to the literature by using probabilistic data to estimate a more precise relationship between expectations and post-secondary education outcomes.

Data

The data derive from the National Longitudinal Survey of Youth 1997 (NLSY97). The motivation behind the NLSY97 was to provide researchers the opportunity to identify characteristics that define the transition from schooling to the labor market for youth during this

period. It was designed such that participants would be representative of the United States in 1997 for those born during the years 1980 through 1984.⁶ They first administered surveys in 1997 to 8,984 individuals who were born between January 1, 1980 and December 31, 1984.⁷ Following the initial 1997 survey, the youth were interviewed subsequently on an annual basis up until 2011, after that the survey was and will be administered biennially. This study utilizes data from 1997 up to the first biennial responses from 2013. Overall, approximately 80% of those initially surveyed in 1997 also completed the most recent 2013 survey used.

In the first round of interviews, both the selected youth and one of their parents each participated in a one hour long personal interview. The youth questionnaire focused in detail on their schooling and employment activities, but also included questions regarding their family background, social behavior, health status, and expectations regarding future events. The inclusion of the parent is beneficial in that it provided more extensive and reliable information regarding the youth's family background, household environment, and general history.⁸

In addition to the round 1 interview, the youth were also incentivized to take a computer-adaptive version of the Armed Services Vocational Aptitude Battery (ASVAB) at a separate time. The ASVAB is composed of 10 subject tests that are meant to measure knowledge and skill of participants in a number of key areas including mathematics, science, and language.⁹ The exam was given to individuals in groups of five to ten under standardized conditions at more than 280 sites across the the country. The tests were primarily administered at Sylvan Learning Centers, but some temporary testing sites were also established at hotels, community centers, and libraries when a testing center was not available. Respondents were

⁶Bureau of Labor Statistics, U.S. Department of Labor (2013)

⁷Respondents ranged from 12 to 17 years old at the date they were initially interviewed

⁸Detailed information was collected on a number of topics, including: education, training, and achievement scores; employment; household characteristics; income, assets, and program participation; health; crime and substance abuse; and attitudes, expectations, non-cognitive ability measures, and general activities.

⁹Testing was conducted by the Department of Defense (DOD) and took place from the summer of 1997 through the spring of 1998; the DOD used the NSLY97 participants as part of a larger effort to establish new norms for the test, which is primarily used for military enlistment screening.

given \$75 to participate and in the end 7,712 or 79.3% of the sample took the computerized ASVAB.^{10 11} While the ASVAB is generally used as a proxy for ability, I utilize it as a source of information for the students and use the students grades from 8th grade and high school as measures of academic ability.¹²

Students' high school transcript data was collected by the BLS directly from the schools in two waves for those who provided written permission. These data can be broken down into individual terms for which the student was in school and contains their courses taken, grades, and school related activities in which they were involved. These data are available for 6,232 or 69% of the individuals who participated in the first round of the study. The low reporting level could signal possible selection by either parents or students who chose to release their grades. However, students were asked in the survey to self-report their grades from both 8th grade and high school and the response rate was much higher than the school collected grades. While Black et al. (2011) find measurement error issues for individuals reporting their highest level of education attainment, when comparing the students' self-reported grades to their transcripts they were relatively precise. It is possible that those who selected into providing access to their transcript GPA might differ in their reporting from those who do not, but both measures were used separately. The results suggest that

¹⁰This is considerably lower than the 93.9% completion rate for the pencil and paper format of the comparable Armed Forces Qualifying Test given to the National Longitudinal Survey of Youth 1979 participants. Per Donna Rothstein and Mark Loewenstein at the the BLS, this is likely the result of a number of factors. First, the NLSY97 test was given in the first year of the survey compared to the second year for the the NLSY79. Second, \$50 was offered to participants in 1979 compared to \$75 in 1997; \$50 in 1979 would have been worth approximately \$117 in 1997. The non-takers in the NLSY97 were primarily Hispanic. Third, a Spanish-language version might not have been available which would have hurt the NLSY97 participants more as it is composed of approximately 21% Hispanics compared to just 15% for the NLSY79. Also, in the NLSY79, special needs students and non-English speakers were excluded which was not the case for the NLSY97. The data supports this as the non-takers were primarily Hispanic.

¹¹In an attempt to maximize sample size, those that did not take the ASVAB were given the mean score of their ethnic group and an imputation indicator variable was included in all regressions where the ASVAB percentile was used.

¹²Neal and Johnson (1995) argue that it is an adequate measure of human capital for students that have yet to enter the labor market. They find that its' exclusion lead to biased estimates when examining student outcomes. They focus on wages when the students are 26 to 29, but endogenous selection into schooling is discussed. However, Bollinger (2003) and Black et al. (2011) find that measurement error in the AFQT and ASVAB, when used as proxy for human capital, can lead to biased estimates of coefficients for correctly measured variables that are correlated with human capital.

the inclusion of the transcript grades increases the explanatory power of the model but did not significantly alter the estimates of interest. Therefore, to maximize sample size only those models that used self-reported grades are presented.¹³

The NLSY97 is composed of 2 subgroups. The first, which is referred to as the cross-sectional sample, consists of 6,748 respondents that are representative of those born in United States during the period of January 1, 1980 to December 31, 1984. The second, which is referred to as the supplemental or over-sample, includes 2,236 respondents that are either Hispanic/Latino or African American. The full sample is 51% male and 49% female, and 52% Non-black/non-Hispanic (categorized as white), 26% Black non-Hispanic, 21% Hispanic or Latino, and 1% Mixed or Multirace.¹⁴ The models below utilize the full sample, and includes an indicator for whether or not the individual was in the representative portion.

This research relies on two subsets of these data. The first includes only those who were 15 or 16 years old as of December 31st, 1996. These 3,450 respondents were asked to report their educational expectations during round 1 in 1997. The second subset includes 1,902 individuals who were asked to report their educational expectations during round 5 in 2001.¹⁵ Seven hundred were asked in both 1997 and 2001.¹⁶ Summary statistics for the full sample, the two expectation subsets, and the 700 overlapping individuals can be found in table 1. The 1997 and 2001 subsets have a similar gender and race composition compared to the full sample. Each are nearly 51% male and 49% female, and are approximately 52% white. The 1997 subset is 27% black while the 2001 subset is 25.4% black. The two are also 20.8% and 23% Hispanic respectively. When they are broken down into the cross-sectional and over-sample categories, there is more variation from the entire sample. A means test for both race and gender show that the two expectations samples are not statically different

¹³For the estimates that include the transcript data email gray.hunter@uky.edu.

¹⁴The mixed race individuals were classified as Black in this study per the suggestion of Dr. Chris Bollinger.

¹⁵This group was a part of an experiment where a fraction of the individuals were randomly chosen to answer the expectations questions in each appropriate topical section, rather than all together in one section as in the 1997 survey. This was done to examine the difference in response rates as a result of the different grouping and ordering. I have been unable to locate the results of this study.

¹⁶The age range of respondents is larger for the second group as it was not restricted like the initial sample.

from the full sample. The 700 overlapping students are of 49% male and 51% female, and 50.8% white, 25.4% black, and 23.7% Hispanic.

Table 1: Summary Statistics

	Sample Type	Individuals	Mean Age	Male	White	Black	Hispanic
Full Sample	Cross-Section	6,614	13.92	51.39%	69.44%	16.84%	13.71%
	Over-Sample	2,162	14.02	51.25%	0%	55.64%	44.36%
	Total	8,776	13.98	51.36%	52.34%	26.40%	21.26%
1997 Sub-sample	Cross-Section	2,582	15.47	50.39%	69.71%	16.77%	13.52%
	Over-Sample	868	15.47	51.84%	0%	57.72%	42.28%
	Total	3,450	15.47	50.75%	52.17%	27.08%	20.75%
2001 Sub-sample	Cross-Section	1,440	18.90	51.67%	68.13%	15.90%	15.97%
	Over-Sample	462	18.96	49.13%	0%	55.19%	44.81%
	Total	1,902	18.92	51.05%	51.58%	25.44%	22.98%
Overlapping 700	Cross-Section	517	15.52	49.32%	68.86%	14.89%	16.25%
	Over-Sample	183	15.42	49.18%	0%	55.19%	44.81%
	Total	700	15.49	49.29%	50.86%	25.43%	23.71%

Notes: Mean age for the full sample, the 1997 sub-sample, and the overlapping 700 is as of December 31st, 1996.

Expectations in the NLSY97

A unique set of questions included in the NLSY97 are those that pertain to the respondent's expectations that future events will take place. The expectations questions range from things as simple as the percent chance the individual eats pizza next year to more complicated issues such as the percent chance they go to jail or die in the following year. While a number of other studies ask questions about respondents' expectations, the NLSY97 is unique in how they are asked to respond. Most surveys ask individuals to respond to the likelihood of an event occurring using a Likert type scale where participants choose from a range of possibilities similar to "never" or "least likely" to "always" or "very likely" or simply ask respondents to report their highest expected level of educational attainment. Examples of this include the Educational Longitudinal Study of 2002, the High School and Beyond, and the National Educational Longitudinal Study of 1988. The benefit of the NLSY97 is

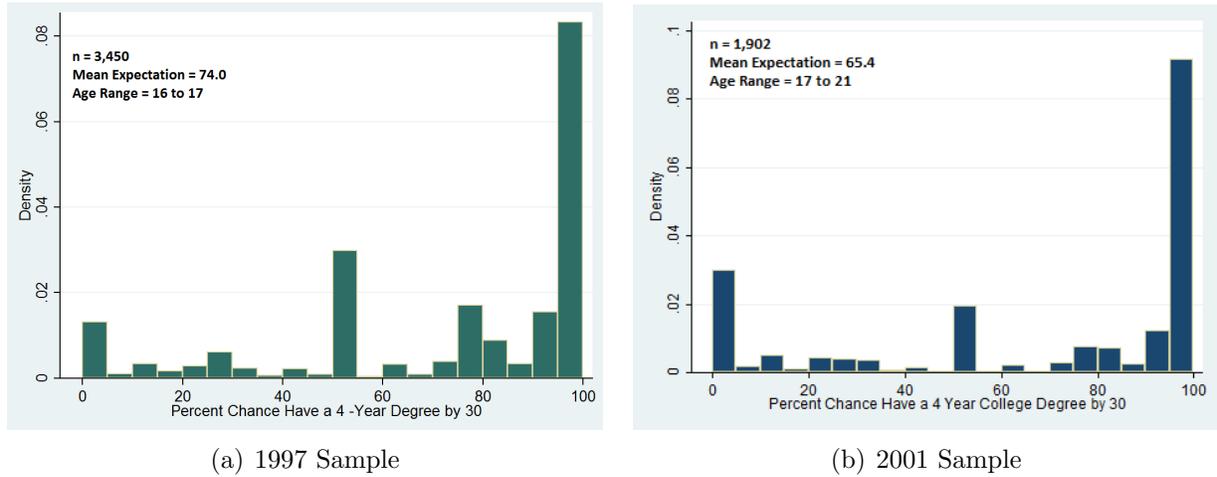
that individuals are asked to report the percent chance between 0 and 100 that they believe they will either graduate high school or college by a certain age. This elicits the student's subjective probability the event will occur and allows for a more detailed analysis.

This study focuses on the respondents' educational attainment expectations, more specifically the belief that they will have completed a four-year college degree by the time they are 30 years old. The survey asks the two subsets of the sample in 1997 and 2001 the following question, "Now think ahead to when you turn 30 years old. What is the percent chance that you will have a four-year college degree by the time you turn 30?" Participants then responded with their belief on the scale from 0 to 100 percent.¹⁷ Figure 1 below presents the frequency distributions of expectations for both the 1997 and 2001 respondents.

The 1997 cohort has a higher mean, similar median, and less variation in their expectations than the 2001 cohort. This could signify that students are updating as they age and progress through schooling, and they are becoming more certain about the outcome whether it be that they will or will not obtain the degree. This idea is reinforced when examining the 700 individuals asked the question in both 1997 and 2001. The average change in expectations was a decrease of 8.6 percentage points, with a standard deviation of 40.1. Updating covered the entire spectrum as some students increased their expectations from 0% to 100%, while others decreased them from 100% to 0%. Overall, 212 students increased their expectations, 272 decreased them, and 216 did not change their expectations during the period. On average, students are overly optimistic about their chance of obtaining a four-year college degree. For the younger, 1997 cohort the average of all of the students' expectations is 72.4% while only 25% of them go on to obtain a four-year degree by age 30. For the 2001 cohort, which is typically two years older than the 1997 cohort, the statistics are 67.3% and 27% respectively.

¹⁷Bureau of Labor Statistics, U.S. Department of Labor (2013)

Figure 1: Frequency Distribution of College Completion Expectations



From the distributions above we see that there is considerable heaping around 0, 50, 75, 90, and 100 percent with the largest share of respondents reporting a 100 percent chance of obtaining a degree.¹⁸ The initial models do not control for this heaping, but as a sensitivity check for misreporting, under, and over confidence I analyze those who heap at 0, 50, and 100% in a later section.

Data Analysis

Prior to discussing how expectations are related to several of the other factors related to post-secondary achievement I first show that, although students are extremely overconfident in their ability, their expectations are still directly related to educational outcomes. First, I calculate correlation coefficient for: expectations and attendance for the 1997 cohort, and expectations and completion for both cohorts.¹⁹ For attendance, the estimated coefficient is .4028 signaling a weak, yet positive association. For the same cohort, the correlation of

¹⁸In 1997, 5.9% of the sample reported a 0 percent chance of completing a four-year degree, 14.7% reported a 50 percent chance, 8.2% believed they had a 75 percent chance, 7.4% believed they had a 90 percent chance, and 35.4% reported a 100 percent chance. In 2001, 13.8% reported a 0 percent chance, 9.7% reported a 50 percent chance, 3.7% reported a 75 percent chance, 5.9% believed they had a 90% chance, and 38.5% reported a 100 percent chance.

¹⁹I do not calculate the correlation coefficient for attendance and expectations for the 2001 cohort as the majority of this sample is already college aged.

.3320 between expectations and completion is smaller but still positive. When the older 2001 cohort is used this correlation jumps dramatically to .4587. This is to be expected as the majority of these students are of college age and should have a better understanding of what it takes to complete a college degree.

Similar to Hurd and McGarry (2002), I also compare the expectations of those who go on to complete the outcome of interest to those who do not. If a relationship is present, we expect that those who report lower(higher) expectations initially are less(more) likely to complete any college and a four-year degree. The results are in the table below. I divide the 2001 sample into 3 categories based on their highest level of academic achievement up to that point in order to further analyze the relationship. In each case, those that go on to both attend and complete college have higher initial expectations on average than those who do not. For example, those in the 1997 sample who go on to attend college have an average expectation of 84.3% while those who do not have an average of 58.53%. For completion, those who go on to complete the degree have an average initial expectation of 90.49% while those who did not have an average of 66.17%. A similar pattern is present in the 2001 sample. Also, columns 2 and 4 suggest a sense of false optimism for those in 1997 sample who do not go on to attend or complete college. When they are compared to those who have completed 12th grade or less from the 2001 cohort, they have noticeably higher expectations. This subset of the 2001 sample is on average a year and half older than the 1997 sample when reporting their expectations, so it is possible that during this time the students who had high expectations but a low probability of being able to complete the degree are receiving a signal that they use to update their expectations downward.

Table 2: Student Expectations and Outcomes

Sample	Student Outcomes			
	Attend College	Does Not Attend College	Completes College	Does Not Complete College
1997 Expectations	84.30% (1,850)	58.53% (1,600)	90.49% (877)	66.17% (2,573)
2001 Expectations				
- ≤ 11th Grade	85.59% (333)	41.47% (459)	93.01% (144)	53.38% (532)
- = 12th Grade	83.49% (292)	43.14% (360)	95.95% (120)	52.69% (648)
- > 12th Grade	n/a	n/a	97.79% (258)	76.93% (200)

Notes: Mean expectations are reported for those associated with each outcome. The number of students in each category is in parenthesis. r below represents the calculated correlation coefficients.

- $r_{attend,exp_{97}} = .4028$

- $r_{complete,exp_{97}} = .3320$, $-r_{complete,exp_{01}} = .4587$

Table 3 divides the expectations for each cohort into eight bins and presents the corresponding measures of matriculation, completion, and persistence. The number of and percent of students in each cohort that attempt and complete a four-year degree are included, as well as a measure of persistence that reports the fraction of students who completed at least one year of college who went on to obtain a four-year degree.²⁰ Excluding those who reported a 100% chance of obtaining a four-year degree, there is a distinct positive relationship between reported expectations and both attempting to obtain a college degree and completion of that degree for both the 1997 and 2001 cohorts; this relationship is also present for expectations and persistence outside of the 0 and 100% bins for both cohorts. The differences in attendance and completion rates between the 91 to 99% and the 100% bins suggest possible issues with overconfidence, misreporting, or a lack of understanding of probability. These issues are examined in the sensitivity analysis section.

²⁰This is used as a rough measure of persistence and is included in the last column of the table.

Table 3: Four Year Degree Expectations and Attainment

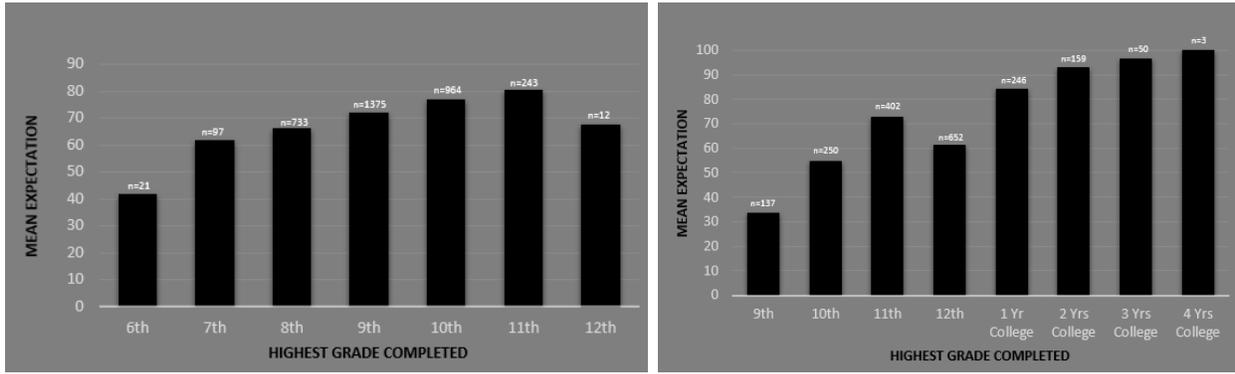
	Percent Chance Have a 4 Year Degree by 30	Number of Students	Number that Attempt Degree	Number that Complete Degree	Percent that Attempt Degree	Percent that Complete Degree	Percent that Complete Degree Conditional on Attempting
1997 Sub-sample	0%	205	30	9	14.6%	4.4%	30%
	1 - 25 %	278	48	5	19.1%	1.8%	9.4%
	26 - 49 %	103	29	3	31.1%	2.9%	9.4%
	50 %	508	159	37	31.3%	7.3%	23.3%
	51 - 75 %	433	240	92	55.4%	21.2%	38.3%
	76 - 90 %	476	315	149	66.2%	31.3%	47.3%
	91 - 99 %	226	173	112	76.5%	49.6%	64.7%
	100 %	1,221	848	470	69.5%	38.52%	55.4%
Total	3,450	1,850	877	53.6%	25.4%	47.4%	
2001 Sub-sample	0 %	263	35	3	13.3%	1.1%	8.57%
	1 - 25 %	179	40	1	22.3%	.6%	2.5%
	26 - 49 %	64	11	1	17.2%	1.6%	9.1%
	50 %	184	53	6	28.8%	3.3%	11.3%
	51 - 75 %	126	70	18	55.6%	14.3%	25.7%
	76 - 90 %	211	124	48	58.8%	22.7%	38.7%
	91 - 99 %	143	119	85	83.2%	59.4%	71.4%
	100 %	732	624	360	85.2%	49.2%	57.7%
Total	1,902	1,076	522	56.6%	27.4%	48.5%	

Notes: Number that Attempt Degree includes those who reported their highest grade completed as 1 year of college or more.

Now that a relationship between expectations and achievement has been identified, it is useful to examine how expectations vary conditional on some of the factors identified in the previous literature also found to be related to post-secondary schooling achievement. First I examine differences based on the highest level of completed schooling when asked about their expectations. Eighty-nine percent of students in the 1997 cohort had completed between 8th and 10th grade when asked their expectations, while for the 2001 cohort ninety percent had completed between 10th grade and 2 years of college.²¹ Figure 3 displays how the students' expectations vary based upon the highest reported level of schooling completed when interviewed.

²¹There is more variation within the 2001 cohort as the respondents were selected from the entire sample.

Figure 2: Expectations by Highest Grade Completed



(a) 1997 Sample

(b) 2001 sample

The figures show that expectations of attaining a college degree rise as students complete more education with the exception of 12th grade, or a high school degree. This seems to be the pivotal point for students in regards to their belief about obtaining a college degree. This is unsurprising as it is the point where many students decide whether or not to pursue a degree or other activities. However, if students are weighing this option in their initial expectations formation, then such a large discontinuity should not be present. It's plausible that students are realizing that the costs, both monetary and psychic, of obtaining a college degree are higher than anticipated. Another possibility is that they discover that the benefits of a high school degree are higher than they believed, which would serve to diminish the believed benefit of a college degree in comparison. This is controlled for and briefly examined in the empirical section below.²² Those who report having completed college at least four years of college in 2001 are not utilized in the models below.

Examining the expectations and completion of the 1997 cohort conditional on gender and race, we find that women have higher expectations than men by roughly 8 percentage points and also go on to complete college 30% of the time compared to just 21% for men. Whites and blacks have similar expectations, yet whites on average are nearly twice as likely to graduate college as blacks. Whites and blacks both have higher expectations and completion

²²There are a myriad of other factors that could also play a role in this revision process that will be explored in future work.

rates than do Hispanics. This pattern is also present when comparing across both race and gender. Overall, women have higher expectations and are more likely to graduate college than men of the same race. These relative patterns are also present for the 2001 cohort. Expectations appear to be heterogenous across race and gender, but are positively related to eventual attainment. The expectations and completion statistics for both cohorts used are broken down based on gender, race, and then race and gender in the table below.

Table 4: Expectations of 4 Year Degree Attainment by Race and Gender

		N	Mean	Median	Std. Deviation	Actual Completion	
1997	Male	1,751	68.14%	75%	32.49%	21.2%	
	Female	1,699	76.69%	94%	30.72%	29.7%	
	White	1,800	74.42%	90%	31.40%	33.2%	
	Black	934	73.42%	90%	32.24%	18.8%	
	Hispanic	716	66.45%	75%	32.11%	14.4%	
	White Male	918	68.72%	80%	32.54%	28.8%	
	Black Male	452	70.85%	80%	32.28%	15%	
	Hispanic Male	381	63.54%	70%	32.23%	10.5%	
	White Female	882	79.80%	95%	29.13%	37.9%	
	Black Female	482	75.83%	95%	32.06%	22.4%	
	Hispanic Female	335	69.75%	75%	31.70%	18.8%	
	Total	3,450	72.35%	85%	31.91%	25.4%	
	2001	Male	971	63.16%	80%	39.08%	23.2%
		Female	931	71.68%	95%	37.45%	31.9%
White		981	69.92%	90%	38.15%	38.2%	
Black		484	67.24%	90%	38.27%	17.6%	
Hispanic		437	61.62%	75%	39.10%	14.2%	
White Male		510	65.69%	85%	38.60%	32.5%	
Black Male		240	62.83%	80%	39.70%	12.9%	
Hispanic Male		221	57.69%	60%	39.12%	12.7%	
White Female		471	74.51%	99%	37.15%	44.4%	
Black Female		244	71.57%	95%	36.37%	22.1%	
Hispanic Female		216	65.64%	85%	38.76%	15.7%	
Total		1,902	67.33%	90%	38.52%	27.4%	

Table 5 presents similar statistics based upon the parent education level for the students. A strictly positive relationship exists between students' college expectations and the amount of education their parents have as well as for completion. Students whose parents both have a high school degree or less not only have the lowest expectations and completion rates, but

they also have the most variation in their beliefs. As the education level for students' parents increases, expectations also increase and less variation is present. For example, in the 1997 cohort the standard deviation of expectations for students' whose parents both have a college degree is approximately half that of students whose parents hold a high school degree or less (18.44% vs 34.35%). This difference is larger for the older, 2001 cohort. Parents who have experienced college can provide their children with certain insights as to the costs and benefits of a degree that a student whose parents have not completed college would not be able to provide. This could be in the form of making sure their children are prepared academically, what to expect the first year on campus, or even through helping them on assignments. Thus, students with more educated parents should have access to more information and therefore have more aligned expectations. The pattern could also be that they are less influenced by the monetary costs of college, as their parents more than likely have higher incomes and more certain they can fund their education. Further testing is needed to disentangle the effects.

Table 5: Expectations of 4 Year Degree Attainment by Parent Education

	Parent's Degree	N	Mean	Median	Std. Deviation	Actual Completion
1997	Both HS or Less	1843	64.28%	75%	34.35%	12.6%
	At Least 1 has Some College	1040	78.77%	90%	27.93%	31.6%
	Both Have Only Some College	177	81.44%	90%	24.78%	36.2%
	At Least 1 has a College Degree	810	86.42%	98%	22.05%	53.7%
	Both Have at Least College Degree	303	90.25%	100%	18.44%	68.3%
2001	Both HS or Less	970	56.27%	58.5%	39.99%	13.3%
	At Least 1 has Some College	622	74.26%	98%	35.92%	33.3%
	Both Have Only Some College	129	79.77%	100%	31.38%	33.3%
	At Least 1 has a College Degree	453	87.70%	100%	25.16%	58.7%
	Both Have at Least College Degree	181	94.33%	100%	14.80%	74%

Notes: Education level is for genetic parents as of 1997. Does not take into account if both of the parents are in the household or not when interviewed. The categories are based on the highest level of achievement for each parent so some overlapping occurs.

Lastly, I briefly discuss the relationship between the ASVAB, students' expectations, and post-secondary outcomes. The data reveal that a higher ASVAB percentile is associated with

both higher expectations and a higher likelihood of both attempting and completing a college degree. For example, those in the 1997 sample that are in the lowest ASVAB quantile have an average expectation of 62%, and while 30% of the this sample completes at least 1 year of college, only 6% go on to complete the degree. As you move up the quantiles, expectations and successful post-secondary outcomes increase accordingly.²³ This pattern is consistent for the 2001 sample. This roughly signals that students understand and incorporate the information tested by the ASVAB into their expectations, both of which are positively related to post-secondary schooling success. There are a number of other factors that are going into each and further analysis is needed. Tables comparing the ASVAB percentile to both expectations and achievement for both samples can be found in the appendix.

Examining the data, it is clear that at least on average students' expectations regarding the attainment of a four-year college degree are heterogeneous based on the highest level of previous academic achievement, gender, race, parent education, and ASVAB percentile. Yet these expectations also appear to be related with students' post-secondary educational outcomes in all instances. On average, those with higher expectations of obtaining a college degree are associated with a higher probability of both attending college and ultimately graduating with a four-year degree. However, it is plausible that after controlling for these, schooling success measures, and other factors identified by the literature that expectations will simply contain the information already included in these controls rather than be another source of information that can be used to predict post-secondary schooling success. The next section tests this hypothesis.

Estimation

I first examine college matriculation or attendance. Next, the completion of a four-year degree is analyzed. For attendance, students are given a 1 if they report completing at least

²³Per Donna Rothstein, the students received their results several months after taking the exam. The results included only their performance on each of the subject tests and did not include the percentile used in this data.

1 year of college by 2013 and 0 otherwise. This was done instead of using any credits received as reporting of credits in the NLSY97 is significantly worse than years of education.²⁴ For completion, those who report having obtained a Bachelor’s degree or higher by 2013 are given a 1 while those who report less are given a 0.²⁵ The estimates of both matriculation and completion will be via maximum likelihood using a probit model. The estimated marginal effects at the mean will be reported in the tables and discussed below.

Each model uses a set of covariates that have been previously identified in the economic literature, many of which are included in the literature review above, as potentially influencing or predicting the outcomes of interest. These include demographics; parent education levels; location; family characteristics; self-reported grades in 8th grade and/or high school; and the students’ peers’ plans for college. The specifics of each category are presented in the table below.

Table 6: Variable Descriptions

<u>Dependent Variables:</u>	i. Student completes at least 1 year of college ii. Student completes a 4 year college degree	
<u>Variable of Interest:</u>	The percent chance the student believes that they will obtain a 4 year college degree by the time they are 30	
<u>Control Variables:</u>	Demographics - Male - White - Black	Parent Education - One parent has some college - Both parents have some college - One parent has a college degree - Both parents have a college degree
	Family - Only child - Household members under 18 - Mother, Father, and both parents absent - Income quantile	Schooling: - Self-reported grades in 8th grade (levels) - Self-reported grades in HS if applicable (levels) -ASVAB Percentile -Enrollment in high school indicator
	Location - Urban - MSA Central City - MSA Non Central City - 3 Census Tracts (W, NE, NC)	Other: - % of peers that plan to go to college

Notes: For each categorical variable, that which is not listed is used as the base case. For grade controls, if a student is still in high school then the 8th grade controls are used, when they have graduated or left school the high school grade controls are used. In each case the student reported their grades in levels. Transcript data for high school gpa is available but low reporting severely reduces sample size. Estimates from specifications that have included it do not differ drastically.

²⁴If anything, this should bias the estimated effect of expectations on attendance downward.

²⁵The NLSY97 has a variable for the highest grade and degree reported ever. So, those who reported completing some college or completing a Bachelor’s degree in previous years that did not respond to the 2013 survey are still included.

Since there are two distinct samples, estimates are obtained using each separately. Recall, the 1997 cohort is composed of 15 to 17 year old students in grades ranging from 8th to 12th. Parents of these students are also asked to report the expectations of their child's probability of completing a college degree. These will be included and discussed in the sensitivity analysis section. The 2001 sample contains some students who are still in high school, college, or neither. Therefore, they will be used only in the completion models and indicators for enrollment in high school, a two-year college, and a four-year college will be included. Approximately 700 students are asked and answer the expectations in both round 1 and round 5. I will estimate the completion model first using their expectations in 1997 and then for 2001 to compare how the predictive power of expectations change as students age and transition out of high school.

Results

For both outcomes, estimates from three difference specifications are presented. The first includes only expectations on the right hand side. The next does not include expectations, but only the controls mentioned previously. The last combines the two above such that both expectations and the other controls are all included. The first two specifications are used to show that expectations and the previously identified controls by themselves have explanatory power. When the two are combined we observed how the coefficient on expectations changes and if the statistical significance is altered. Note, that while the NLSY97 lends itself to a panel analysis it is not being used as such for this paper. The subscript t in the models below is used to indicate the two different subsets of the data being used, and the subscript n denotes the time added until the student's outcome is observed between ages 30 and 32. These estimates suggest that expectations are highly related to both outcomes even when controlling for the various factors associated with college attainment, however they do not imply that a causal relationship exists as they are endogenous and a suitable instrument has

not yet been identified.²⁶ The models and the corresponding estimates are discussed below.

Matriculation

The dependent variable for matriculation is an indicator for whether the student ever completed at least one year of college.²⁷ The probit model below is used for the estimation where Φ represents the cumulative distribution function of the standard normal distribution and X' is a vector of that contains the previously mentioned controls. Since none of the respondents in the 1997 sample had completed high school at the time of their interview, their grades from 8th grade are used. When the 2001 sample is in the completion model, their high school grades are included instead as the majority had completed high school.

$$Prob(AttendCollege_{i,1997+n}) = \Phi(Expectations_{i,1997}, X'_{i,1997}\beta) \quad (1)$$

The related matriculation estimates for the 1997 sample are included in columns 1-3 in the table below. Expectations are found to be a positive and statistically significant factor in predicting the probability that a student attends college. The addition of the control variables decreases the estimated effect by approximately 46% which was expected due to its correlation with several of the other controls. However, it still suggests that, at the means, a 10 percentage point increase in expectations increases the probability that a student completes at least 1 year of college by 3.7 percentage points on average, ceteris paribus. I consider this economically relevant as decreasing the out-of-pocket costs for a student by \$1,000 is found to have a similar impact in the literature.²⁸ The addition of expectations also increases the percent of variation explained by the model from 26.1 to 28.5%. For simplicity, the other estimated coefficients have been suppressed but each had the hypothesized

²⁶Jacob and Wilder (2011) present several reasons why this is the case.

²⁷Ideally, this would indicate whether students simply attended college since a large portion of students that drop out do so after one semester or less. Unfortunately, reporting of individual college credits was incomplete for the majority of students so it could not be utilized.

²⁸See Manski and Wise (1983), Leslie and Brinkman (1988), Dynarski (1999), and Dynarski (2000).

influence excluding one of the race controls.²⁹

Table 7: College Expectations, Attendance, and Completion - Probit Model

	Y = 1 if completed at least 1 year of college		Y = 1 if completed a 4 year college degree			
College Expectations_97	0.00699*** (0.000312)		0.00374*** (0.000346)	0.00589*** (0.000300)		0.00244*** (0.000301)
Demographic Controls	No	Yes	Yes	No	Yes	Yes
Parent Education Controls	No	Yes	Yes	No	Yes	Yes
Location Controls	No	Yes	Yes	No	Yes	Yes
Family Controls	No	Yes	Yes	No	Yes	Yes
Grades in 8th Grade & ASVAB	No	Yes	Yes	No	Yes	Yes
Peer Controls	No	Yes	Yes	No	Yes	Yes
Observations	3450	3450	3450	3450	3413	3413
Pseudo R^2	0.125	0.261	0.285	0.123	0.316	0.332

Notes: Demographic controls include male, white, and black. Parent education include indicators for if one or both of the respondents' parents have some college or a four-year degree. Location controls include rural, 1 of 4 census tracks, and MSA. Family controls include number of members in the household under 18, whether student is only child, and indicators for if one or both parents are missing. Peer controls is the reported percent of peers that the respondent expects to go to college.

-Marginal effects; Robust standard errors in parentheses;

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Completion

The completion model is also estimated by maximum likelihood via a probit model. Similar covariates to the first model are used and, for similar reasons as above, three models are estimated and analyzed. For the 1997 cohort, all students are used as none had completed any college at the time they were asked to report their expectations. For the 2001 cohort, a model is estimated that uses the entire cohort and controls for enrollment.³⁰ The last specification utilizes only the 700 students who answered the questions in both periods to examine how the predictive power of expectations changes for students over time. Again, Φ represents the cumulative distribution function of the standard normal distribution, t

²⁹Whites were estimated to be less likely to attend college than Hispanics holding all other factors constant. This could be that after controlling for all of these factors there are more minority scholarships available that allow Hispanics to attend college at a higher rate. The western united states as also has large fraction of Hispanics that go and complete a two year college. This does not persist for the completion estimates.

³⁰A heckman selection model was attempted, but I was unable to identify a factor that influences attendance but not completion.

represents the cohort used in the estimation, and the other estimated coefficients have been suppressed for simplicity.

$$Prob(Complete4YrDegree_{i,t+n}) = \Phi(Expectations_{i,t}, X'_{i,t}\beta) \quad (2)$$

Columns 4 to 6 in table 7 above contain the estimates from the 1997 sample. When the controls are added the coefficient diminishes considerably but remains positive and statistically significant. Also, the estimated impact on completion is smaller than for attendance but not as much as one might expect. It suggests that a 10 percentage point increase in a student's expectations is associated with a 2.44 percentage point increase in the probability that they go on to complete a four-year degree. This is only about 1.5 percentage points lower than for attendance and is likely a byproduct of using completing at least 1 year of college as the measure for attendance considering the drop-out rate is much higher for students in the first year.³¹ Again, the model fit also increases when the expectations are included in the set of controls.

Table 8 below contains the estimates from two specifications that use the 2001 sample. The first, who's estimates are in columns 1-3, does not contain enrollment controls. I'll focus the discussion primarily on the the second specification which contains these controls, who's estimates are in columns 4-6, as it appears to be the best fit. In both cases expectations are estimated to be statistically significant predictors of college completion. The most comprehensive model suggest that a 10 percentage point increase in expectations results in an increase in the probability that the student completes college by 4.15 percentage points even when controlling for current enrollment status. The enrollment controls are each positive, statistically significant, and follow a logical sequence with those enrolled at a four-year college having a larger impact on future completion of a four-year degree than those enrolled at a two-year school. The same is true when comparing enrollment at a two-year college

³¹While it is only a 1 percentage point difference, it is nearly 35% smaller so there is a non-negligible difference.

and high school. When expectations are added, the estimated impact of these enrollment controls diminishes suggesting that some of the information included in the expectations variable is associated with current schooling status. Also, the estimated impact of expectations on completion from the 2001 sample is considerably larger than the one obtained from the 1997 sample which suggests that as students age and progress through schooling the predictive power of their expectations increase.

Table 8: College Expectations and Completion with 2001 Sample - Probit Model

Y = 1 if completed a 4 year college degree						
College Expectations_01	0.00719*** (0.000287)	0.00508*** (0.000322)	0.00719*** (0.000287)		0.00415*** (0.000358)	
Enrolled in 2yr College (d)					0.143** (0.0443)	0.0722* (0.0350)
Enrolled in 4yr College (d)					0.515*** (0.0365)	0.344*** (0.0415)
Demographic Controls	No	Yes	Yes	No	Yes	Yes
Parent Education Controls	No	Yes	Yes	No	Yes	Yes
Location Controls	No	Yes	Yes	No	Yes	Yes
Family Controls	No	Yes	Yes	No	Yes	Yes
Grades in High School	No	Yes	Yes	No	Yes	Yes
Peer Controls	No	Yes	Yes	No	Yes	Yes
Observations	1902	1850	1850	1902	1850	1850
Pseudo R^2	0.254	0.327	0.425	0.254	0.429	0.479

Notes: Demographic controls include male, white, and black. Parent education include indicators for if one or both of the respondents' parents have some college or a four-year degree. Location controls include rural, 1 of 4 census tracks, and MSA. Family controls include number of members in the household under 18, whether student is only child, and indicators for if one or both parents are missing. Peer controls is the reported percent of peers that the respondent expects to go to college.

-Marginal effects; Robust standard errors in parentheses; (d) for discrete change of dummy variable from 0 to 1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To confirm the above hypothesis, I estimate the completion model again using the 700 students who answered the expectations questions in both 1997 and 2001. In 1997 respondents were between 15 and 17 and thus of high school age. When asked five years later they

should have graduated and either proceeded into college or the work force. The estimates are included in table 9 below and loosely confirm the hypothesis from above. When the respondents are in high school, a 10 percentage point increase in their expectations is associated with a 3.32 percentage point increase in the probability that they complete a four-year degree. Five years later when they report their expectations, the estimate increases to 3.77 percentage points. Although the difference is only .45 percentage points, statistically the two estimates are not equal. These results indicate that the predictive power of expectations increases as students age and graduate from high school.

Table 9: College Expectations and Completion in High School, then After w/ Enrollment Controls - Probit Model

Y = 1 if completed a 4 year college degree						
	1997 Sample			1997 Sample 5 Years Later		
College Expectations_97	0.00686*** (0.000704)		0.00332*** (0.000756)			
College Expectations_01				0.00700*** (0.000559)		0.00377*** (0.000549)
Enrolled in 2yr College (d)				0.265*** (0.0708)	0.140* (0.0572)	
Enrolled in 4yr College (d)				0.634*** (0.0485)	0.386*** (0.0662)	
Demographic Controls	No	Yes	Yes	No	Yes	Yes
Parent Education Controls	No	Yes	Yes	No	Yes	Yes
Location Controls	No	Yes	Yes	No	Yes	Yes
Family Controls	No	Yes	Yes	No	Yes	Yes
Grades in 8th Grade	No	Yes	Yes	No	No	No
Grades in High School	No	No	No	No	Yes	Yes
Peer Controls	No	Yes	Yes	No	Yes	Yes
Observations	700	673	673	700	679	679
Pseudo R^2	0.135	0.337	0.358	0.299	0.532	0.575

Notes: Demographic controls include male, white, and black. Parent education include indicators for if one or both of the respondents' parents have some college or a four-year degree. Location controls include rural, 1 of 4 census tracks, and MSA. Family controls include number of members in the household under 18, whether student is only child, and indicators for if one or both parents are missing. Peer controls is the reported percent of peers that the respondent expects to go to college.

-Marginal effects; Robust standard errors in parentheses; (d) for discrete change of dummy variable from 0 to 1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The estimates above suggest that student expectations contain some form of private information that is not captured by the standard economic controls. These findings demonstrate that even at a young age students possess information regarding their potential for schooling success that econometricians have not considered. They also show that students gain relevant information regarding their true potential to complete a four-year college degree as they age. They use this information to form more accurate predictions of their ability to complete a

college degree. Below I test several other ways in which expectations might influence both college attendance and completion.

Sensitivity Analysis

This section examines other possible variations in which student expectations and post-secondary educational outcomes are related and what, if any, role parent expectations might play in student achievement. I test a quadratic and cubic relationship first, then follow this up with indicators at the heaping points of 0, 50, and 100%. Last, I include parent expectations in the models and discuss the findings. Again, I use a probit model to examine each outcome and discuss the marginal effects evaluated at the means.

Quadratic and Cubic Expectations

It is plausible that the relationship between student expectations is quadratic or cubic. Thus, I use both a quadratic and cubic specification to test this theory. Similar to above, I estimate the attendance model using only the 1997 sample and the completion model using both. The estimates from each are included in the the table below. For attendance, columns 1 and 2 contain the quadratic and cubic specifications respectively. The completion estimates for the 1997 and 2001 cohorts are in columns 3-4 and 5-6 respectively. For each outcome, across all specifications, the estimates suggest that a quadratic and cubic relationship is inappropriate.³²

³²The estimates from a linear probability model also suggest the findings above.

Table 10: Quadratic and Cubic Expectations

	<u>Y = 1 if completed at least 1 year of college</u>		<u>Y = 1 if completed a 4 year college degree</u>			
	<u>1997 Sample</u>		<u>1997 Sample</u>		<u>2001 Sample</u>	
College Expectations_97	0.00553*** (0.00138)	0.00142 (0.00338)	0.00231 (0.00134)	-0.000170 (0.00327)		
Expectations ²	-0.0000154 (0.0000114)	0.0000846 (0.0000775)	0.00000102 (0.0000104)	0.0000572 (0.0000688)		
Expectations ³		-0.000000621 (0.000000479)		-0.000000333 (0.000000404)		
College Expectations_01					0.00306 (0.00209)	-0.00252 (0.00525)
Expectations ²					0.00000926 (0.0000175)	0.000138 (0.000110)
Expectations ³						-0.000000759 (0.000000631)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Parent Education Controls	Yes	Yes	Yes	Yes	Yes	Yes
Location Controls	Yes	Yes	Yes	Yes	Yes	Yes
Family Controls	Yes	Yes	Yes	Yes	Yes	Yes
Grades in 8th Grade	Yes	Yes	Yes	Yes	No	No
Grades in High School	No	No	No	No	Yes	Yes
Peer Controls	Yes	Yes	Yes	Yes	Yes	Yes
College Enrollment Controls	No	No	No	No	Yes	Yes
Observations	3450	3450	3413	3413	1850	1850
Pseudo R^2	0.285	0.286	0.332	0.332	0.479	0.479

Notes: Demographic controls include male, white, and black. Parent education include indicators for if one or both of the respondents' parents have some college or a four-year degree. Location controls include rural, 1 of 4 census tracts, and MSA. Family controls include number of members in the household under 18, whether student is only child, and indicators for if one or both parents are missing. Peer controls is the reported percent of peers that the respondent expects to go to college.

-Marginal effects; Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Heaping Indicators

I anticipate that those who reported a 0, 50, or 100% chance of obtaining a degree differed significantly from those who did not due to over or under-confidence, a lack of understanding of probability, or because they lacked focus or dedication when answering the survey question. Again, I estimate the attendance model using only the 1997 sample and the completion model using both. The estimates are presented in the table below. The only marginally significant coefficient from the heaping indicators comes from the model estimating degree completion that uses the 2001 sample. In this case, those who report a 100%

chance of completing a four-year degree are estimated to be less likely to go on to complete it than those who do not. Specifically, those who report a 100% chance of obtaining a degree are 4.844 (.586-5.43) percentage points less likely to go on to complete the degree than those who report a 99% chance.

Table 11: Expectations Indicators

	Y = 1 if completed at least 1 year of college		Y = 1 if completed a 4 year college degree	
	1997 Sample	1997 Sample	2001 Sample	
College Expectations_97	0.00394*** (0.000567)	0.00292*** (0.000530)		
Expect=0_97 (d)	-0.0249 (0.0614)	0.101 (0.0785)		
Expect=50_97 (d)	-0.0392 (0.0306)	0.00247 (0.0304)		
Expect=100_97 (d)	-0.0439 (0.0287)	-0.0135 (0.0187)		
College Expectations_01			0.00586*** (0.000796)	
Expect=0_01 (d)			0.297 (0.166)	
Expect=50_01 (d)			0.0131 (0.0589)	
Expect=100_01 (d)			-0.0543* (0.0222)	
Demographic Controls	Yes	Yes	Yes	
Parent Education Controls	Yes	Yes	Yes	
Location Controls	Yes	Yes	Yes	
Family Controls	Yes	Yes	Yes	
Grades in 8th Grade	Yes	Yes	No	
Grades in High School	No	No	Yes	
Peer Controls	Yes	Yes	Yes	
College Enrollment Controls	No	No	Yes	
Observations	3450	3413	1850	
Pseudo R^2	0.286	0.333	0.482	

Notes: Demographic controls include male, white, and black. Parent education include indicators for if one or both of the respondents' parents have some college or a four-year degree. Location controls include rural, 1 of 4 census tracts, and MSA. Family controls include number of members in the household under 18, whether student is only child, and indicators for if one or both parents are missing. Peer controls is the reported percent of peers that the respondent expects to go to college.

-Marginal effects; Robust standard errors in parentheses; (d) for discrete change of dummy variable from 0 to 1
 -* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Parent Expectations

In the first round of interviews, both the selected youth and one of their parents each participated in a one hour long personal interview. During this interview they also asked the parent about the likelihood that the student would go on to complete a four-year college degree by the time they turned 30. It is possible that at such a young age, 15 to 17 when answering the questions, parents might possess more information regarding the probability that their child will go on to obtain a college degree. Since the parents are only included in the initial survey, only those students who answer in both periods will be available from the 2001 sample. For this reason I estimate four models. The first two use the full 1997 sample to examine matriculation and completion. The second two limit the sample to those who answer in both periods and focus on college completion. When the limited sample is used, the first specification includes the students' expectations reported in 1997, while the second includes their expectations in 2001. This was done to examine how the roles of each change as the student ages.

The estimates from the models containing parent expectations are below. The addition of parent expectations results in a number of changes. For attendance, the estimated predictive power of the students' expectations decreases by .07 percentage points from the initial estimates, but the overall model fit increases.³³ Also, the predictive power of parent expectations is nearly equivalent to the students'. Examining completion, when the parent expectations are included using both the full and limited 1997 sample, the estimated predictive power of the parents' expectations dominates the expectations of the student. However, when the student is asked 5 years later, the parents' expectations have zero predictive power while the students' expectations remain positive and statistically significant.

³³The .07 percentage point decrease is an 19% percent decrease.

Table 12: Parent Expectations and College Achievement

	<u>Y = 1 if completed at least 1 year of college</u>	<u>Y = 1 if completed a 4 year college degree</u>		
	<u>Full 1997 Sample</u>	<u>Full 1997 Sample</u>	<u>1997 and 2001 Respondents</u>	
College Expectations_97	0.00308*** (0.000420)	0.00195*** (0.000377)	0.00254** (0.000921)	
College Expectations_01			0.00384*** (0.000534)	
Parent Expectations_97	0.00276*** (0.000415)	0.00223*** (0.000364)	0.00294** (0.000917)	0.000971 (0.000637)
Demographic Controls	Yes	Yes	Yes	Yes
Parent Education Controls	Yes	Yes	Yes	Yes
Location Controls	Yes	Yes	Yes	Yes
Family Controls	Yes	Yes	Yes	Yes
Grades in 8th Grade	Yes	Yes	Yes	No
Grades in High School	No	No	No	Yes
Peer Controls	Yes	Yes	Yes	Yes
College Enrollment Controls	No	No	No	Yes
Observations	2938	2905	584	554
Pseudo R^2	0.304	0.348	0.382	0.580

Notes: Demographic controls include male, white, and black. Parent education include indicators for if one or both of the respondents' parents have some college or a four-year degree. Location controls include rural, 1 of 4 census tracts, and MSA. Family controls include number of members in the household under 18, whether student is only child, and indicators for if one or both parents are missing. Peer controls is the reported percent of peers that the respondent expects to go to college.

-Marginal effects; Robust standard errors in parentheses

-* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The findings suggest that parents also possess some valuable information regarding their child's schooling preferences and ability not captured by either the student's own expectations or the standard economic controls. The predictive power diminishes as the student ages which suggests that students are learning some of the private information that their parents have and are utilizing it when making predictions about their future success. Also, although student and parent expectations are highly correlated the addition of parent expectations does not drive the estimated predictive power of the students' expectations to zero.³⁴ This further indicates that students have some form of private information that is orthogonal to all the standard economic controls and the educational expectations of their parents when

³⁴Parent and student expectations are highly correlated. The correlation coefficient between the 1997 parent and student expectations is .57. It is .40 between parent expectations in 1997 and the students expectations in 2001.

it comes to predicting their post-secondary success.

Conclusion

The recent increase in the cost of attending college has led to a number of discussions, including how to reduce the costs and if a degree is still a worthwhile investment. However, the matriculation decision process and elements which contribute to both students' attendance and success are seldom discussed. I present new estimates of the factors that predict a student's decision to attend college and whether or not they complete a four-year degree that econometricians have not previously found. I focus on the predictive power of students' self-reported probability of obtaining a four-year college degree. Later I consider how the predictive power of these expectations vary as the student ages and progresses through different levels of schooling.

Manski (1993) outlined the importance of understanding and identifying how a student's expectations influence their decision to attend college or not. He notes that the estimated returns to education could be flawed when economists rely on the assumption of rational expectations instead of attempting to understand the process by which students use information to form expectations about the costs and benefits of a college degree. Dominitz and Manski (1996) and Manski (2004) both examine these issues, but are limited by small sample sizes and cross-sectional data. Jacob and Wilder (2011) are able to observe students over time and study the role of expectations in degree attainment, but do not examine whether these also influence the individual's decision to attend college. They are also limited by the nature of their data as their measure of expectations is imprecise.

A unique set of questions from the National Longitudinal Study of 1997 (NLSY97) allows for an examination of how a student's subjective probability of completing a college degree changes as they progress through schooling and what role it plays in their outcomes. A subset of students report their subjective probability of obtaining a four-year college degree

once while they are in high school and again when some have entered either college or the workforce. This is an improvement on the previous literature, as most other studies use data where students respond on a Likert scale regarding their educational expectations, or are simply asked the highest level of education they expect to complete without any notion of how confident they are in realizing that outcome.

The results indicate that, even after controlling for the factors previously identified in the economics and education literature, a student's self-reported probability of completing a four-year college degree plays a role in predicting whether they attempt to complete a college degree and if they are successful in obtaining a four-year degree. As students age and progress through more schooling, their expectations play a larger role in predicting completion. Also, parent expectations are estimated to be influential when students are young, but their predictive power diminishes as the student ages and gathers information related to the costs and benefits of a college degree. The estimates signal that, as Manski (1993) hypothesized, students possess private information, captured by their expectations in this case, that is helpful in determining post-secondary schooling attendance and success that isn't being captured by the previously identified controls, and suggests that youth and econometricians may possess and utilize different data when making schooling decisions.

Understanding the process by which students both form and update their expectations is essential to understanding the underlying role that these expectations play in the attainment process. The use of interpretable subjective data on expectations, combined with choice data, can help solve the identification problem in the returns to education literature by relaxing the rational expectations assumption. To accomplish this, a study that examines how students form and update their expectations about whether or not to attend college is necessary and forthcoming. Lastly, if the information that makes students more accurate predictors of their future success can be identified, it is possible we could also increase the matriculation and graduation rates at post-secondary schools.

References

- Bailey, Martha J and Susan Dynarski**, “Gains and gaps: Changing inequality in US college entry and completion,” Technical Report, National Bureau of Economic Research 2011.
- Bernheim, B Douglas**, “How do the elderly form expectations? An analysis of responses to new information,” in “Issues in the economics of aging,” University of Chicago Press, 1990, pp. 259–286.
- Black, Dan, Seth Sanders, and Lowell Taylor**, “Measurement of higher education in the census and current population survey,” *Journal of the American Statistical Association*, 2011.
- Bollinger, Christopher R**, “Measurement error in human capital and the black-white wage gap,” *Review of Economics and Statistics*, 2003, 85 (3), 578–585.
- Bureau of Labor Statistics, U.S. Department of Labor**, *National Longitudinal Survey of Youth 1997 cohort, 1997-2011 (rounds 1-15)* Produced by the National Opinion Research Center, the University of Chicago and distributed by the Center for Human Resource Research, The Ohio State University 2013.
- Dominitz, Jeff and Charles F Manski**, “Eliciting student expectations of the returns to schooling,” *Journal of Human Resources*, 1996, 31, 1–26.
- **and** — , “Using expectations data to study subjective income expectations,” *Journal of the American Statistical Association*, 1997, 92 (439), 855–867.
- Dynarski, Susan**, “Hope for whom? Financial aid for the middle class and its impact on college attendance,” Technical Report, National bureau of economic research 2000.

- Dynarski, Susan M**, “Does aid matter? Measuring the effect of student aid on college attendance and completion,” Technical Report, National bureau of economic research 1999.
- Fischhoff, Baruch, Aandrew M Parker, Wändi Bruine de Bruin, Julie Downs, Claire Palmgren, Robyn Dawes, and Charles F Manski**, “Teen expectations for significant life events,” *Public Opinion Quarterly*, 2000, pp. 189–205.
- Forbes**, “America’s Top Colleges: 2016 Ranking,” 2016.
- Hilmer, Michael J**, “Post-secondary fees and the decision to attend a university or a community college,” *Journal of Public Economics*, 1998, 67 (3), 329–348.
- Hoxby, Caroline M and Christopher Avery**, “The missing “one-offs”: The hidden supply of high-achieving, low income students,” Technical Report, National Bureau of Economic Research 2012.
- Hoxby, Caroline, Sarah Turner et al.**, “Expanding college opportunities for high-achieving, low income students,” *Stanford Institute for Economic Policy Research Discussion Paper*, 2013, (12-014).
- Hurd, Michael D and Kathleen McGarry**, “The predictive validity of subjective probabilities of survival,” *The Economic Journal*, 2002, 112 (482), 966–985.
- Jacob, Brian A and Tamara Wilder**, *Whither Opportunity? Rising Inequality, Schools, and Children’s Life Chances*, Russell Sage Foundation, 2011.
- Leslie, Larry L and Paul T Brinkman**, *The Economic Value of Higher Education. American Council on Education/Macmillan Series on Higher Education.*, ERIC, 1988.
- Lochner, Lance**, “Individual perceptions of the criminal justice system,” *The American Economic Review*, 2007, 97 (1), 444–460.

- Manski, Charles F**, “Adolescent econometricians: How do youth infer the returns to schooling?,” in “Studies of supply and demand in higher education,” University of Chicago Press, 1993, pp. 43–60.
- , “Measuring expectations,” *Econometrica*, 2004, 72 (5), 1329–1376.
- **and David A Wise**, *College choice in America*, Harvard University Press, 1983.
- Neal, Derek A and William R Johnson**, “The role of pre-market factors in black-white wage differences,” Technical Report, National Bureau of Economic Research 1995.
- Perna, Laura Walter**, “Differences in the decision to attend college among African Americans, Hispanics, and Whites,” *Journal of Higher Education*, 2000, pp. 117–141.
- Smith, V Kerry, Donald H Taylor, and Frank A Sloan**, “Longevity expectations and death: Can people predict their own demise?,” *The American Economic Review*, 2001, 91 (4), 1126–1134.
- Zafar, Basit**, “College major choice and the gender gap,” *Journal of Human Resources*, 2013, 48 (3), 545–595.