An Analysis of the Labor Market Returns to Community College and Vocational Training

April 14, 2023

Author	Supervisor
Eli Levine	V. Joseph Hotz
Duke Department of Economics	Duke Department of Economics
Undergraduate Student	Professor

Abstract

Education and training are fundamentally linked with labor market performance. There is a significant body of work analyzing the role of education in wages with an emphasis on a comparison between a college degree and a high school diploma. However, as states have begun to shift their education policies to make community college and trade school more accessible, it is important to understand the expected labor market returns to these forms of education. In this paper, using data from the National Longitudinal Survey of Youth's cohort that began in 1997, the returns for different levels of education using the Mincer equation are found. While there was a data limitation surrounding trade school, it was possible to analyze the impact of adding a vocational license or a training certificate to a high school diploma. When controlling for experience in three different ways, specifically by age, time at highest training and labor market experience, it was found that returns to a training certificate relative to high school are between 18.7% and 36.3% higher than a high school diploma. Furthermore for community college, the wage returns are between 26.4% and 45.8% higher relative to a high school diploma. These findings highlight that additional training and certification can be an effective tool for increasing labor market returns for high school graduates even without a bachelor's degree.

1 Introduction

The relationship between education and income is frequently analyzed. Education is often considered as a way for a person to improve their respective labor market performance (Houthakker, 1959). Much of this research has been concentrated on the economic returns to university degrees, especially when compared to a high school diploma. A university degree has remained a significant prerequisite for many opportunities, especially due to the relative scarcity of university graduates. In 2021, 91.1% of Americans had completed a high school education or a GED, while only 37.9% of Americans had a university degree (Alonzo, 2022).

However, the emphasis on this difference has led to other forms of education and certification being less focused on academically. In this paper, the author is considering the question of what are the labor market returns to community college and vocational training. As more Americans, especially low income citizens, attend trade schools and pursue these types of certification, it is important to understand what the returns are relative to a more traditional path of high school and university (Holland & Deluca, 2016). In this article literature surrounding the returns to community college, trade school and vocational training are analyzed, and a methodology centered around the Mincer equation is used to display what the associated returns to these types of training are for Americans currently.

2 Literature Review

The decision to attend college can be a life-changing course of action. A college degree is often seen as a silver bullet that enables a higher standard of living. There are two dominant perspectives on the value of college education, specifically the human capital theory and the signaling theory.

The human capital theory, outlined by Dr. Gary Becker in Investment in Human Capital: A Theoretical Analysis, argues that college is an investment in human capital. Therefore, it is seen that earnings over a lifetime are higher for those who have attended college, even when considering the cost of a four-year degree (Becker, 1962). In this paper, Becker laid the foundation for considering a college degree as an investment in one's ability to specialize and thus receive a premium on earnings associated with this extra training.

The signaling theory was developed by Dr. Michael Spence and explained in Job Market Signaling. This theory states that there is an issue of information asymmetry between employers and potential employees. Employers do not know how skilled an employee is, and by extension do not know their productivity. Thus, a degree operates as a signal to employers that an applicant is competent (Spence, 1973). While these theories differ in their mode of delivering a higher standard of living, they agree that a college degree is an investment that provides strong returns in the labor market. This encourages us to look deeper and ask which barriers prevent people from obtaining a college degree.

Regardless of which theory is more convincing, there is little theoretical debate on whether education has positive long-run benefits on earnings. To test if this expectation is met, Dr. Michael Hout analyzed the social and economic returns of a college education. He found that, even beyond what we would expect from a degree operating only as a signal, education improves every aspect of life. He founds there are substantial personal and social returns to attending college, including improvements in one's health, bringing out more productivity in coworkers and that the present value of the wage increases continues to outweigh the costs (Hout, 2012). Thus through Dr. Hout's work, it is clear that education is a good indicator of personal welfare, and can often be responsible for a higher quality of life.

However, by looking broadly at the impacts of college education, Hout does not focus enough on the marginal student. While on average, college graduates are more likely to enjoy a higher standard of living, it is important to understand which benefits exist for students who struggled academically in secondary school and thus were accepted to a college by being only slightly over a cutoff for admittance. Thus Dr. Seth Zimmerman looked at this hypothetical marginal college student in Florida. By trying to identify college's effectiveness at improving people's income over their lifetime, Zimmerman analyzed students accepted to university who resided on an admission fringe. This was his strategy for identifying the marginal student. Florida International University, a representative university in the Florida State University System, has a more generous way of calculating GPA so students who qualified for a tuition subsidy program by being the first in their family to attend college were more likely to be admitted to FIU than any other school in the Florida public school system. Zimmerman found that this marginal college student, who was admitted under the generous approach by FIU, still earns 22% more than their counterpart who were just below the threshold (Zimmerman, 2014). It is worth noting that a comparison between all high school graduates and college graduates found a 50% income premium associated with a degree. This reveals that college is not as effective for everyone but attending still provides a clear benefit compared to not attending.

Thus, due to the clear benefits associated with attending college, education can be a means of generating economic mobility. In industrialized nations, to understand the relationship between inequality and mobility we have to understand more specifically the institutions that operate in that country (DiPrete, 2020). Therefore, this paper recognizes the value in understanding a more expansive view of the institution of education and training.

While the benefits of a marginally capable student attending college are clear, it is worth considering that colleges are not uniform in their quality. There are significant differences, both in academic curricula offered, and in resources provided to students. Thus, it is important to understand the distinction between an elite college and a standard college. Elite colleges are identified by their rigor and breadth of resources. Elite colleges have not increased their class size proportional to the larger number of applicants. This selectivity has come with clear economic benefits. Graduates of elite colleges earn 12% more than graduates of an average university (Hout, 2012). This premium for elite colleges gives high school students an element beyond prestige to strive for in their applications. A more thorough analysis of the distinction between graduates of elite institutions and those who graduate from other academic institutions, displays that SAT scores and high school GPAs are heavily correlated with future earnings, and due to the selective nature of elite academic institutions, their student body consists almost entirely of this demographic (Dale & Krueger, 2011). However, when controlling for average SAT scores for admission at each college it appears that an institution's status has little effect on earnings. Nonetheless, there is a notable exception, namely for students whose parent's received little education, Hispanic and Black students. These three groups all see much larger than expected economic returns after graduating from a prestigious university (Dale & Krueger, 2011).

However, despite this premium, elite colleges have yet to be as effective at promoting widespread intergenerational mobility. Access to attendance at an elite university is usually a luxury for children of rich parents. Children whose parents are in the top 1% of income earners are 77 times more likely to attend an Ivy League institution (Chetty et al., 2017). Despite the unlikelihood of attending an elite institution, if an individual is not born into wealth, these colleges are the most effective at helping people from the bottom quintile reach the top 1% of income earners in the United States (Chetty et al., 2017). However large public universities are incredibly effective at promoting this mobility, especially in California where 20% of each undergraduate class must come from the bottom quintile. Furthermore, regardless of family income before attendance, students tend to have comparable earnings with their peers after graduation (Chetty et al., 2017). The discrepancy between elite institutions and average colleges is noteworthy because it appears necessary that a large barrier to this mobility is the increased likelihood of attending a premier institution given one's parent's class.

While it is clear that four-year colleges are effective at promoting mobility, it is worth analyzing some of the other forms of education, specifically community colleges and trade schools. For many community college students, community or junior colleges are a more affordable pathway toward higher education. While critics argue that students who attend junior college usually stop after their Associate's Degree, Dr. Cecilia Rouse found that there is a relationship between living near a community college and pursuing more years of education (Rouse, 1995). This means that the community college system is able to accomplish what it is designed to do, specifically promote additional avenues of training.

However, assuming that the more academically qualified students do transition to a four-year degree program upon completion of their schooling, this narrows the interpretation of the results for returns to community college. In an analysis of labor market returns for a 2- year degree compared to 4-year degrees it was found that each credit faces similar opportunity costs (Kane & Rouse, 1993). This means that when hiring, Rouse found that there was a proportional relationship between wages received by applicants with a 2-year degree and a 4-year degree. Employers did not penalize attendees of a community college While this does lend support for the human capital model of education, it still implies that community college students should still earn substantially less than their peers who attend a four-year institution.

Furthermore, since community college often includes students who will transfer to an institution that grants four-year degrees, it is worth analyzing those who attend when they are capable of pursuing a four-year degree and those who attend community college when they would otherwise halt further human capital develop after completion of high school. Dr. Jack Mountjoy analyzed this "democratized" (increasing access to education) and "diversion" (diverting students from a four-year institution) population of junior college students. He found that community college had positive return to both income and mobility for the democratized and negative returns for the divergent population (Mountjoy, 2022).

An analysis of trade schools and vocational programs shows that returns are not as clear or definite as graduates of a community college or four-year academic institution. In many central European countries, it is common that while in secondary school, a student might decide to pursue a vocational path instead of an academic one. This provides a successful environment to compare the returns of vocational education with an academic curriculum. While initially, those who attend vocational programs see higher returns, over a lifetime these are diminished (Hanushek et al., 2016). This is largely due to many skills taught in vocational schools being automatable, so while a standard post-secondary path provides a skillset that can be re-applied to a number of different industries, many vocational programs focus on a single skill which might not translate well in a labor market affected by automation (Hanushek et al., 2016). Thus Hanushek argues that a college degree prepares student to address a variety of potential problems their employers might face, while trade school does not provide similarly effective training, creating someone skilled in a specific issue.

In spite of the relative scarcity of literature surrounding trade school in the United States, there has been some analysis surrounding vocational licenses or training certificates. For example, in an analysis of the WorkAdvance program, which was designed to provide vocational training to low income workers, it was found that this training not only significantly increases employment rates, but also income (Katz et al., 2022). Additionally further work has shown that when workplaces provide training, their employees are more likely to continue this independent training regimen on their own accord (Osterman, 2022). These two sources make it abundantly clear that training certificates can be an effective means at improving one's labor market returns.

In this paper, the author focuses on the returns to the different levels of education. This will be done by comparing the labor market returns to other levels of education with community college and trade schools when considering the demographic profile of students. In the United States, there is not a significant amount of work relating to the returns of attending vocational training, nor much analysis of the returns of an Associate's degree. This paper hopes to fill this gap in the literature.

3 Methodology

In order to analyze the labor market returns to community colleges and vocation specific training, it is important to understand how the returns to education broadly are measured. When estimating the rate of return to schooling, the typical study uses a wage differential equation. A rudimentary example of such an equation is discussed in chapter 4 of Dr. George Borjas's Labor Economics textbook:

$$\log(\mathrm{w}) = \beta \text{ s} + \text{other variables}$$

Where w represents wages, β is the coefficient that represents the percentage change in wages for one additional year of schooling, and s signifies the number of years of schooling (Borjas, 2008). Most of the time the relevant "other variables" represent demographic or geographic factors that will affect wages but are not captured by the number of years of schooling. It should be noted that "β" implies that students face the same wage-student locus, meaning that they are of comparable ability. This education-centric model for wages would tell us the benefit of one extra year of schooling for workers of comparable stature (Borjas, 2008).

While this elementary model is potentially useful as an introduction to the relationship between wages and education, it is by no means the standard model of analyzing the income returns to education. Dr. Jacob Mincer recognized in the middle of the twentieth century that the present value of one's lifetime earnings is a function of their training and their experience. However, Mincer identified that this attributed income disparities disproportionately to age and experience in the labor market (Mincer, 1958). Thus in his book, Schooling, Experience, and Earnings, Dr. Mincer further expanded this model in consideration with the data accessible at the time, stating that

$$\ln(E_t) = \ln(E_0) + \mathrm{rs} + \beta_1 \mathrm{t} + \beta_2\, t^2$$

where E_t is a person's earnings at time t, E_0 is the earning capacity without the presence of schooling (this variable essentially represents one's earnings as a function of ability), "rs" is the rate of return for each additional year of schooling, and t represents time in the labor market (Mincer, 1976). Further model specification has been done to calculate E_0 as the capacity for one's earnings without the presence of education, considering demographic variables as well as ability.

A model of wage differentials was applied in an analysis of vocational programs in Europe. Hanushek and his co-authors found that when comparing those placed in a "school-to-work" apprenticeship program initially saw increases in both employment and wages, measured by a least squares regression equation and by a Mincer equation respectively. However, over time those in a general education program caught up to and surpassed their colleagues in both factors. The researchers hypothesized that this is in part due to general education recipients being more adept in retraining programs that might be necessary at later stages in life (Hanushek et al, 2016).

However, wage differential models and the Mincer equation are not without their critics. Dr. Robert LaLonde expressed that the most appropriate way to compare the impact of training or education is only by analyzing experiments. Thus he analyzed data from a natural experiment conducted by the United States for a program called the National Supported Worker Demonstration (NSWD). Participants were either randomly assigned to receive training (treatment), while others were placed immediately in employment (control). The treatment and control groups did not have statistically significant differences in age, education, wage or hours worked before the experiment. It was found that men in the treatment group made around $886 more than those in the control section by the end of the study, while the difference for women was $851 (LaLonde, 1986). When trying to recreate these findings using econometric methods at the time, LaLonde found that models overstated the training program's impact. The only models that were close to predicting the impact of the program violated the least squares assumption of specificity (LaLonde, 1986). While this random assignment provides rationale for an experimental approach toward analyzing the returns to education, it ignores the potential that non-experimental measures can have in replicating the findings of the experimental approach. For example certain characteristics were involved in a person's decision to enroll in the NSWD initially (Heckman & Hotz, 1989). This paper and its ensuing critique are relevant because they highlight the importance of considering all potential demographic information in order to determine the impact of training on earnings.

A common method of analyzing non-experimental data is through the use of instrumental variables. The instrumental variable method involves using a variable that is correlated with an independent variable but not with the error term of an initial equation. It is used in order to address issues surrounding endogeneity of a variable. When using an instrumental variable, the coefficient applied to the independent variable is the covariance between the instrumental variable and the dependent variable divided by the covariance between the independent variable and the dependent variable (Wooldridge, 2012). This approach has been used repeatedly to identify the returns to education.

In an analysis of the returns to community college, Dr. Thomas Kane and Dr. Cecilia Rouse employed ordinary least squares to longitudinal datasets in order to find the expected wage return per credit (Kane & Rouse, 1993). It is worth considering that in their analysis they used different sets of controls during different phases of their analysis. Firstly, they looked at the relationship between wage years after graduating high school and education measured by the number of credits when controlling for the region of the country, work experience, and race. Next, they added controls for personal background such as family income, test scores, and class rank. In their final model, they used education level as an instrumental variable to address that different schools measured different credits differently (Kane & Rouse, 1993).

Dr. Kane and Dr. Rouse's idea to employ instrumental variables to signify education was applied to the Mincer-style equation in a working paper prepared by the IZA - Institute of Labor Economics. Dr. Christopher Jepsen, Dr. Kenneth Troske, and Dr. Paul Coomes looked at administrative data from a community college in Kentucky that offered three different types of signals of program completion, specifically associate's degrees, certificates, and diplomas. Certificates usually only consist of 1-2 semesters of coursework and are primarily awarded in technical programs such as electrician and IT Network administration, while diplomas take around a year to complete and are usually found in technical fields, such as surgery technology, accounting, and nursing (Jepsen, Troske, & Coomes, 2012). In order to evaluate the returns to these three different certification types, the researchers used a simple and multivariate regression. The simple regression used earnings as the dependent variable with demographic information and an instrumental variable indicating what award they received for their studies. In their multivariate approach, they try to analyze the fixed effects of a student by comparing those who completed their award to those who dropped out of their program. They added in this multivariate regression a coefficient for enrollment which is a function of the opportunity cost of attending, increase in earnings, wages two quarters prior to attendance, and wages one quarter prior to attendance. They also attended an instrumental variable called intent which measures a student's program intent at the beginning of the program (Jepsen, Troske, & Coomes, 2012). When analyzing the returns to different degree types, especially for those within the same institution these can be useful classifications in order to represent the variability of academic paths in community colleges.

A similar style of analysis was used by Dr. Di Xu and Dr. Madeline Trimble, in their analysis of community colleges in North Carolina and Virginia. They used a similar approach of analyzing the different completion awards as instrumental variables, using fixed effects to track similar students, as well as a function that considers the opportunity costs for each individual enrolling (Xu & Trimble, 2016). These fixed effect models have been shown to be more effective at representing the labor market returns to these kinds of degrees than a Mincer equation because of how fixed effects represent biases that result from time-invariant characteristics (Xu & Trimble, 2016). The researchers discuss that the standard Mincer equation consistently provides a smaller approximation for both the short-term and long-term effects of these non-degree awards. This lends credibility to the idea that fixed-effect regression equations might more accurately portray the labor market returns of non-degree awards from community colleges.

To estimate the returns to education, other regression variants have been employed. For example, in his analysis of how effective community colleges are in promoting upward mobility, Dr. Jack Mountjoy, used a two-sided sum of least squares regression equation. Dr. Mountjoy's equation used the returns to different community college awards where the independent variables were the dependent variables of regression equations based on the democratization and diversion effect of community colleges. This enabled him to assess the difference between the benefits and the costs of community college (Mountjoy, 2022).

Furthermore, in order to analyze the returns to education for a marginal student, Dr. Seth Zimmerman employed a regression discontinuity to identify the increase in earnings for the marginal student relative to the marginal non-student when analyzing students from Florida (Zimmerman, 2014). A regression discontinuity allows one to analyze the causal effects since those directly below and directly above are comparable. When comparing individuals who have a trade school certificate or community college degree, once controlling for ability, a regression discontinuity would be an effective tool for comparing the marginal benefit of different certifications.

4 Data

4.1 Data Preparation

In order to analyze the labor market returns to differing levels of education, the National Longitudinal Survey of Youth, specifically the 1997 cohort was used. The National Longitudinal Survey of Youth is a dataset composed by the Bureau of Labor Statistics that follows a cohort of the American people over period of time. This dataset is nationally representative and consists of people who were between the ages of 12 and 16 before the start of 1997. This dataset was used because among the survey questions included, information concerning not only demographic features, such as ethnicity and gender, but also education and annual income is available. However, the data provided in this survey required some transformations.

The first section of transformations pursued involved recoding many of the variables in the data frame. In many cases, the manner in which the data is collected for survey purposes is not applicable for a regression model. For example, in ethnicity, 1 indicates whether someone is Black, 2 represents Hispanic, 3 is a mixed racial or ethnic identifier and 4 is all other races or ethnicities. Thus, these variables had to be recoded into four distinct variables that were 1 if their observation was consistent with that category and 0 otherwise. This transformation had to be pursued for Sex as well.

The next step involved generating binary variables for education, because, similar to previously discussed variables, education was stored in a manner convenient to collecting a survey. In order to generate similar binary variables for education, a few more steps were necessary along with certain assumptions. This is because, in order to represent each observation's most relevant training to their employment, it was assumed that only their highest educational attainment should be considered. This decision was made under the assumption that for an individual their highest level of training was most relevant to their employment, and by extension their wages. In most cases, a person with a Bachelor's degree will have a High School degree. Thus, in order to represent one's highest level of education, the education variable for the final round of interviews, which were conducted throughout 2019, and sometimes into early 2020, was one-hot encoded. However, this dataset did not contain trade school as a designation for education, only whether an individual had ever completed either a training certificate or received a vocational license (it does not specify which or in what discipline). Thus, to satisfy the assumption that only one's highest level training was relevant to their wages, if someone had completed either a training certificate or vocational license, and their highest level of education was either no degree, high school or a GED, then their highest level of training relevant to their employment is likely that certification. While this means that the returns to trade school specifically are inaccessible, it is possible to compare the labor market returns to those who received this training in addition to a high school diploma to those in the labor market without additional training.

In order to accurately assess the returns for each additional training, it is important to control for each observation's skill level. Someone with a Bachelor's degree likely has more skills than a high school dropout, so it is important to control for their skill level in order to isolate the impacts of their respective training. This was controlled for by incorporating each observations ASVAB score in 1999. This metric was chosen as it is a standardized test that the majority of the respondents took and represents their percentile score relative to each other.

The final transformation that was necessary for all of the models prepared was to log-transform income in the final time period, specifically, annual income in 2019. This was done in accordance with the Mincer equation, and it has the benefit of minimizing the skewness of the income data so that it was more likely to follow a normal distribution. Since the dependent variable has been normalized the linearity assumption can be applied for the regression models pursued throughout this analysis. However, to log-transform income in 2019, it was necessary to remove all N/A observations in the dataset for that variable. These not applicable observations are due to a variety of reasons. The first is refusal, which occurred 66 times. The second is that the respondent did not know their income. In this case, they were asked a follow up question to estimate the range. This estimation was input into the income in 2019 column. The next two reasons are “valid skip” and “non-interview.” It can be assumed that a valid skip meant that they did not work in the time period. These observations also had to be removed. Additionally, if someone was not interviewed, their observations had to be removed. The scope of the analysis was narrowed to only those who participated in the labor market in the final time period and had available data. It is important to note that this process biased the results as only observations that worked in the last time period were counted. However, despite this selection bias, it is important to consider active participants in the labor market to analyze the returns to different levels of training.

The model employed in this analysis was the Mincer equation. After recoding each demographic variable, it became possible to employ this model in order to determine the extent of the correlation between each level of education and the log of income. The first model only considers how the education variables alone are correlated with income. The second model represents demographic information, education level and an individual's skill (as represented by their ASVAB score in 1999) as independent variable and the log-transformation of income as the dependent variable. The third model includes their age to control for some experience, as well as their age squared in order to normalize the value. The fourth model replaces age with the amount of time since each individual in the dataset obtained their highest level of education. This variable was created by using the months having each level of education (this variable was provided in the dataset), as well as the date of a training certificate in order to find the age at which each observation received their highest training. This age was then subtracted from their age on the day of their interview in 2019. This model is used because it indicates the level of experience they have achieved in their field after training. The longer someone has been in their highest training, the more time they have had to refine their skills, knowledge, and expertise, which can contribute to better job performance and career advancement.The fifth and final model uses labor market experience up to 2019. This variable was more computationally expensive to generate as more often than not people would leave the labor market for periods of time in order to pursue education. Thus, the approach identified each age in which an observation received more than $10,000 per year (approximately the 10th percentile of the income distribution) and counted any observation over that as one entering the labor market. This eliminated the observations where one would earn money unrelated to their training while pursuing other education (specifically while they are a high school student). Then the difference between their age at this observation and the final observation represented their time in the labor market. There are some limitations to this approach which will be discussed in the appropriate section.

4.2 Exploratory Data Analysis

The first aspect of the data analyzed was whether or not a training certification or vocational license was a terminal degree. Thus, the data was subset to focus on the population that had received vocational training at any point of their career. Then, the age at which each observation that had received community college and/or university education was subtracted from the age at which that observation had received training. If the value was negative, this meant that the respondent was younger when they received their training than they were at their next stage in education, showing that a training certificate was not terminal. For those who had pursued training and had an associate's degree, 55.6% received their vocational training before their degree. For those with a bachelor's degree, the number was 44.8%. These results suggest that for those who received either an Associate's degree or a Bachelor's degree, about 50% received further vocational training later in their career. Additionally it highlights an opportunity for future research that analyzes the returns for vocational training for those who have a degree.

The next step was to generate a table displaying summary statistics for each degree type in the dataset. The below table includes the median, the mean, the standard deviation, and the count of observations for each degree type. There is a large standard deviation of the data at each level of training, however, that is likely a factor of the shape of the distribution. Additionally, it is clear that further training increases the mean of income, with the exception of the increase between a master's degree and a doctorate. The mean high school graduate earns \$37731.28 per year, which is less than the mean of a high school graduate with a training certificate and the mean of an associate's degree holder who earn \$50,266.77 and \$51,556.84 respectively.

	Degree Type	Median	Mean	Standard Deviation	Count
1	Drop Out	26000.00	33254.59	31620.42	275
2	High School Diploma	33000.00	37731.28	28606.56	495
3	Training Certificate	40000.00	50266.77	44166.18	1378
4	Associate's Degree	42000.00	51556.84	43612.76	517
5	Bachelor's Degree	58000.00	71521.00	56592.65	1203
6	Master's Degree	70000.00	81457.16	56239.38	541
7	Doctorate	72000.00	76300.24	37968.63	55
8	Professional Degree	130000.00	174632.52	114794.35	95

Table 1: Summary of Income for Each Degree Type

After looking through the summary statistics between income and degree type, it is important to visualize the distribution. Thus, the next step of the exploratory data analysis was to graph the appropriate distributions. The following graphs represent the distribution of income for the entire dataset, followed by subdivisions for high school graduates and those with vocational licenses, associate's degrees and bachelor's degrees.

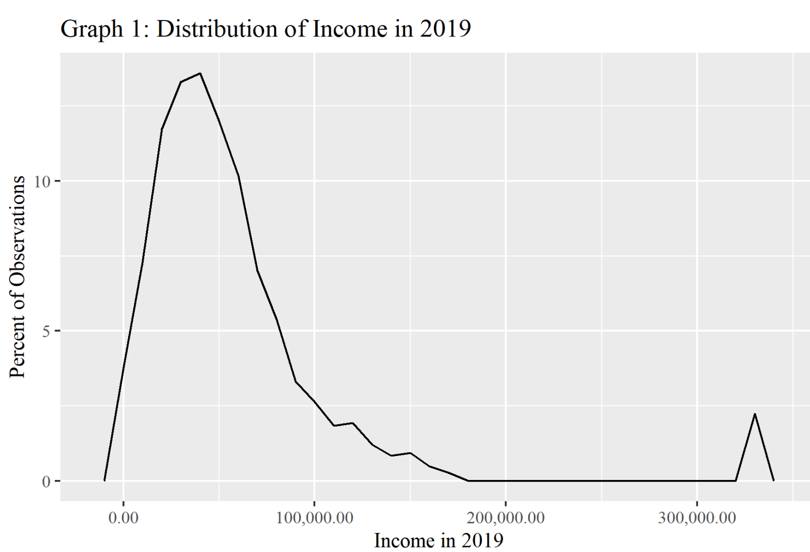

Figure 1: The Frequency of Income of the Respondent in 2019

The distribution of income in 2019 is displayed in the graph above. As is expected the graph has a right skewed distribution, as earnings are clustered toward the lower end of the distribution. The mean of this distribution is \$57,600 and the median is \$46,000. The standard deviation of this data is \$52,491. The first quartile of the distribution is at \$28,000 and the third quartile is at \$70,000. The following graphs subset this data by different levels of training, specifically, high school, trade school, associate's degrees and bachelor's degrees.

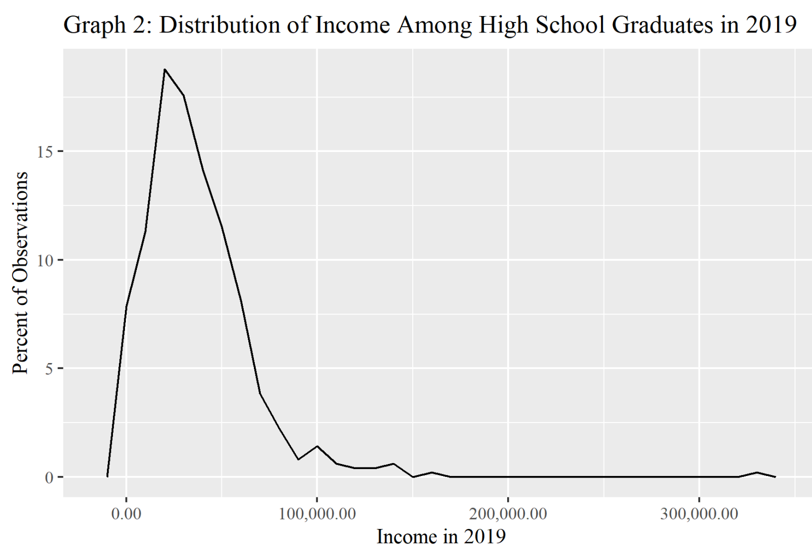

Figure 2: The frequency of Income in 2019 Among those Whose Highest Training is High School

Similar to the graph of total income in the dataset, the income of high school graduates has a right skewed. The mean income is \$37,731.28 and the median income is \$33,000. The standard deviation of this distribution is \$28,606.56. As is expected, the graph of high school graduates has both a lower median and mean than the total population.

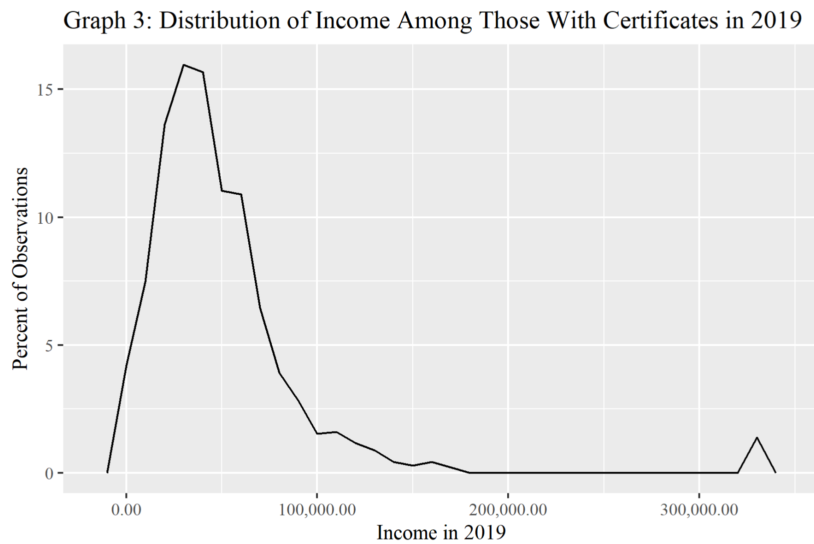

Figure 3: The frequency of Income in 2019 Among those Whose Highest Training is a Training Certificate or Vocational License

The distribution of income of those with a high school education and a vocational license or training certificate is shown in the above graph. Similarly, the data has a right skew. The mean and median income among respondents in this criteria is \$50,266.77 and \$40,000 respectively. The standard deviation of this distribution is \$44,166.18.

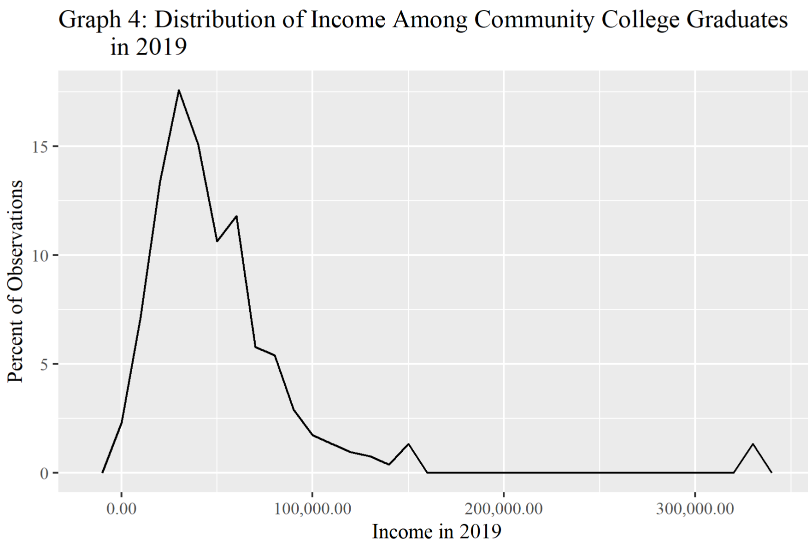

Figure 4: The Frequency of Income in 2019 of Respondents Whose Highest Level of Education is an Associate's Degree

The distribution of income for respondents with their highest level of training being community college or junior college is shown above. The median of this distribution is $42,000. The mean earnings among these respondents is $51,556.84 and standard deviation is $43,612.76. It should be noted that this frequency plot has a similar shape and comparable mean to high school graduate who supplement their education with a training certificate.

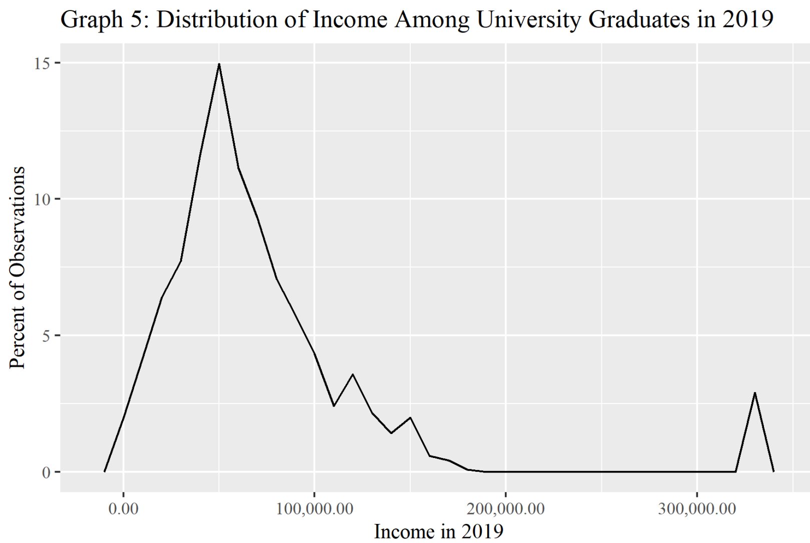

Figure 5: The frequency of Income in 2019 among Respondents Whose Highest Level of Education is a Bachelor's Degree

The final distribution that indicates the returns to education is shown above, specifically of respondents whose highest level of training is a Bachelor's degree. While this distribution has a similar shape, its mean and median are considerably higher than the prior distributions at $71,521 and $58,000 respectively. The standard deviation is around $56,592.65.

Across the above distributions it is evident that the labor market returns increase with further training. However, through this exploratory data analysis, it is clear that the average labor market returns between community college and training programs are not significantly different from each other. This discrepancy will be further explored when applying the Mincer equation to model these returns after controlling for demographic information, skill level and experience.

The next step of this exploratory data analysis is to visualize the distribution of skill, as measured by the ASVAB test. It is important to consider that the scores are calculated relative to each other.

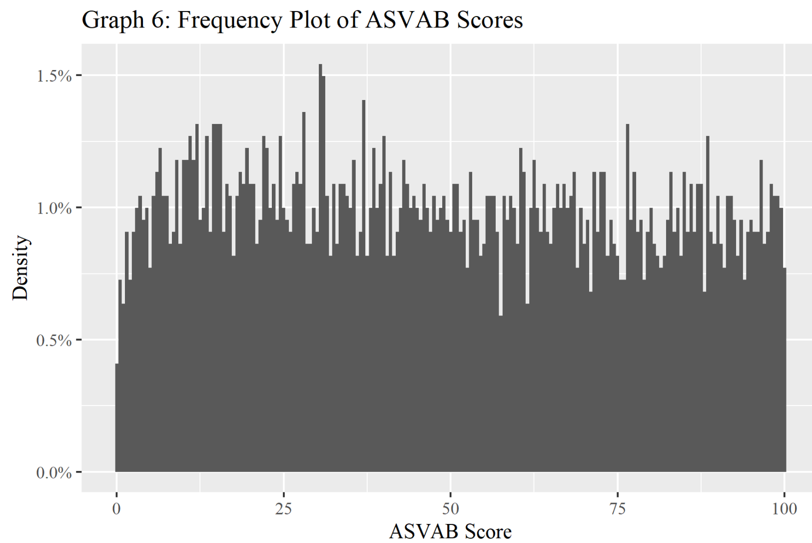

Figure 6: The Distribution of ASVAB score in the Data Set

The above graph shows the density of ASVAB score of respondents in 1999. Since the percentile was measured relative to other respondents, as opposed to the population as a whole, it is clear that the results roughly follow a uniform distribution.

The final step of this exploratory data analysis was to display the representation of different races and ethncities in the dataset.

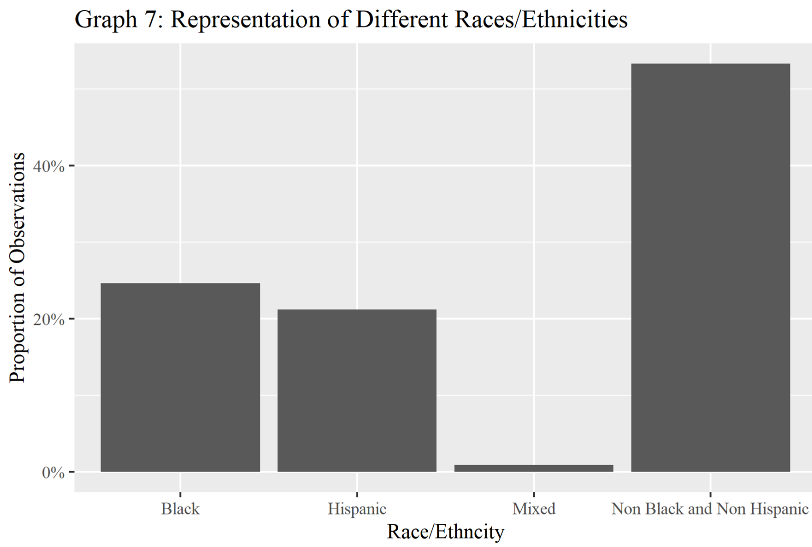

Figure 7: The Proportion of Races and Ethncities in the Data

The above graph displays the proportion of different races and ethnicities represented in the dataset. It is worth noting that the NLSY over represents both Black and Hispanic Americans, as they comprise 24.6% and 21.1% of the data respectively compared to their national representation of 12.1% and 18.7%.

5 Results

The below table displays the findings for the five regression models, specifically, one with only education variables, one with demographics included, and the Mincer equation where time is measured in three different ways, specifically by Age, time at one's highest level of training and labor market experience. All of these models are statistically significant predictors as measured by the F statistic. While only using education predicts roughly 14.9% of the variation, as measured by R^2, however including controls for demographics and experience explains roughly 30.4% of the variation.

There are certain trends across all variables, specifically that women earn less than men in the dataset and that ethnic minorities earn less than non Black and non Hispanic observations. Depending on the controls, women earn between 45% and 36.5% less than men. Similarly Black people in this dataset earn between 10.9% and 4.8% less than the non Black and non Hispanic observations. However, it is worth noting that the coefficients for Hispanic and mixed race are not statistically significant.

The correlation coefficient of the ASVAB is consistently statistically significant across models, as each percentile increase is correlated with between a 0.004% and 0.002% increase in earnings. Strangely, not taking the ASVAB in 1999 was also a statistically significant positive predictor of earnings. However, unless there is a discernible pattern of those who did not take this exam, this correlation is likely spurious.

	Dependent variable:				
	Log of Income in 2019				
	Only Education	Demographic Variables	Age	Time At Highest Training	Labor Market Experience
	(1)	(2)	(3)	(4)	(5)
Drop out	−0.228***	−0.187***	−0.189***	−0.231***	−0.007
	(0.065)	(0.062)	(0.062)	(0.067)	(0.058)
Training Certificate Vocational License	0.295***	0.296***	0.292***	0.417***	0.187***
	(0.045)	(0.043)	(0.043)	(0.064)	(0.039)
Associate's Degree	0.355***	0.367***	0.367***	0.510***	0.264***
	(0.054)	(0.052)	(0.052)	(0.071)	(0.047)
Bachelor's Degree	0.692***	0.648***	0.647***	0.777***	0.571***
	(0.046)	(0.046)	(0.045)	(0.066)	(0.041)
Master's Degree	0.911***	0.891***	0.887***	1.057***	0.796***
	(0.053)	(0.053)	(0.053)	(0.074)	(0.048)
Doctorate	0.874***	0.813***	0.810***	1.004***	0.721***
	(0.122)	(0.118)	(0.118)	(0.130)	(0.105)
Professional Degree	1.605***	1.505***	1.501***	1.663***	1.469***
	(0.096)	(0.094)	(0.094)	(0.107)	(0.084)
Female		−0.450***	−0.449***	−0.450***	−0.365***
		(0.025)	(0.025)	(0.025)	(0.022)
Black		−0.109***	−0.109***	−0.107***	−0.048*
		(0.032)	(0.032)	(0.032)	(0.029)
Hispanic		0.060*	0.060*	0.061*	0.029
		(0.033)	(0.033)	(0.033)	(0.030)
Mixed		−0.036	−0.033	−0.029	−0.060
		(0.125)	(0.124)	(0.124)	(0.110)
ASVAB Score		0.004***	0.004***	0.004***	0.002***
		(0.0001)	(0.0001)	(0.0001)	(0.0001)
Did not take ASVAB		0.175***	0.174***	0.170***	0.135***
		(0.042)	(0.042)	(0.042)	(0.038)
Age (2019)			−0.294		
			(0.438)		
Age Squared (2019)			0.004		
			(0.006)		
Time at Highest Training (2019)				0.011***	
				(0.004)	
Time at Highest Training Squared (2019)				−0.0001	
				(0.0001)	
Labor Market Experience (2019)					0.071***
					(0.005)
Labor Market Experience Squared (2019)					−0.001***
					(0.0002)
Constant	10.217***	10.277***	15.226*	10.071***	9.586***
	(0.039)	(0.049)	(8.066)	(0.074)	(0.061)
Observations	4,559	4,559	4,559	4,559	4,481
R^2	0.149	0.227	0.228	0.230	0.304
Adjusted R^2	0.148	0.225	0.226	0.227	0.301
Residual Std. Error	0.858 (df = 4551)	0.818 (df = 4545)	0.818 (df = 4543)	0.817 (df = 4543)	0.725 (df = 4465)
F Statistic	113.799***	102.518***	89.640***	90.301***	129.868***
	(df = 7; 4551)	(df = 13; 4545)	(df = 15; 4543)	(df = 15; 4543)	(df = 15; 4543)

Note: *p<0.1; **p<0.05; ***p<0.01

Furthermore, in this Mincer equation, it is shown that when a high school graduate or GED recipient pursues a training certificate or vocational license, they should earn between 41.7% more and 18.7% more than if they had not pursued further training. Additionally those with an Associate's degree earn between 51% and 26.4% more than their peers who stop education or training after high school.

It is important to note that these estimates vary wildly. It is worth considering that when comparing across models, the one that controls for labor market experience is consistently more conservative in its correlation coefficient than the other models, especially the one that controls for experience by looking at the time each individual has their highest degree. Since it is also the most statistically significant model, as measured by its F-Statistic and R^2, it is likely the most accurate reflection. By using this model as the standard, it can be shown that a high school graduate should expect to earn roughly 18.7% more by pursuing additional training by means of a vocational license or training certification or 26.4% more by pursuing an associate's degree, after controlling for time in the labor market, skill and demographics.

6 Limitations

While this paper's findings are consistent with literature, specifically that there is a wage premium associated with specialization outside of a university institution, there remain certain limitations of this project surrounding data, methodology and interpretation. The first issue with the data concerns the methodology of the NLSY97. Since this dataset is survey data, the interviewer is relying on the respondent to account for information, specifically income. This question is cause for concern because respondents might not feel comfortable sharing their income with the interviewer, or might round the income to reflect approximations. The approach for data wrangling around missing values was to ignore them under the assumption that regardless of income, respondents were equally likely to be uncomfortable sharing their income.

A methodological limitation of this paper concerns how "not applicable" values were handled. In the survey for 2019, there were roughly 1200 valid skips that were displayed as N/A but there were also 2037 "non-interviews." All of these observations had "not applicable" listed as their income in 2019. In order to generate values for the natural log of income in 2019, those who did not work (no income) and those who did not answer the income question in 2019 were removed. Alternative approach could have been using mean imputation to approximate their income. While this approach could have helped create a successful proxy for income, it could potentially bias the results of the analysis, as each observations education level, sex, and race/ethnicity would likely be used to approximate their income. Thus the regression equations would be overfit to the data, and the coefficients would be less interpretable.

Furthermore, a limitation of this paper is that it uses the ASVAB score taken in 1999 as an indicator for a person's skill. While the ASVAB tests students on a variety of topics, such as math and comprehension, it also tests on mechanical skills and assembling skills. Since the dataset only includes respondents' scores in 1999, it ignores the potential that these respondents might improve in their ability relative to each other. However, since the percentile grade is relative to other respondents, it does provide a metric of skill relative to the other respondents in the NLSY in a close to initial time period.

Another limitation concerns the fifth model, specifically labor market experience. This is because, after 2010, interviews were only conducted every other year. Thus, if a respondent exited the labor market for the duration of a year and returned prior to their next interview, this would be lost in the survey. While the surveyors did ask for month by month breakdowns, wrangling this data led to computational issues that limited this research to the proxy for labor market experience. This means that the labor market experience, might not necessarily reflect years in the labor market. The approach to handle this issue taken in this paper, was to subtract the respondents age in 2019 from their youngest age in which they earned at least $10,000 per year. While for majority of the respondents this reflected their labor market experience, it assumes that people do not leave the labor market once they enter, which is not always the case. Additionally, since a "not applicable" value could also signify that a respondent did not answer the question, as opposed to had no income, it is difficult to be certain whether a lack of value means they exited the labor market or not. While the approach taken for creating and coding this variable had some limitations, it is a reasonable approximation of a respondents' labor market experience in the presence of other omitted variables.

An additional limitation of this data is that there was no provided definition of a vocational license or training certificate. The respondents were only asked: "R has attended a training program and received certificate/license/degree." Thus how the respondent chose to interpret that question can influence their response. For example one respondent could see Microsoft office certification as indicative of vocational certification while another does not. This limitation is a product of the survey and can be addressed through the use of alternate data sets.

A major assumption of this analysis is that one's highest degree or training is most relevant to their labor market performance. While this might be true for those with education beyond a bachelor's degree, it might not be true for those who attended high school and then pursued a vocational license or training certificate. This is especially true since information on what the license or certificate is in is not provided, nor is the respondents job. This limitation can be addressed in future research by adding additional follow-up questions to identify the extent of the certificate and its relevance to each respondents current labor market status.

An additional limitation of these results surrounds the racial or ethnic information collected. It is difficult to interpret the results of the correlation coefficients related to race due to a variety of different ethnicities, such as White, East Asian, South Asian, Native American, etc, all being lumped into a single category. Therefore it is difficult to understand how ethnicity intersects with labor market returns for community college graduates or high school graduates with a training certificate. This can be addressed by either changing the question in future versions of the National Longitudinal Survey of Youth or through using alternate datasets.

7 Discussions

While this paper contributed to a further understanding of the labor market returns to pursuing additional training as well as community college, it was unable to complete its initial intention of understanding the labor market returns to trade school compared to a high school degree. This question can be further explored through further data collection. The Current Population Survey, American Community Survey, the Panel Study of Income Dynamics and the National Longitudinal Survey of Youth all lack data on whether or not a respondent attended trade school. While individual states' Department of Education has access to this administrative data, as seen in Dr. Mountjoy's analysis of the marginal community college student, it is difficult to identify national level data on this attainment (Mountjoy, 2022). Thus there is potential for future work in developing a dataset that measures trade school enrollment and thus allow for a more accurate analysis of the returns of this form of occupational education.

Additionally, these findings provide a basis for future research on how policies related to cost intersect with enrollment in training programs and community college. Certain states have moved to either dramatically reduce tuition or eliminate these costs all together. Thus, this analysis can be built upon by further analyzing how the different returns to these forms of training intersect with different state specific policy.

Similarly, as Dr. Rouse discussed, geography plays a role in enrollment in community colleges, specifically how far someone lives from such an institution (Rouse, 1995). Due to data limitations, this facet was not considered in the analysis. However, this presents avenues for future research, specifically about how proximity intersects with enrollment in training programs or community college. Furthermore, there is potential that the labor market returns are different as proximity might cause a higher proportion of marginal students to sort into these training programs.

While this paper looked at how additional vocational training can benefit those with only a high school diploma, it did not dive as much into how this training can be beneficial to people of different levels of educational attainment. Thus more work concerning the economic benefits of this vocational training for workers with different levels of education can lead to a more rigorous understanding of the benefits toward labor market returns of pursuing additional training.

However, this paper is able to identify that once controlling for labor market experience, there are significant labor market returns associated with high school graduates pursuing further training, either at a community college or through pursuing more specialized vocational licenses and training certificates. This means that high school graduates can be able to dramatically improve their labor market returns through means beyond the traditional path of pursuing a four year university degree.

8 References

Alonzo, F. (2022, February 24). Census Bureau releases New Educational Attainment Data. Census.gov. Retrieved from https://www.census.gov/newsroom/press-releases/2022/educational-attainment.html

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. Journal of Political Economy, 70(5, Part 2), 9–49. https://doi.org/10.1086/258724

Chetty, R., Friedman, J., Saez, E., Turner, N., & Yagan, D. (2017). Mobility Report Cards: The role of colleges in intergenerational mobility. https://doi.org/10.3386/w23618

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the United States. https://doi.org/10.3386/w19843

Dale, S., & Krueger, A. (2011). Estimating the return to college selectivity over the career using administrative earnings data. https://doi.org/10.3386/w17159

DiPrete, T. A. (2020). Annual Review of Sociology, 46, 379–398. https://doi.org/10.1146/annurev-soc-121919-054814

Hanushek, E. A., Schwerdt, G., Woessmann, L., & Zhang, L. (2016). General Education, vocational education, and labor-market outcomes over the lifecycle. Journal of Human Resources, 52(1), 48–87. https://doi.org/10.3368/jhr.52.1.0415-7074r

Heckman, J. J., & Hotz, V. J. (1989). Choosing among alternative non-experimental methods for estimating the impact of Social Programs: The case of manpower training. Journal of the American Statistical Association, 84(408), 862–874. https://doi.org/10.1080/01621459.1989.10478848

Holland, M. M., & DeLuca, S. (2016). "why wait years to become something?" low-income African American youth and the costly career search in for-profit trade schools. Sociology of Education, 89(4), 261–278. https://doi.org/10.1177/0038040716666607

Hout, M. (2012). Social and economic returns to college education in the United States. Annual Review of Sociology, 38(1), 379–400. https://doi.org/10.1146/annurev.soc.012809.102503

Houthakker, H. S. (1959). Education and Income. The Review of Economics and Statistics, 41(1), 24–28. https://doi.org/10.2307/1925454

Jepsen, C., Troske, K. R., & Coomes, P. A. (2012). The labor-market returns to Community College degrees, diplomas, and certificates. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2161239

Kane, T., & Rouse, C. (1993). Labor market returns to two- and four-year colleges: Is a credit a credit and do degrees matter? https://doi.org/10.3386/w4268

Katz, L. F., Roth, J., Hendra, R., & Schaberg, K. (2022). Why do sectoral employment programs work? Lessons from WorkAdvance. Journal of Labor Economics, 40(S1), S249-S291.

LaLonde, R. J. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. The American Economic Review, 76(4), 604–620. https://doi.org/https://www.jstor.org/stable/1806062

Mincer, J. (1958). Investment in human capital and personal income distribution. Journal of Political Economy, 66(4), 281–302. https://doi.org/10.1086/258055

Mincer, J. (1976). Schooling, experience and earnings. National Bureau of Economic Research.

Mountjoy, J. (2022). Community Colleges and upward mobility. American Economic Review, 112(8), 2580–2630. https://doi.org/10.1257/aer.20181756

Osterman, P. (2022). How American adults obtain work skills: Results of a new national survey. ILR Review, 75(3), 578-607.

Rouse, C. E. (1995). Democratization or diversion? the effect of community colleges on educational attainment. Journal of Business & Economic Statistics, 13(2), 217–224. https://doi.org/10.2307/1392376

Spence, M. (1973). Job market signaling. The Quarterly Journal of Economics, 87(3), 355. https://doi.org/10.2307/1882010

Wooldridge, J. M. (2012). Instrumental Variables Estimation and Two Stage Least Squares. In Introductory Econometrics: A Modern Approach (5th ed., pp. 512–525). essay, Southwestern Cengage Learning.

Xu, D., & Trimble, M. (2016). What about certificates? evidence on the labor market returns to Nondegree Community College Awards in two States. Educational Evaluation and Policy Analysis, 38(2), 272–292. https://doi.org/10.3102/0162373715617827

Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. Journal of Labor Economics, 32(4), 711–754. https://doi.org/10.1086/676661