capsee

CENTER FOR ANALYSIS OF
POSTSECONDARY EDUCATION
AND EMPLOYMENT

What About the Non-Completers? The Labor Market Returns to Progress in Community College

A CAPSEE Working Paper

Matthew Zeidenberg
Abt Associates

Marc Scott
New York University

Clive Belfield
Queens College, City University of New York

February 2015

Abstract

Of the copious research on the labor market returns to college, very little has adequately modeled the pathways of non-completers or compared their outcomes with those of award holders. In this paper, we present a novel method for linking non-completers with completers according to their program of study. This method allows us to calculate the labor market returns to programs of study, accounting both for those who obtain an award and those who do not. We use a large dataset of community college transcripts matched with earnings data. We find that different classification systems—by algorithm, intent, or goal—yield very different enrollment patterns across programs. These classifications make a substantial difference to earnings patterns. Our results show that returns vary not only by program completion, but they also vary by program non-completion. Consequently, combining completers and non-completers yields a new pattern of returns. For some awards, this leads to wider earnings differentials. However, we find that the variance in returns by subject of study is reduced when we combine data on completers and non-completers. Finally, we find that progression in a program *per se* does not lead to higher earnings for students who do not complete (even as it demonstrably does for students who complete their program). If validated, these findings have significant implications for policies on program choice and on student retention policies.

Table of Contents

1. Introduction

While copious attention has been paid to the labor market returns to particular education credentials (Altonji, Blom, & Meghir, 2012; Belfield & Bailey, 2011), most students in two-year colleges and many in four-year institutions do not complete a degree or certificate program. For students who first begin in community college, only one-third earn a credential from any institution within six years (another one-fifth are still enrolled after six years). Moreover, the completion rate has been declining over time (Bound, Lovenheim, & Turner, 2010; Radford, Berkner, Wheeless, & Shepherd, 2010). Non-completers thus compose the majority of community college enrollees, yet little attention has been paid explicitly to non-completers and their labor market outcomes. Moreover, many community colleges and policymakers are exploring ways to improve completion rates and help students choose appropriate college programs; to gain a clear picture of the impacts of programs, however, it is necessary to examine both completers and non-completers. It is important to know how labor market outcomes vary according to the students' status at point of exit from college, which varies according to the number of credits they have accumulated, the courses taken, and their progression toward completing an award. Equally important, the outcomes of non-completers are essential when considering the returns to awards. Typically, estimates are reported only for the returns of those who actually complete a particular award and not for those who had intended to complete it but do not do so. As such, these are *ex post* estimates of returns. But the optimal information for decision-making is the *ex ante* return to a college program, or the return an enrolling student can expect. Only by combining evidence for non-completers and completers of a particular program is it possible to estimate the *ex ante* return. As well, non-completers vary in how far they have progressed in a program, or even whether they have enrolled in a program at all. For students who do not complete, it is unclear whether those who almost completed a specific program have higher returns than those who only accumulated college credits that are not directed at a particular field of study.

Identifying which programs non-completers are actually in is a challenge. Students fail to complete a college award for many reasons, and many non-completers make very little progress toward completion. Broadly, there are two methods for categorization: stated preference (i.e., what the student proposes as their program of study) and revealed preference (i.e., the courses the student is actually enrolled in). Existing studies have largely identified students according to their declared program, with little research using revealed preference to determine student programs (Compton, Laanan, & Starobin, 2010). Furthermore, most of these studies examined completion rates and not labor market outcomes over the longer term.

Community colleges have low average completion rates, and non-completers are more numerous and varied than at other institutions. In this paper, we look at labor market returns to completion and non-completion at community colleges using a two-stage approach. First, we utilize a novel method to determine a non-completer's program of study by examining her actual coursework. This method looks at the transcripts of all students who obtained an award for each

program of study. It then classifies each non-completer with an award and subject corresponding to that of the set of completing students that each non-completer's transcript most closely resembles. In effect, the student is "revealing" her program based on the courses she took in relation to courses needed to complete. Once each non-completer's program of study is identified, we apply a second method to determine how far along each student has progressed in her program of study. We then compare this revealed preference approach with approaches based on stated preferences. In the second part of our investigation, we merge these data with data on labor market outcomes. This allows us to identify the labor market returns both to non-completion pathways and to progress along each pathway. Also, by combining the sample of non-completers and completers, we are able to estimate the *ex ante* returns to awards and programs of study.

Our analysis is structured as follows. First, we describe the alternative approaches to categorizing non-completion and the challenges of categorization. We then discuss our algorithmic approach. Next, we describe the data for analysis: a large-scale dataset of community college students with linked transcript and wage data over multiple cohorts and spanning multiple years in the labor market. We report information on non-completers using the algorithm and then estimate labor market returns both for non-completers (by status and by progress) and in combination with completers to estimate *ex ante* results.

Briefly, our results are as follows. We find different enrollment patterns across the measures to capture student progress, and these in turn affect completion rates. We find differences between *ex post* and *ex ante* returns and find that these vary across awards and subjects: awards and fields that are relatively lucrative remain so after adjustment for non-completion, but the differential shrinks. Finally, we find no evidence that progression matters: controlling for the number of credits, students who are further along in an award program have labor market outcomes that are no better than students who have just accumulated general college credits.

2. Understanding College Non-Completion

Categorizing Students Who Fail to Complete Community College

Conventionally, students are categorized by award received, with the residual put into a single group (called *some college* or *college dropouts*). Yet many—almost all—of these students intended or expected to complete their program (by *program* we refer to two elements: the award the student is aiming for and the field of study or major; we focus first on awards and then on subjects). Programs vary in their completion rates, but they also vary in their progression rates: some students will drop out in the introductory courses; others will fail the upper-level courses. Thus, it is necessary to identify which program a student is in, even if (especially if) he does not complete that program; this can be done using either stated or revealed preference methods.

Stated preference methods identify non-completing students according to their declared major or their proposed program (see, e.g., Bailey, Alfonso, Scott, & Leinbach, 2004; Choy & Horn, 1992; Compton et al., 2009; Jacobson & Mokher, 2009; Stuart, 2009). However, not all students declare a major, and declarations are often made late in a student's college career. For some non-completers, a declared major might not correspond to their intended pathway but instead serve as a placeholder (e.g., to satisfy college requirements). Even for completers, we have found that the last declared major does not always match their actual completed major. Other related approaches rely on students' declarations of intent or goal. However, these declarations may be inaccurate or missing: questions about intentions or goals are often presented in closed form with mixed options (e.g., the student might have to choose between "associate degree" or "transfer" responses); and sometimes the declarations have a default category for assignment. As Bailey, Jenkins, and Leinbach (2006) noted, it is often difficult to determine the intentions of a student. In fact, they may be unclear even to the student herself, especially when starting out. Generally, economists are skeptical about stated preferences, particularly in the context of an experience that has so many uninformed consumers (for information deficiencies of students, see Zafar, 2011. Economic models focus on what people do, not what they say they are going to do.

Conventionally, revealed preference methods—looking at what students do—are preferable. The most common approach is to describe dropouts according to the number of credits they have obtained (Melguizo, Kienzl, & Alfonso, 2011). However, this approach assumes that all credits are created equal in how they relate to progress toward completion: no account is taken for repeated courses, courses that cannot be applied to graduation, or general college requirements. Other studies have created taxonomies (e.g., Bahr, 2010). But these taxonomies are often restricted to a few categories and may not correspond to the labels that colleges themselves use to categorize students. Moreover, few studies have looked at labor market outcomes with a focus on categorization of student pathways. In a recent and detailed investigation of the labor market using the National Longitudinal Survey of Youth (NLSY79), Agan (2014) separated out decision nodes based on college sector choices to yield eight different pathways. However, this method did not allow for non-completers to be linked to particular groups of completers such that *ex ante* returns could be calculated for programs. In most of these studies, dropouts are identified *ex post* and, as a combined group, are separated from all completers.[1]

Our approach—emphasizing the revealed preference method—looks at student course-taking patterns and assigns non-completing students to a particular program based on projected awards and subjects of study. In other words, students are matched according to the amount of

---

[1] We are not focused on demographic or college characteristics that are associated with non-completion. This literature is extensive (DesJardins, Ahlburg, & McCall, 2002; Doyle, 2009). This literature could be used to link completers to non-completers to estimate, for example, the *ex ante* returns to particular demographic groups. That said, we are not aware of research that has extensively performed this type of analysis.

postsecondary human capital they have obtained. Using a classification algorithm, we can combine samples of non-completers with completers to calculate labor market returns *ex ante*; we can also look at how far non-completing students progress relative to completers.

The Classification Algorithm

In order to perform the classification, we use an algorithm commonly used in machine learning, the Naïve Bayes algorithm (Mitchell, 1997). This algorithm classifies entities with features into categories. In our case, the entities are the students as represented by their transcripts, the features are the individual courses that students take, and the categories are the awards that students earn. The job of the classifier is to take the transcripts of students who have not completed an award and predict what award they would have earned if they had continued on the pathway that the transcript implies. For the classifier, the input is a list of all features (courses) and whether or not each course is present or not; that is, a long vector of zeros and ones (a bit vector). Since there are many more unique courses in the community college system than any one student will have taken, the bit vector for any given student will be mostly zeros, with ones to indicate the particular courses the student has taken.

The algorithm makes its predictions by compiling the probabilistic relationships between individual courses taken and the award earned, for those students who have earned an award. What makes the algorithm "naïve" is that it is based on the assumption that the features (here, courses taken) are independent of one another; that is, the probability of taking a given course is independent of taking another one, or the probability of seeing them together is simply the product of their individual probabilities. In reality, this is far from the case; for instance, taking Nursing 101 greatly raises a student's probability of following up with Nursing 102. However, machine learning researchers have found that the performance of the algorithm is quite good in many contexts, in spite of this naïve assumption, and therefore it is widely used (Rish, 2001). There are other learning algorithms that have been shown to have better performance (Caruana & Niculescu-Mizil, 2006), but the Naïve Bayes algorithm has the advantage of being relatively easy to understand and can be used as a "proof of concept" for the usefulness of classification of student transcripts.

Formally, the algorithm can be expressed as follows:

$$Y \leftarrow \arg\max P(Y = y_k) \prod_i P(X_i \mid Y = y_k)$$

Here, the X_i represents indicators for each feature (course) and y_k represents the possible categories (awards). For each award, the algorithm computes the product of the probabilities of each course having the realized values, given that award in each case, and multiplies it by the overall probability of the award. Award Y (corresponding to a particular value of k, where k is the subscript over which the awards range) with the maximum value of this product is then selected by the algorithm.

4

To "train" the classifier, we use ten-fold cross validation (Kohavi, 1995). This means that we divide the training data (here, the completer's transcripts represented as bit vectors) into ten equal-sized sets, randomly. We train the algorithm ten times, each time with a different tenth of the data left out of the training. We then test the performance of the algorithm on each corresponding remaining tenth of the data. The prediction for each completer is the prediction for the run of the algorithm in which it was left out. This cross-validation guards against over-fitting to specific realizations of the data. From these data, the Naïve Bayes algorithm is able to create a prediction of what award appears most likely given every student's transcript data, completer and non-completer alike.

We examine the validity of this revealed preference approach in several ways. For the completers, one test is the rate at which the prediction matches the actual award. (It is not possible to assess its performance on the non-completers, since we do not know what their "true" award would be, had they completed.)

To check the error rate for the model, we compare the final declared major and the actual award for completers. For the non-completers, we compare the intended or final declared major (where available) and the predicted award. Of course, a gap between intent and our empirically deduced major does not indicate that our method is incorrect. Also, for many students lacking a declared major, the algorithm is the only way we can discover any major information at all (other than comparing the student's transcript to the requirements for every degree at their college).

The closeness of the match is also important. This method can make either mild or severe errors of classification. For instance, if the method erroneously classifies a student into the Associate of Science (AS) degree program in Business Administration instead of the Associate of Applied Science (AAS) degree program in Business Administration, that is a mild error, because these two programs are very similar. On the other hand, if the student is classified instead into the AAS in Nursing or the AAS in Mechanical Engineering, those are much more severe errors. For the completers, we can get a sense of the relative number of mild and severe errors. Because there are so many possible classifications, it is difficult to do this for all of the misclassifications; we therefore do this for some of the most common awards.[2]

There are several factors that might affect the predictive accuracy of the algorithm. First, many students do not progress very far in their studies. The fewer courses with which the algorithm has to work, the less precise it will be in assigning the student to a specific pathway. (Only if a non-completer took only one course that no completer ever took does the model fail to predict a pathway.) Of course, the fact that these students make little progress may suggest they

---

[2] The accuracy of the algorithm depends on the number of bins into which awards and majors are categorized; the more bins, the more possibilities for errors, all things being equal. Typically, colleges offer awards in the same field at different levels. One approach is to assign a student to one based on the relative proportions of the completer population: if most completers get AAS degrees, then most non-completers will be assigned to this award. One alternative is to collapse all awards into a given field and not be concerned about the specific award. However, there is a big difference in labor market outcomes between an AAS and a certificate.

have no specific pathway or intent other than "being in college." Second, many programs have "back-loaded" course requirements: for instance, most of the courses associated with a business program, such as accounting and economics, are taken in the second year of an AS program. Only for students who progress to take some of these required courses will the algorithm be accurate. Third, students may initially declare "generic" or "default" majors such as liberal arts; they may plan to change their major when their interests get more specific but instead drop out. When colleges have many such declared majors, the model will be less precise. Fourth, many students take credits that are not required, either because they are unsure of the program requirements or because they cannot access the necessary courses (Romano, Losinger, & Millard, 2011). This might introduce more noise into the data the algorithm uses. However, the lack of precision of the algorithm may be a consequence of students' lack of specificity regarding their pathways rather than a failure of the algorithm to identify specific pathways.

As well, we estimate students' progression along particular programs. Within their identified program, we define progression as the ratio of credits earned in the program's course by a non-completer to the median credits earned by all completers of that program. (Because completers do not all take the same courses—due largely to options and electives in each program—we use the heuristic of counting the credits earned only among the courses that account for the top 90 percent of enrollments in the selected program.) Straightforwardly, the progression ratio tells us how far along the program the non-completer got as a percentage of what is required for completion.

We note two aspects of our method. First, our predictions do not necessarily reveal "preferences" in the sense that students choose their programs. The predictions are based on what students actually do (rather than what they want to do). This does not imply intentionality: students may simply take courses without much consideration as to what program they are in. But intentionality is not essential to revealed preference, which strictly speaking is "what happens" rather than what students want to happen (or think is happening). Second, our predictions can be discrepant from what the student is actually awarded. That is, the model may assign a student into a diploma program even if their award is actually a certificate. In this sense, the model is classifying awards in a different way to how colleges classify awards. It is classifying them based on the human capital they have accumulated through their coursework rather than on the college requirements for graduation.

Estimating Labor Market Returns for Non-Completers

The evidence on the labor market returns to community college credentials is now voluminous. In their summary, Belfield and Bailey (2011) reported strongly positive returns across most awards: based on 18 studies, the average earnings premium for an associate degree compared with a high school diploma was 13 percent for men and 21 percent for women. Recent studies have estimated the returns to awards offered by community colleges using large-scale datasets (Andrews, Li, & Lovenheim, 2012; Dadgar & Trimble, 2014; Jacobson, LaLonde, &

Sullivan, 2005; Jepsen, Troske, & Coomes, 2014; Liu, Belfield, & Trimble, 2014). These newer studies have made comparisons within samples of postsecondary students (not between postsecondary students and high school graduates who never attended college). But the overall findings are similar: returns to associate degrees are high, particularly in the health sciences, and gains for female students exceed those of male students. However, these studies revealed more heterogeneous returns to certificates and diplomas, as well as to variations across particular fields of study.

The studies that looked explicitly at community college non-completers used an approach based on evidence of credits accumulated. A few studies have found earnings gains from credits or years of study that do not lead to a completed degree; gains are identifiable for as little as a semester's worth of credits (Jacobson et al., 2005). As noted above, however, this approach does not provide a link between those who complete a program and those who do not. It also assumes that all credits are equal in terms of getting students through their program.

We estimate the returns to pathways using a standard Mincerian earnings equation:

$$\ln Y_t = \alpha + \theta AF_{t\text{-}k} + \beta X_{t\text{-}j} + \gamma Z_{t\text{-}k} + \delta EXP_t + \varphi EXP_t^2 + \varepsilon \tag{1}$$

Earnings Y at quarter t are a function of: a vector of college awards/fields, AF; a vector of pre-college characteristics, Z; a vector of pre-college personal and ability-related characteristics, X; and work experience, EXP. We report quarterly earnings as percentage gains over baseline.

The coefficient θ represents an estimate of the earnings premium from the program followed by the student in community college. For completers, the program is either an award (associate degree, certificate, or diploma) or award-with-subject (based on CIP categories). For non-completers, the program is identified using our classification algorithm. As a comparison, the program is identified using the stated preferences of students about their award intentions and goals. Using this method we separate students into those who intended to get an award and did and those who intended to get an award but did not. (We undertake an analogous separation based on student goals.)

The coefficient θ is an unbiased estimate of the gains from education programs if there is no omitted variable bias or selection/endogeneity bias (Brand & Xie, 2010). Although our earnings function does not fully address these biases, results using this method have proved to be highly robust and minimally biased (Carneiro, Heckman, & Vytlacil, 2011; Jepsen et al., 2014). As our sample includes all those initially enrolled at a community college, we might expect less omitted variable bias than we would expect using high school graduates as the baseline group (e.g., all subjects were motivated enough to enroll in college). Endogeneity bias may be present in our sample. All students faced similar labor market options because they began college at the same time (and in the same state), but clearly some entered the labor market before others, and there may be unobservable reasons for this. Yet, it seems unlikely that the bias on unobservables is as large as the bias on observables (see Webber, 2014), and earlier analyses have found the bias on observables to be very small (Liu et al., 2014). Moreover, it is important to note that we

are making comparisons between program completers and all program entrants, and so the extent of the bias should be smaller.

Our estimation approach is as follows. First, to make equivalent comparisons, we identify as our program default category those students who obtained a diploma; all coefficients are therefore relative to these completers. Next, we estimate the returns to completers and the general class of non-completers, as per the general literature. We then estimate returns where non-completers are separated according to their program. Here the default category is individuals in diploma programs. This yields the returns to not completing a specific program.

We then combine the coefficients on completers and non-completers to estimate the *ex ante* returns. Specifically, we estimate these *ex ante* returns R as the weighted sum of the returns to completers C and non-completers N:

$$R_{ExAnte} = \varphi R_C + (1-\varphi) R_N \tag{2}$$

Here, φ is the probability of completing the award or program. So, the *ex ante* returns to a pathway depend on the returns to completers, the returns to non-completers, and the probability of completing.

We expect that program estimates for the pooled group of non-completers with completers will yield lower rates of return to college. However, non-completers may have had more opportunity to work. The complementarity between education and experience may be such that over the window of analysis here the non-completers actually experienced faster wage gains. Non-completers might also be inter-temporally substituting: taking employment when easily available with the expectation that they can return to college later if their earnings do not increase.

As well, we expect that the adjustment to *ex ante* returns will affect pathways differentially. Relatively lucrative award choices may not be so valuable once we adjust for differences in completion rates. Most likely, the adjustment should compress earnings gaps if high-earning programs are also those with the highest non-completion rates. For example, an associate degree in nursing leads to relatively high earnings; but if this degree program has a relatively high failure rate, then the net earnings advantage is attenuated.

Finally, we look at student progression to examine whether students who progressed further along a given program have higher returns. We estimate a version of equation (1) for non-completers but include an indicator for progression *PROG*:

$$\ln Y_t = \alpha + \beta X_{t-j} + \gamma Z_{t-k} + \delta EXP_t + \varphi EXP_t^2 + \lambda PROG_t + \zeta CR_t + \varepsilon \tag{3}$$

Progression is almost certainly positively associated with earnings because students who have progressed further have more credits. Therefore, we include credits *CR* in our estimation. The coefficient λ is therefore identifying whether student progression toward an award yields higher earnings than simply accumulating college credits.

We expect that students who have progressed further on an award will have higher earnings than those whose progress was weaker, even after controlling for credits. Students who are progressing are most likely to have taken upper-level (harder) courses and mastered more material. They may also be more goal-directed (such that the earnings gains identify unobserved motivation). Greater progression should lead to higher earnings. Alternatively, lower-level courses may produce more human capital (with more general value in the labor market). Students who drop out might benefit from being generalists rather than specialists: taking courses in principles of economics and principles of sociology may be more useful than principles of economics and advanced economics, for example. Also, lower-level and non-program courses are more readily available in the academic calendar; such courses can thus enable students to more easily integrate their studies with their labor market activities.

3. Data

Datasets

Our dataset is composed of all first-time-in-college, credit-seeking students in the North Carolina Community College System (NCCCS) in the academic years 2002–03 through 2004–05.[3] The NCCCS includes 58 colleges and enrolls approximately 100,000 new curriculum (award-seeking) students each year. The transcript dataset contains information on individual students, including full college transcripts (e.g., courses taken, grades earned, awards received, duration of study), basic personal information (e.g., age, gender, race/ethnicity), and financial aid received (loans and grants per semester). The transcript data do not include direct information on socioeconomic status or on prior academic achievement.

We merge the college transcript data with student-level data from the National Student Clearinghouse (NSC). The NSC tracks students as they transfer to other Title IV eligible colleges, as more than one third of all community college students do (Hossler et al., 2012). The NSC dataset includes information on awards subsequent to enrollment within the NCCCS, but field of study is not reliably recorded across all colleges.

Our analysis focuses on the highest award obtained by each student at any institution within the designated time period. NCCCS awards include certificates, with 12–18 semester-hour credit requirements, which are intended to be completed in one year of full-time study or less; diplomas, with credit requirements ranging from 36–48 semester-hour credits, with a general education requirement; and associate degrees (mostly in applied science), with a required 64–76 semester-hour credits, including a cross-disciplinary general education requirement of 15

---

[3] We exclude students who were not enrolled in designated curriculum programs leading to awards but in customized training, personal enrichment courses, or other noncredit programs. For further details on the dataset, see Liu et al. (2014).

semester-hour credits. The NSC data allow for the identification of NCCCS students who eventually obtained a bachelor's degree or higher (these students are excluded from our sample).

Using Social Security numbers, we then merge this combined student dataset with earnings data obtained from the North Carolina Department of Commerce Unemployment Insurance (UI) records. The UI earnings data are collected on a quarterly basis from UI-covered employers and include total earnings from all jobs, as well as Standard Industrial Classification (SIC) and North American Industry Classification System (NAICS) information for each job (hours of work are not available). We use earnings data for the most recent available year (2011). Earnings are the average across quarters with non-zero earnings in 2011. All earnings are adjusted for inflation to be expressed in 2010 dollars using the quarterly Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W).[4]

Descriptive Frequencies

Table 1 shows the sample for analysis and information based on application of the algorithm (see also Appendix Figure A.1). Our sample includes students who received an award at the community college level and then transferred to a four-year college and includes students who transferred but received no award. We exclude students who transferred and then received a bachelor's degree without receiving a credential at the community college level. We assume these students are not intending to complete an award at the community college level and so their progression status is not relevant.

As is the case across the community college sector, fewer than half of the students in our sample received an award (excluding bachelor's degrees). For associate degrees, 19 [13] percent of the female [male] students have received an award. Based on the algorithm, 49 [50] percent of students were in associate degree programs. The completion rate for associate degrees is therefore 39 [26] percent. For certificates, 4 [7] percent of students received an award. Yet 31 [34] percent of students were in certificate programs. The completion rate for certificates is therefore 13 [21] percent, considerably below that for associate degrees. Finally, 3 [2] percent of students completed diplomas, with 20 [16] percent of students in such programs. The graduation rate for the period under study is therefore 15 [13] percent. Interestingly, these graduation rates are quite different from rates derived from IPEDS data (which do not count students who transfer) and from comparative analyses of colleges based on their proportions of awards that are certificates (Bailey & Belfield, 2012).

---

[4] Unlike sample survey data, UI earnings data are not affected by biases due to imputation, self-reporting, and nonresponse. However, the UI data do not include all workers; they exclude independent contractors, military personnel, some federal personnel, and those working in the informal sector (e.g., casual laborers). In most states, including North Carolina, state UI datasets do not include workers who moved out of state. However, in our sample over 90 percent have at least one UI wage record.

Table 1: Descriptive Statistics

	Female	Male
Awards		
Associate degree	.19	.13
Certificate	.04	.07
Diploma	.03	.02
Students predicted per award program track		
Associate degree	.49	.50
Certificate	.31	.34
Diploma	.20	.16
Intentions		
Associate degree (intended and award)	.04	.01
Associate degree (intended but no award)	.13	.05
Certificate (intended and award)	.04	.07
Certificate (intended but no award)	.57	.48
Intended other	.32	.42
Goals		
Associate degree (goal and award)	.09	.05
Associate degree (goal but no award)	.24	.19
Certificate (goal and award)	.01	.01
Certificate (goal but no award)	.04	.03
Goal other	.63	.71
Program match	.28	.31
Credits earned		
Associate degree—award	73.0 (25.3)	68.3 (25.4)
Certificate—award	35.0 (22.8)	26.6 (20.0)
Diploma—award	59.5 (20.4)	53.4 (23.2)
Associate degree—predicted but no award	25.2 (23.1)	22.4 (22.2)
Certificate—predicted but no award	9.9 (12.9)	7.8 (10.9)
Diploma—predicted but no award	18.0 (18.9)	14.3 (17.0)
Progression		
Associate degree—predicted but no award	0.32 (0.29)	0.29 (0.29)
Certificate—predicted but no award	0.30 (0.37)	0.25 (0.34)
Diploma—predicted but no award	0.29 (0.30)	0.25 (0.31)
Ln(Quarterly Earnings)[a]	8.16 (1.20)	8.35 (1.24)
Total observations	256,615	

Note. [a] Non-zero earnings observations only.

Two stated preference approaches are available in our dataset. One uses a statement of each student's intentions; using a closed form question, students could declare their intention as earning an associate degree or occupational award (which we interpret as a certificate), or other intentions. The other is based on each student's goals; also a closed form question, students could declare an associate degree or skills upgrading (which we interpret as a certificate).

As shown in Table 1, these stated preference metrics yield much lower rates of associate degree program enrollment and imprecise rates of certificate program enrollment. They also yield many uninformative "other" codings; as shown in Appendix Figure A.1, a large proportion of students cannot be identified using these metrics. These factors suggest caution in using stated preferences as indicators of what students do. Given the ambiguity between intentions/goals and awards, we focus on the revealed preference measures for associate degree students (and not students in other programs).

Finally, Table 1 shows the credits earned and progression rates for students by award. Credits earned are the typical indicator of student progress. For each award program, non-completers have about one third of the credits of completers. However, this credit comparison may be misleading if the credits are not moving students toward completing a program. A more accurate guide is the progression rate. This is calculated as the ratio of credits earned in a program to the median credits earned by all completers of each program.[5] Looking at progression, non-completers do not get very far along their pathway: on average, non-completers only progressed about 30 [26] percent along the way to completing their intended award.

4. Results

Model Accuracy

We begin by examining the model accuracy. Our first observation based on the computation of the algorithm is that there are very large numbers of awards and majors within the college system: we estimate there are over 600 combinations. In turn this suggests that students are unlikely to know *ex ante* which program to choose and, in their first year, which program they are making the most progress toward.

Second, only 10 percent of students do not declare a major (the declaration rate is actually slightly *lower* for completers). This rate is sufficiently low that administrators may prefer to rely on the major for making program assignments. However, the final declared major of completers only matches their final award 61 percent of the time. So major is not a very good guide to what awards students end up with. It is likely to be even less useful a guide as to the

---

[5] As noted above, credits earned in a program are only those from courses accounting for the top 90 percent of enrollments in the program.

pathways of non-completers. By contrast, the algorithm correctly classifies completers based on their courses 71 percent of the time.

Given these rates and measurement errors, the match between the algorithm and the declared major is low. The declared major matches the predicted major only 15 percent of the time. However, this low match rate is a function of both the bin numbers and the noise factors referred to above, as well as the trained target of the algorithm, which is the actual major. There are 637 different bins, many of which are very close to each other. The main factor reducing precision is that most of the non-completers made very little progress (9 credits at the median).

The model does generate predictions that are, for some students, discrepant from their actual award. However, these discrepancies are few. Only 5 percent of students with associate degrees were not predicted to receive this award; 10 percent of students who received a certificate were not predicted to receive one; and 5 percent of students who received a diploma were not predicted to receive one. In our analysis, we use our predictions of awards rather than the actual awards.

Labor Market Returns for Non-Completers

Table 2 shows coefficients for earnings gain relative to students whose highest award is a diploma from community college (see table note for covariate controls). The sample includes students who dropped out of community college and students who transferred to another institution, unless those transfer students then went on to complete a bachelor's degree at a four-year institution (and so are better matched with four-year students). Model (1) aggregates all non-completers into a single group. The other three models differ in how they classify students *ex ante*: Model (2) uses the prediction from the algorithm, and models (3) and (4) apply versions of students' proposals of intent or goals.

For female students, model (1) shows the gains over diploma-holders for associate degree holders were +5 percent, but for certificate holders they were –33 percent. Grouped together, non-completers earned 32 percent less than diploma-holders. Therefore, we identify very high returns to *completing* an associate degree (+37 percent) or diploma (+32 percent) and insubstantial returns to a certificate (–1 percent). This last result is plausible: non-completers and certificate-holders have similar numbers of credits. These premiums correspond to the *ex post* premiums that are typically reported in rate of return studies.

Model (2) shows earnings gaps accounting for the predicted program of each student (again relative to diploma-holders). Students with associate degrees or certificates obtain approximately the same premium/deficit as in model (1). But the non-completers are disaggregated according to which award program they intended to complete based on the algorithm. All non-completers earn less, but the deficit varies according to program track.

Table 2: Returns to Awards: Actual, Predicted, and Stated Preference
Average Quarterly Earnings Gains in 2011 for 2002–05 NCCCS Cohorts

	Female				Male			
	(1) Actual	(2) Predicted	(3) Intended	(4) Goal	(1) Actual	(2) Predicted	(3) Intended	(4) Goal
Relative to student with diploma								
Associate degree	0.042** [0.018]	0.092*** [0.016]			0.159*** [0.029]	0.157*** [0.025]		
Certificate	–0.401*** [0.024]	–0.339*** [0.023]			–0.033 [0.032]	–0.041 [0.027]		
Did not complete an award	–0.379*** [0.017]				–0.007 [0.028]			
Predicted assoc. degree (no award)		–0.268*** [0.017]				0.022 [0.024]		
Predicted certificate (no award)		–0.439*** [0.017]				–0.043* [0.025]		
Predicted diploma (no award)		–0.345*** [0.017]				–0.063** [0.026]		
Proposed assoc. degree (awarded)			0.439*** [0.016]	0.163*** [0.024]			0.124*** [0.041]	0.158*** [0.033]
Proposed assoc. degree (no award)			–0.000 [0.014]	–0.230*** [0.023]			0.025 [0.032]	0.003 [0.031]
Proposed certificate (awarded)			–0.036 [0.026]	–0.191*** [0.068]			0.014 [0.043]	–0.097* [0.058]
Proposed certificate (no award)			0.103*** [0.018]	–0.222*** [0.026]			0.079** [0.039]	–0.015 [0.034]
Proposed other			0.091*** [0.020]	–0.165*** [0.023]			0.084** [0.040]	0.025 [0.031]
Observations	*96,221*	*96,221*	*96,221*	*96,221*	*62,769*	*62,769*	*62,769*	*62,769*
R-squared	0.130	0.133	0.116	0.119	0.160	0.160	0.158	0.158

Note. Log average quarterly earnings in 2011 as dependent variable. Education coefficients are relative to student with diploma award. Cohorts of FTIC students from fall 2002 to summer 2005. Model includes: race (3 groups), single parent, HS graduate, disability, enrollment age (2 groups), experience, experience squared, financial aid EFC, grants, loans, and aid, college GPA first semester. Robust standard errors in brackets.

*** $p < .01$. ** $p < .05$. * $p < .1$.

Non-completers on the associate degree track earn 24 percent less than diploma holders, those on the certificate track earn 36 percent less, and those who did not complete their diploma earn 29 percent less. In other words, for students who terminated their postsecondary education with no award, those who were predicted to complete certificate programs have the worst labor market outcomes.

An alternative approach to estimating *ex ante* returns is to look at student proposals of intent or goal. As noted above, these proposals may be questioned: some students do not make them, make them late, or are assigned them by default; and they reflect students' stated preferences rather than revealed behaviors. We focus on results for associate degree students. Models (3) and (4) report earnings gaps based on students' proposed intent and goal. For female students, there are very high returns (+55 percent in model [3]). Analogously, model (4) shows there are positive returns to students whose "goal" was an associate degree and who actually earned an associate degree (but not as strong, at +18 percent). Unsurprisingly, students who did not complete do considerably worse. Students who "intended" to get an associate degree and did not get one have earnings that are indistinguishable from diploma-holders (model 3). But students who failed in their "goal" of earning an associate degree earn considerably less (–21 percent). Thus, the penalty for non-completion is very different depending on whether intent or goal is used. There is a similar inconsistency between intent and goals in the returns to certificates. Of course, these two indicators classify students at very different rates: 17 percent of students intended to get an associate degree, whereas 33 percent had this award as their goal. By comparison with our revealed preference approach, the outlier approach is *intent*: this last measure yields a zero penalty for non-completion relative to a diploma; the other two measures yield similar penalties (of 24 percent and 21 percent).

The results for males are quite different from those of female students. However, we caution that these differences may be because the returns to diplomas may be different across the genders. For males, *ex post* returns to associate degrees are very high (+17 percent); returns to certificates and to the general group of non-completers are not statistically significantly different from diploma-holders. As shown in model (2), the returns to predicted non-completers in each award group are similar to those of diploma-holders. For students who did not complete, it does not appear to matter what they failed to complete. More starkly, the only pathway that did yield an earnings premium is completion of an associate degree; for all other pathways (diplomas, certificates, non-completion) there is modest or no statistically significant difference.

For the stated preference results, the premiums are very similar for students who either intended to, or had the goal of, earning an associate degree and were successful (+13 percent and +18 percent respectively). This is despite these two declarations yielding very different proportions of completers (Table 1). As with the revealed preference approach, other pathways were not statistically significantly different.

Labor Market Returns *Ex Ante*

We now turn to calculations of the *ex ante* returns to each award. Using the specifications in Table 2 we can calculate the expected quarterly earnings across the combined group of both award completers and those predicted for that award. Based on the coefficients in Table 2, it is clear that the combined earnings are going to be much lower than those for completers.

To accurately identify the differences between *ex post* and *ex ante* premiums, we estimate the specification as per Table 2 but define the default group as all students who were predicted to be on a diploma track (not just those who completed one). We focus on the relative earnings for associate degree programs and diploma programs. The results are given in Table 3.

The results in model (1) show that, relative to students on a diploma track, students who earned associate degrees have much higher subsequent earnings. For women, the gap is 41 percent and for men it is 23 percent. These premiums are much higher than those in Table 2. This is unsurprising because the default group now includes students who failed to complete a diploma. Nevertheless, even those who completed a certificate have lower earnings than those who enrolled in a diploma program (although the effect is not statistically significant for males).

Our primary comparison is between the actual results for associate degree holders and the evidence for those who are predicted, intended, or had the goal of being an associate degree recipient. These results are given in models (2), (3), and (4), respectively. Compared with those on the diploma track, the earnings gains for those predicted to be in an associate degree program are still positive and statistically significant. They are 15 percent [12 percent] higher for female [male] students. This figure is much lower than for the actual completers (model [1]): *ex ante* returns are always lower than *ex post* returns. But critically, this figure is much higher than the relevant model (1) comparison in Table 2. The *ex ante* return to an associate degree is significantly positive and greater than the *ex post* return. We emphasize that this is all relative to a diploma. Looking only at students who completed associate degrees and diplomas *understates* the premium to the former. Adjusting for the probability of completing these respective awards, the associate degree premium is *higher* (coefficients in models [1] of Table 2 and [2] of Table 3).

An analogous result holds for certificates. For female students, there is a large penalty for completing a certificate relative to completing a diploma (–29 percent, model [2] of Table 2). But this penalty is much lower when we adjust for who is in which program. The penalty for being in a certificate program relative to being in a diploma program is only –14 percent. That is, completing a certificate appears to be the worse type of award to complete, but this disadvantage is lessened when we account for the higher probability of completing one. This results holds for female students; for male students, the results are inconclusive as few pathways appear relatively lucrative.

Table 3: Returns to Awards: *Ex Ante* Actual, Predicted, and Stated Preference
Average Quarterly Earnings Gains in 2011 for 2002–05 NCCCS Cohorts

	Female				Male			
	(1) Actual	(2) Predicted	(3) Intended	(4) Goal	(1) Actual	(2) Predicted	(3) Intended	(4) Goal
Relative to diploma enrollee								
Associate degree (awarded)	0.343*** [0.011]				0.204*** [0.016]			
Certificate (awarded)	−0.097*** [0.020]				0.012 [0.021]			
Non-completer (excl. diploma)	−0.070*** [0.010]				0.049*** [0.013]			
Associate degree (awarded or not)		0.140*** [0.010]	0.098*** [0.013]	0.080*** [0.012]		0.110*** [0.013]	0.069*** [0.023]	0.074*** [0.016]
Certificate (awarded or not)		−0.149*** [0.011]				0.009 [0.015]		
Other group			0.030*** [0.011]	0.038*** [0.012]			0.074*** [0.013]	0.065*** [0.015]
Observations	96,221	96,221	96,221	96,221	62,769	62,769	62,769	62,769
R-squared	0.127	0.120	0.124	0.123	0.160	0.159	0.159	0.159

Note. Log average quarterly earnings in 2011 as dependent variable. Education coefficients are relative to student predicted to be on diploma track. Other group includes all other students except those on diploma track. Cohorts of FTIC students from fall 2002 to summer 2005. Model includes: race (3 groups), single parent, HS graduate, disability, enrollment age (2 groups), experience, experience squared, financial aid EFC, grants, loans, and aid, college GPA first semester. Robust standard errors in brackets.

*** $p < .01$. ** $p < .05$. * $p < .1$.

These findings are partly corroborated when we use the proposed measures of intent and goal. When we estimate an equation where diploma enrollees are compared with those who intended to and did get an associate degree, the returns to the degree are very large and comparable to those in column (1) of Table 3. The same is true for those whose goal was an associate degree and were successful in this goal (these results are not reported). Of more interest is the comparison between those who intended to get a diploma and those who intended to get an associate degree (regardless of their success).

As shown in columns (3) and (4) of Table 3, for women [men] the gains for associate degrees are +11 percent [+8 percent] based on intent and +8 percent [+8 percent] based on goals. For female students, at least, this result accords with those described above. When only looking at completers, the returns to an associate degree over a diploma are +5 percent. After adjusting for probability of completion, the returns increase to 8–11 percent. Looking only at completers therefore understates the returns to an associate degree. However, for male students, the opposite conclusion is evident: gains for associate degree holders are overstated, compared with the gain for associate degree enrollees.

As an alternative exposition, we estimate the predicted earnings for each group. We estimate the models in Table 2 to generate predicted earnings with a binary indicator for each award status (controlling for covariates). These predicted earnings were quarterly earnings in 2011, or seven to nine years after initial enrollment. Overall, these results showed *ex ante* returns to be considerably below *ex post* returns.[6]

Table 4 provides an overall summary of the results, comparing *ex ante* and *ex post* returns to diplomas with those of other awards. For female students, those with associate degrees have higher returns than those with diplomas (R_C) and those who intended to complete an associate degree also have higher returns (R_N). The associate degree track also has a higher completion probability. The net result is that the returns to associate degrees are even greater after adjusting for non-completion. For certificates, the converse is true: relative to diplomas, the returns are lower for non-completers (as well as completers) and the completion rate is lower. Hence, the returns are even lower after adjusting for non-completion. For male students, the results for associate degrees are equivalent to those for female students: returns are even greater when we adjust for non-completion. However, the results for certificates are distinctive. Men who completed a certificate earn less than men who completed a diploma. But men who did not complete a certificate earn more than those who did not complete a diploma. As well, the completion rate for certificates is higher than for diplomas. Given the sizes of the coefficients and probability, we calculate as per equation (2) above that the lower *ex post* returns are fully offset by the relatively higher returns to non-completion and the higher probability of completion. Overall, for men, we therefore find no difference in returns for those beginning a certificate versus those beginning a diploma.

---

[6] Details are not reported but are available from the authors.

Table 4: Returns to Awards: Ex Post and Ex Ante

	Returns relative to diplomas		Completion probability	Interpretation
	R_C	R_N	β	
Female:				
Associate degrees	>	>	>	Returns even <u>greater</u> after adjusting for non-completion
Certificates	<	<	<	Returns even <u>lower</u> after adjusting for non-completion
Male:				
Associate degrees	>	>	>	Returns even <u>greater</u> after adjusting for non-completion
Certificates	<	>	>	Lower *ex post* returns fully offset by higher completion and returns to non-completion

Note. Based on coefficients from Table 3.

Labor Market Returns by Subject

We now estimate returns by subject, separately identifying those for completers and the group who intended to complete. The subjects are collapsed from the CIP codes and follow the categories used by Dadgar and Trimble (2014); the default subject is "missing/other." First, for each sample grouping we estimate the returns to subjects for students who have any award (degree, certificate, or diploma) in that subject. All students who did not complete an award are grouped together. Therefore, all "award" coefficients are relative to students who got an award in a missing subject. In the second specification, all students in each subject (labeled *field*) are grouped together regardless of whether they received an award or not; that is, all non-completers are re-assigned to their field of study. Again, the default subject is missing/other (including those who completed and those who did not complete). By comparing the two specifications, we are able to see the effect of including all students in a subject grouping versus only those students who passed their subject-specific award. We combine all students across award types; we do this partly to make the analysis manageable and partly because of sample size issues. However, to check the results we also restrict the sample to just students in the associate degree programs.

Table 5 shows the returns across subjects using the two specifications (*award* and *field*). For female students who received awards, there are few earnings differences by subject (column [1] of Table 5). Most students who earned an award get similar earnings, at least similar to the returns for students for whom the subject award is unknown. The clear exception is nursing, where earnings gains for nursing credentials are enormous (+104 percent); with an alternative

Table 5: Returns to Subjects: Actual and Predicted, Average Quarterly Earnings Gains in 2011 for 2002–05 NCCCS Cohorts

	Female				Male			
	All Students		Associate Degree Students (Predicted)		All Students		Associate Degree Students (Predicted)	
	Award	Field	Award	Field	Award	Field	Award	Field
Nursing	0.711***	0.315***	0.992***	0.708***	0.968***	0.327***	0.647***	0.252***
	[0.131]	[0.050]	[0.179]	[0.092]	[0.108]	[0.059]	[0.143]	[0.084]
Health	0.189	0.114**	0.381**	0.266***	0.647***	0.266***	0.256*	0.083
	[0.131]	[0.050]	[0.179]	[0.092]	[0.106]	[0.061]	[0.143]	[0.082]
Humanities / Social Sci.	–0.030	0.046	0.060	0.081	0.306***	0.147***	–0.127	–0.118
	[0.131]	[0.050]	[0.179]	[0.092]	[0.103]	[0.055]	[0.139]	[0.076]
Math / Science	–0.520	–0.106	–0.479	–0.216	0.201	0.097	–0.304	–0.227
	[0.351]	[0.074]	[0.385]	[0.171]	[0.288]	[0.097]	[0.303]	[0.160]
Inf. Science	–0.054	0.008	0.060	0.049	0.416***	0.196***	–0.006	–0.088
	[0.136]	[0.053]	[0.183]	[0.095]	[0.105]	[0.057]	[0.141]	[0.077]
Engineering Mech.	0.039	0.111**	0.173	0.200**	0.548***	0.306***	0.174	0.084
	[0.140]	[0.056]	[0.186]	[0.096]	[0.103]	[0.056]	[0.140]	[0.077]
Mechanics	–0.257	–0.169**	–0.200	–0.090	0.388***	0.171***	0.100	–0.001
	[0.170]	[0.075]	[0.277]	[0.147]	[0.103]	[0.056]	[0.142]	[0.078]
Protective Services	0.139	0.013	0.185	0.038	0.703***	0.407***	0.274*	0.141*
	[0.134]	[0.053]	[0.182]	[0.095]	[0.102]	[0.056]	[0.140]	[0.078]
Construction	–0.409	–0.139	–0.474	0.045	0.249**	0.087	0.260*	0.052
	[0.276]	[0.093]	[0.789]	[0.270]	[0.107]	[0.058]	[0.147]	[0.085]
Business / Marketing	–0.030	0.020	0.111	0.133	0.334***	0.196***	–0.059	–0.045
	[0.132]	[0.051]	[0.180]	[0.093]	[0.107]	[0.057]	[0.144]	[0.079]
Education / Child Care	–0.194	–0.215***	–0.025	–0.062	0.284*	0.078	–0.271	–0.241**
	[0.132]	[0.050]	[0.180]	[0.093]	[0.146]	[0.059]	[0.205]	[0.100]
Transportation	–0.284	–0.257**	0.600**	0.403*	0.166	–0.059	–0.021	0.030
	[0.181]	[0.102]	[0.304]	[0.225]	[0.110]	[0.063]	[0.317]	[0.138]
Cosmetology	–0.219*	–0.208***	0.063	–0.005	0.265**	0.009	–0.062	–0.180**
	[0.133]	[0.052]	[0.183]	[0.095]	[0.124]	[0.064]	[0.162]	[0.087]
Did not complete an award	–0.218*		–0.042		0.333***		–0.071	
	[0.130]		[0.179]		[0.101]		[0.138]	
Observations	*96,221*	*96,221*	*48,885*	*48,885*	*62,769*	*62,769*	*33,444*	*33,444*
R-squared	0.144	0.128	0.162	0.141	0.164	0.164	0.153	0.151

Note. Log average quarterly earnings in 2011 as dependent variable. Cohorts of FTIC students from fall 2002 to summer 2005. Model includes: race (3 groups), single parent, HS graduate, disability, enrollment age (2 groups), experience, experience squared, financial aid EFC, grants, loans, and aid, college GPA first semester. Robust standard errors in brackets.

*** $p < .01$. ** $p < .05$. * $p < .1$.

comparison group, awards in health would probably also yield statistically significant gains. Unsurprisingly, as shown in the final row of column (1) of Table 5, non-completers have lower earnings than completers whose subject is unknown (by –20 percent).

Column (2) of Table 5 shows the pattern of returns for students who have been assigned to a field regardless of whether they completed an award or not. With this specification, the gains for nursing and health programs are substantially reduced (+37 percent and +12 percent respectively). This diminution suggests that, although nursing programs are highly lucrative, they are less so when one accounts for the lower probability of completion. Moreover, new gaps emerge for the lowest earning subjects: programs in mechanics, education/child care, transportation, and cosmetology are now relatively less lucrative than they have been in the past. Thus, given their relatively high failure rates, these subjects now appear to yield statistically significant penalties (compared with the unknown category).

As shown in columns (3) and (4) of Table 5, the results are similar when the sample is restricted just to associate degree programs. Again, nursing and health programs are the most lucrative, but their premiums—albeit remaining positive—are attenuated when we adjust for the probability of completing an associate degree in these fields. Looking across the sample of students who are following a path toward an associate degree, we find that adjusting for the likelihood of completion reduces subject-specific earnings gaps. For example, relative to those with associate degrees in unknown/missing subjects, returns to associate degrees in nursing/health are 170 percent/46 percent higher, and the returns to all non-completers are 6 percent lower. When these non-completers are assigned to their predicted track, the returns to associate degrees in nursing/health over those in unknown subjects fall to 103 percent/30 percent. Although the gap remains very large, the *ex ante* returns to nursing and health programs are one-third lower than the *ex post* returns for female students.

For male students, the patterns in Table 5 are different because subject-specific enrollment patterns are very different by gender within community colleges. For the full sample of students there are high returns to engineering mechanics and protective services, as well as to the health subjects (although other subjects appear to be high-earning, they are not when compared with the group of non-completers, who also have relatively high earnings compared with those who completed an award in the missing field category). As for the female sample, the returns to all students within a given field are more closely clustered than are the returns only to completers within that field. Consistently, therefore, the greater earnings in some fields partly reflect the relative difficulty in program completion. For the sample of students in associate degree programs, subject-specific variations in earnings are much lower. Few subjects show relatively high returns. Again, earnings gaps by subject award are much greater than by subject field. The greatest effect is for nursing: the very large relative earnings premiums for these

awards are reduced by a factor of three when we include all the students who were predicted to be in a nursing program but who did not complete their award.[7]

Labor Market Returns for Progressing in a Program

Next we estimate the effects of students' progression on their future earnings to see whether students who have gotten further in their program do better in the labor market than those students who accumulated general credits. As completers have progressed fully, the sample is restricted to non-completing students. The estimates on progression are reported in Table 6.

Table 6: Returns to Credits and Program Progression, Average Quarterly Earnings Gains in 2011 for 2002–05 NCCCS Cohorts

	Female			Male		
	(1)	(2)	(3)	(1)	(2)	(3)
Credits earned	0.003*** [0.000]		0.004*** [0.000]	0.001*** [0.000]		0.003*** [0.001]
Progression in program		0.101*** [0.014]	–0.127*** [0.025]		0.035*** [0.025]	–0.117*** [0.033]
Observations	67,781	67,771	67,771	46,978	46,977	46,977
R-squared	0.126	0.125	0.126	0.164	0.163	0.164

Note. Log average quarterly earnings in 2011 as dependent variable. Cohorts of FTIC students from fall 2002 to summer 2005. Sample restricted to students who did not complete an award. Model includes: race (3 groups), single parent, HS graduate, disability, enrollment age (2 groups), experience, experience squared, financial aid EFC, grants, loans, and aid, college GPA first semester. Robust standard errors in brackets.

*** $p < .01$.

Model (1) estimates the returns to credits earned: for each credit, female [male] earnings are subsequently higher by 0.3 [0.1] percent. These are significant gains, showing that dropouts still benefit from having more human capital. Model (2) estimates the returns to progression toward an award: these returns are also positive. Students who progressed further in their program have higher earnings.

---

[7] As with female students, the returns are reduced significantly for male students. For female students, relative to those with associate degrees in unknown/missing subjects, returns to associate degrees in nursing/health are 170 percent/46 percent higher and the returns to all non-completers are 5 percent lower. Reassigning the non-completers reduces the returns to 103 percent/30 percent. For male students, relative to those with associate degrees in unknown/missing subjects, returns to associate degrees in nursing/health are 91 percent/29 percent higher and the returns to all non-completers are 5 percent lower. Reassigning the non-completers reduces the returns to 29 percent/9 percent.

However, because progression is strongly (although not perfectly) associated with more credits, it is not surprising that it is positively associated with earnings. Of interest here is whether, conditional on their credit accumulation, a student who has progressed further toward completion will have higher earnings. Model (3) tests for the net effect of progression: it shows that, among non-completers, controlling for total credits earned, students who have progressed further in their program of study courses earn *less* than students who have taken general credits. The penalty for progression is similar for both female and male non-completers: students who progressed twice as far as the average non-completer earn 4–5 percent less (and the effect of credits has not attenuated in this broader model). To our knowledge there is no prior evidence to corroborate this finding. It may be that there are diminishing returns to subject matter (over general courses) or that taking upper-level subject-specific courses is inconvenient. It may be that these individuals have had less time in the labor market. Overall, this finding does not support theories that students who progress further are more motivated or have higher ability (conditional on them not completing their program). Instead, it suggests a high-stakes scenario where, unless a student is going to graduate, their labor market outcomes would be improved if they dropped out of a program as quickly as possible.

5. Discussion and Conclusion

Most community college students do not complete a credential, even among those who enter college indicating that this is their intention. There is substantial current interest in boosting college completion rates. Along with the increased emphasis on completion, there has also been increased interest in programs of study and on getting students to adopt and stick to a program of study relatively early in their college career, with the view that this will boost completion (Dadgar, Venezia, Nodine, & Bracco, 2013; Jenkins & Cho, 2012). Yet college faculty, administrators, and staff do not typically have a good sense of what credentials their students are pursuing and what amount of progress those students have made toward that goal (Gardenhire-Crooks, Collado, & Ray, 2006; Grubb, 2006). Furthermore, there is little information on the economic value of student progression—as opposed to the wealth of information on the economic value of award completion. There is very little information in the literature on the expected return to pursuing a program, adjusted for the probability of completing it, which is what these *ex ante* results provide.

This information is needed for two reasons. First, if colleges are to raise completion rates, colleges will need a good sense of where their students are in their programs in order to help them move along their program pathways. Even in the case in which a student has explicitly informed the college as to what credential she is pursuing, her actual coursework may not correspond well to her declared major. Second, given low completion rates, the economic output of community colleges is driven partly by the value of human capital accumulated by these non-completers, not just that of the relatively few completers alone.

The challenge is to classify students who do not complete an award. Both stated preference and revealed preference methods may be appropriate. Given students' lack of understanding about program requirements and the complexity of programs offered by colleges, we argue for a greater focus on revealed preference methods to interpret student progress. Certainly, different classification systems—by algorithm, intent, or goal—yield very different enrollment patterns.

More importantly, these classifications make a substantial difference to earnings patterns. Not only does completion of a particular award yield a specific payoff, but failure to complete does as well. For example, we find that across the non-completers, failing a certificate program appears to carry the heaviest penalty. Combining the analysis for completers and non-completers also makes a difference. Completion of an associate degree yields higher earnings than completion of a diploma. And after we account for the probability of completion, this premium actually increases.

By contrast, the results by subject show that returns are compressed once we adjust for the probability of completion. For example, the returns to nursing are extremely high, but the completion rate is very low such that the *ex ante* returns—although still substantial—are significantly diminished. In general, therefore, the widespread finding that subject of study matters a lot to returns to college—although correct—is probably an overstatement of the true differential (Webber, 2014). From a policy perspective, it is important for students to take into account not only the returns to a program, but the probability of actually completing that program.

Finally, we find that progression *per se* does not lead to higher earnings for students who do not complete (even as it demonstrably does for students who complete their program). This result, which is surprising but not implausible, may help explain why so many students do drop out. After taking the relatively valuable general courses at college, these students may find it inconvenient to continue to attend college or may have insufficient motivation, time, or funds for the upper-level courses, which are no more valuable in the labor market but are required for graduation. Viewed from a social perspective, this finding suggests that college personnel in charge of programs might face a high-stakes dilemma: forcing students into a structured program is beneficial if they complete; but if they do not complete, it is more efficient for them to accumulate general credits. Regardless of how rationally students make these decisions, however, considerably more attention should be paid to the human capital obtained by non-completers.

References

Agan, A. (2014). *Disaggregating the returns to college*. Princeton, NJ: Princeton University.

Altonji, J. G., Blom, E., & Meghir, C. (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics*, *4*(1), 185–223.

Andrews, R. J., Li, J., & Lovenheim, M. F. (2012). *Quantile treatment effects of college quality on earnings: Evidence from administrative data in Texas* (NBER Working Paper No. 18068). Retrieved from National Bureau of Economic Research website: www.nber.org/papers/w18068

Bahr, P. R. (2010). The bird's eye view of community colleges: A behavioral typology of first-time students based on cluster analytic classification. *Research in Higher Education*, *51*(8), 724–749.

Bailey, T., Alfonso, M., Scott, M., & Leinbach, T. (2004). *The educational outcomes of occupational postsecondary students.* Washington, DC: U.S. Department of Education, National Assessment of Vocational Education.

Bailey, T., & Belfield, C. R. (2013). Community college occupational degrees: Are they worth it? In L. Perna (Ed.), *Preparing today's students for tomorrow's jobs in metropolitan America* (pp. 121–148)*. Philadelphia, PA: University of Pennsylvania Press.

Bailey, T., Jenkins, D., & Leinbach, T. (2006). *Is student success labeled institutional failure? Student goals and graduation rates in the accountability debate at community colleges* (CCRC Working Paper No. 1). New York, NY: Columbia University, Teachers College, Community College Research Center.

Belfield, C. R., & Bailey, T. (2011). The benefits of attending community college: A review of the evidence. *Community College Review*, *39*(1), 46–68.

Bound, J., Lovenheim, M. F., & Turner, S. (2010). Why have college completion rates declined? An analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics*, *2*(3), 129–157.

Brand, J. E., & Xie, Y. (2010). Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education. *American Sociological Review*, *75*(2), 273–302.

Carneiro, P., Heckman, J. J., & Vytlacil, E. J. (2011). Estimating marginal returns to education. *American Economic Review*, *101*(6), 2754–2781.

Caruana, R. & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. Proceedings of the 23rd International Conference on Machine Learning, pp. 161–168.

Choy, S. P., & Horn, L. J. (1992). *A guide to using postsecondary transcript data and an overview of course taking in less-than-four-year postsecondary institutions.* Berkeley, CA: National Center for Research in Vocational Education.

Compton, J., Laanan, F. S., & Starobin, S. (2010). Career and technical education as pathways: Factors influencing postcollege earnings of selected career clusters. *Journal of Education for Students Placed at Risk, 15*, 93–113.

Dadgar, M., & Trimble, M. J. (2014). Labor market returns to sub-baccalaureate credentials: How much does a community college degree or certificate pay? *Educational Evaluation and Policy Analysis*. Advance online publication. doi: 10.3102/0162373714553814

Dadgar, M., Venezia, A., Nodine, T., & Bracco, K. R. (2013). *Providing structured pathways to guide students toward completion.* San Francisco, CA: WestEd.

DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (2002). A temporal investigation of factors related to timely degree completion. *Journal of Higher Education*, *73*(5), 555–581.

Doyle, W. R. (2009). The effect of community college enrollment on BA completion. *Economics of Education Review*, *28*, 199–206.

Gardenhire-Crooks, A., Collado, H., & Ray, B. (2006). A whole 'nother world: Students navigating community college. New York, NY: MDRC.

Grubb, W. N. (2006). "Like, what do I do now?": The dilemmas of guidance counseling. In T. Bailey and V. Morest (Eds.), *Defending the community college equity agenda* (pp. 195–222). Baltimore, MD: Johns Hopkins University Press.

Hossler, D., Shapiro, D., Dundar, A., Chen, J., Zerquera, D., Ziskin, M., & Torres, V. (2012). *Reverse transfer: A national view of student mobility from four-year to two-year institutions* (Signature Report No. 3). Retrieved from National Student Clearinghouse Research Center website: http://www.studentclearinghouse.info/signature/3/NSC_Signature_Report_3.pdf.

Jacobson, L., & Mokher, C. (2009). *Pathways to boosting the earnings of low-income students by increasing their educational attainment*. Washington, DC: Hudson Institute Center for Employment Policy.

Jacobson, L., LaLonde, R., & Sullivan, D. G. (2005). Estimating the returns to community college schooling for displaced workers. *Journal of Econometrics*, *125*(1–2), 271–304.

Jenkins, D., & Cho, S. W. (2012). *Get with the program: Accelerating community college students' entry into and completion of programs of study* (CCRC Working Paper No. 32). New York, NY: Columbia University, Teachers College, Community College Research Center.

Jepsen, C., Troske, K., & Coomes, P. (2014). The labor market returns to community college degrees, diplomas and certificates. *Journal of Labor Economics*, *32*(1), 95–121.

Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection.* Proceedings of the International Joint Conference on Artificial Intelligence, pp. 1137–1143.

Liu, Y. T., Belfield, C. R., & Trimble, M. J. (2014). *Estimating the medium-term labor market returns to community college awards: Evidence From North Carolina* (A CAPSEE Working Paper). New York, NY: Center for Analysis of Postsecondary Education and Employment.

Melguizo, T., Kienzl, G., & Alfonso, M. (2011). Comparing the educational attainment of community college transfer students and four-year college rising juniors using propensity score matching methods. *Journal of Higher Education*, *82*(3), 265–291.

Mitchell, T. (1997). *Machine learning*. New York, NY: McGraw Hill.

Radford, A. W., Berkner, L., Wheeless, S. C., & Shepherd, B. (2010). *Persistence and attainment of 2003–04 beginning postsecondary students: After 6 years* (NCES 2011-151). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Rish, I. (2001). *An empirical study of the Naive Bayes classifier.* Proceedings of the IJCAI-01 Workshop on Empirical Methods in Artificial Intelligence, pp. 41–46.

Romano, R. M., Losinger, R., & Millard, T. (2011). Measuring the cost of a college degree: A case study of a SUNY community college. *Community College Review*, *39*(3), 211–234.

Stuart, G. R. (2009). *A benefit/cost analysis of three student enrollment behaviors at a community college: Dropout, transfer and completion of an associate's degree/certificate* (Doctoral dissertation). Retrieved from http://etd.ohiolink.edu.

Webber, D. A. (2014). The lifetime earnings premia of different majors: Correcting for selection based on cognitive, noncognitive, and unobserved factors. *Labour Economics*, *28*(0), 14–23.

Zafar, B. (2011). How do college students form expectations? *Journal of Labor Economics, 29,* 301–348.

Appendix

Appendix Figure A.1

Assignment of Students to Programs

Legend:
- Other
- Dipl: Predicted+noaward
- Dipl: Predicted+award
- Cert: Predicted+noaward
- Cert: Predicted+award

Categories: Algorithm, Intent, Goal