I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14483

Labor Market Signaling and the Value of College: Evidence from Resumes and the Truth

Daniel Kreisman
Jonathan Smith
Bondi Arifin

JUNE 2021

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14483

Labor Market Signaling and the Value of College: Evidence from Resumes and the Truth

Daniel Kreisman
Georgia State University and IZA

Jonathan Smith
Georgia State University and IZA

Bondi Arifin
Ministry of Finance, Republic of Indonesia

JUNE 2021

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Labor Market Signaling and the Value of College: Evidence from Resumes and the Truth*

How do college non-completers list schooling on their resumes? The negative signal of not completing might outweigh the positive signal of attending but not persisting. If so, job-seekers might hide non-completed schooling on their resumes. To test this we match resumes from an online jobs board to administrative educational records. We find that fully one in three job-seekers who attended college but did not earn a degree omit their only post-secondary schooling from their resumes. We further show that these are not casual omissions but are strategic decisions systematically related to schooling characteristics, such as selectivity and years of enrollment. We also find evidence of lying, and show which degrees listed on resumes are most likely untrue. Lastly, we discuss implications. We show not only that this implies a commonly held assumption, that employers perfectly observe schooling, does not hold, but also that we can learn about which college experiences students believe are most valued by employers.

JEL Classification: J01, J24

Keywords: signaling, resume, employer learning, statistical discrimination, jobs board

Corresponding author:
Daniel Kreisman
Department of Economics
Georgia State University
Atlanta, GA
30302-3992
United States
E-mail: dkreisman@gsu.edu

* This work does not represent opinions of the College Board, the National Student Clearinghouse, or any other organization.

1 Introduction

Identifying productive from unproductive workers at hire is among the most difficult and costly challenges firms face. A broad literature in economics acknowledge as much in large part by focusing on how employers learn about workers' productivity only after a hire is made. In an effort to distinguish themselves in the hiring process, workers can take steps to signal their productivity to employers, possibly by investing in schooling.

The degree to which this schooling imparts skills to workers or whether it primarily serves as a signal of their pre-existing ability is a perennial question in economics, one that has nearly spawned its own literature. Policy concerns are not far behind. If schooling does little more than help employers distinguish productive from unproductive workers *ex ante*, then large public subsidies for schooling are hard to justify (Spence, 1973), even if there are efficiency gains to be had (Stiglitz, 1975).

Much of the literature on returns to schooling has focused on the value of earning a college degree. Yet the reality is that for every five students who enroll in college, fewer than three graduate. In fact, over the past 20 years more than 31 million Americans have left college without a diploma.[1] Determining the value of those college credits, apart from the signal a degree affords, is a difficult task. We take a novel approach by asking how college non-completers themselves value their schooling, and what implications this has for the literature on labor market signaling, employer learning and returns to college. We do so by focusing on how job-seekers choose to disclose schooling to employers using the most salient and recognizable signal for new hires: resumes.

Building on existing frameworks, we assume that attending college but not completing sends employers two countervailing signals. One positive – attending some schooling and potentially gaining associated human capital, and one negative – dropping out. If the latter outweighs the former, job seekers may elect not to signal to employers that they ever attended college in the first place. Learning whether job seekers strategically omit partially completed schooling tells us if they perceive it to be a negative signal on balance. By observing what characteristics predict this, such as school type, quality, and duration, we can further ask at what level or type of schooling job seekers believe the value to be positive. This results in a test of a commonly held assumption in the literature, that employers accurately observe job-seekers' schooling at the time of hire. If job-seekers regularly omit, or misrepresent, their education to employers, empirical work estimating the return to schooling, in particular for non-completers, is misspecified.

Testing these hypotheses requires a match of two rare sources of data, one showing what job seekers actually reveal to employers and another with the truth. For the former, we use a large sample of recent resumes scraped from a leading nationwide online jobs board. For the latter, we match these resumes to data from the College Board and National Student Clearinghouse (NSC), which contain records of college enrollment, student demographics, and measures of academic ability. To maximize matching across datasets, and to eliminate ambiguity about transfers and multiple

[1] Shapiro et al. (2014).

enrollments, we focus on (male) job-seekers who attended only one college.[2]

We find that one out of every three non-completers in our matched sample of early career job-seekers omit that only college experience from their resume. We also find that omission is systematically related to schooling characteristics, in line with straightforward theoretical predictions. Students who enroll for fewer than two years, and those who attend less competitive institutions, are most likely to omit. In addition, each year of potential experience in the labor market increases the likelihood of omitting, as the schooling signal is less important for workers with more experience. Yet, we find job-seeker race and ability, measured by PSAT scores, have no predictive value.

Job-seekers can also lie about their education on their resume. We find that over 20 percent of our sample of job-seekers lie either about a degree or merely attending, and further demonstrate that both college quality and field of study listed on resumes are statistically related to the probability of lying. For example, students are least likely to lie about earning a degree in easily verifiable fields, such as health care, education and technical trades, and are most likely to lie about a degree in business, humanities or "general studies". Like strategic omissions, this provides suggestive evidence of which types of schooling impart more human capital, and which might be disproportionately signals.

Our main result, that a significant share of job-seekers strategically omit schooling, has implications for policy, for firms, and for the literature more broadly. First, take policy. Our simple takeaway is that many students who complete fewer than two years of schooling in lower quality schools, particularly in two-year colleges, feel they are more likely to get a job if they say they never went in the first place. This concerning statistic fits with an emerging literature showing that many lower quality, often for-profit, schools have little benefit for students, for example Cellini and Turner (2016) and Deming et al. (2016). It also accords with literature on the importance of soft skills in the labor market, for example Heckman et al. (2006) and Deming (2017), of which not completing might provide a negative signal. Results also have implications for a literature on rising degree requirements in response to minimum wage increases (Clemens et al., 2018), increasing skill requirements in response to macro-economic shocks and technological advancements in the economy (Hershbein and Kahn, 2018), and on negative signals on resumes (Kroft et al., 2013). Importantly, it suggests that pushing students into schools from which they are not likely to graduate may be counterproductive, particularly if they accumulate significant debt.

Next, take the economics literature. Our results directly test an assumption in a host of economic models that employers accurately observe schooling for early career workers. For example, employer learning and statistical discrimination models, in the spirit of Farber and Gibbons (1996) and Altonji and Pierret (2001), assume employers have less information about worker productivity at time of hire than a retrospective looking researcher, allowing for the latter to test the signaling value of schooling. A long literature has followed suit.[3] A related literature on "sheepskin" effects, and

[2] Scraping resumes of males increases the likelihood that we match to administrative records as the likelihood of a surname change is much lower. Further, Including female job-seekers introduces sample selection issues as we could only match females who did not change their last name.

[3] Araki et al. (2016); Arcidiacono et al. (2010); Bauer and Haisken-DeNew (2001); Kahn (2013); Kahn and Lange

similarly returns to college credits, suggest that workers see additional returns to each college credit they earn, and receive a benefit above and beyond that for completing a degree (Dynarski et al., 2017; Jaeger and Page, 1996; Jepsen et al., 2014; Kane and Rouse, 1995). In each of these and other related literatures, it is either implicitly or in most cases explicitly assumed that schooling is readily and accurately observed by employers. While a smaller literature deals with survey misreporting (Black et al., 2003; Kane et al., 1999), there is scant work asking whether job-seekers strategically misreport schooling to employers. We test this very assumption and demonstrate in a straightforward manner not only that it does not hold, but that its violation has meaningful implications for the class of models described above and others.

These results also have implications for firms. Screening on schooling is a common practice among employers looking to narrow down large applicant pools. The growth of online jobs boards, increasing the number of applications per job, has allowed for (or possibly led to) explicit credential screening. That non-trivial shares of job seekers omit or lie on resumes suggests that firms would be wise to exercise caution when using schooling to screen resumes. Our data suggest they may be weeding out candidates with college experiences, and letting in many who are not what they claim.

Perhaps the most novel implication of our analysis is a demonstration of the power of an untapped resource for researchers in resumes. While work has begun to emerge using data from online jobs boards (Clemens et al., 2018; Deming and Kahn, 2018; Helleseter et al., 2018; Hershbein and Kahn, 2018; Kuhn and Shen, 2012; Marinescu, 2017), these studies use job posting information, providing little insight into what job-seekers actually put on resumes, and hence what employers observe. How job-seekers present themselves to employers holds vast potential for study, and the explosion of online jobs boards provides untold information on tens of millions of job-seekers. In addition to ours, only a few papers have taken up this opportunity, though in different circumstances. Shen and Kuhn (2013) and Kuhn and Shen (2015, 2016) use data from a private sector jobs board in China and in Mexico to study hiring dynamics. A recent working paper by Schubert et al. (2019) uses U.S. resumes to track mobility, though these are not linked to administrative records. Thus, we aim not only to provide insights into key questions about returns to skill in the labor market, but also to demonstrate that online resume postings are a potential source of "big data" for future research on employment, skills and returns to human capital investments.

The paper proceeds as follows. In Section 2, we discuss our data sources and matching process. Then we discuss the construction of our analytic samples and additional variables. Section 3 presents a theoretical and empirical framework on omitting schooling on a resume. Section 4 presents main results, which includes descriptive statistics on omitting, along with regression analyses. Section 5 briefly presents a framework and results for lying on resumes about colleges attended and degrees earned. In Section 6 we discuss implications for several classes of empirical models estimating returns to schooling. Section 7 briefly concludes with suggestions for future work.

(2014); Lange (2007); Light and McGee (2015a,b); Mansour (2012); Oettinger (1996); Schönberg (2007).

2 Data

2.1 Resumes: From an online jobs board

We collect resumes from a large, national online jobs board. The board allows employers to list vacancies and also serves as an aggregator of job postings elsewhere on the web. For job-seekers, the service is free to use. To access the site, job-seekers sign up with their name, location (zip code), an email address and phone number (these last two we cannot observe), and can then choose whether to make their resume private or public. We do not know what share of resumes are private, but the volume of public resumes suggests that the private share is low. All resumes made public can then be searched by potential employers, while private resumes are only seen by employers when applicants apply to a specific posting.

We initially collect over 556,000 resumes from the online jobs board in the fall of 2016 and spring of 2017. Our scraping procedure identifies the most recent 1,000 resumes from each zip code in each of the largest 100 U.S. cities. We then normalize the number of unique resumes taken from each city to the relative size of the city, allowing us to economize on scraping time and to generate a representative draw from the sample frame.[4] Using first names, we scrape resumes that are most likely to be male.[5] The reason for this is that females are more likely to change their surnames in marriage, which would make matching selective to those who never changed their name, potentially inducing bias and lowering match rates. Among matched resumes, we observe self-selected gender as entered by test-takers in the College Board. Because job-seekers enter all resume information into uniform fields rather than uploading created resumes, we are able to use the web site's meta-tags to parse out each field with precision.

Job-seekers are asked to list work experience sequentially, including job title, company, location, start and end dates, and a description of duties and accomplishments. They are also asked to create an entry for each school they attended, including school name, degree, field of study, location, and start and end dates. There are then several additional fields job-seekers can fill out, including skills, their objective, their eligibility to work in the US, willingness to relocate, and an option to add additional information. We rely only on resumes from job-seekers who filled out the education section for matching and analysis purposes, implying that omissions are not simply the result of skipping part of the resume. Additionally, we note that the education section is in the middle of the resume, and that all resumes we use contain information both before and after that section. Finally, we remove the small number of resumes that are largely incomplete from our sample.

[4] For example, if 77449 (Houston) is the most populous zip code, we scrape the most recent 1,000 resumes from that zip. Then if zip code 30307 (in Atlanta) is 16 percent as populous as Houston, we scrape 160 resumes from that zip. An alternative would have been to scrape the most recent 1,000 from all zip codes and weight the regression. Our procedure economizes on scraping time and weights locations proportionately while maintaining representativeness in urban areas.

[5] We identify names that are almost exclusively given to male children according to social security files by first extracting names from a website query and keeping only those which have a probability near one of being male according to social security.

2.2 The Truth: Administrative Educational Records

We make use of College Board data from the graduating high school cohorts of 2004 to 2014, which provide demographic and background information on students who ultimately become job seekers. These records contain information from the over two million students per high school cohort, over 9 million males in total, who take at least one of the PSAT, SAT, or AP exams. These data contain students' self-reported race and ethnicity, among other demographic characteristics, including high school and cohort. We use students' PSAT scores as our primary measure of academic ability.[6]

These College Board data are merged with records from the National Student Clearinghouse (NSC), which contain information on college enrollment for approximately 94 percent of college students in the U.S.[7] The most notable deficiency is for-profit college enrollment, though many of the largest for-profits are included.[8] The data track all spells of enrollment at participating colleges, measured in days, whether students graduate, and if they graduate, their field of study.

We supplement these records with information about colleges students attended through the Integrated Postsecondary Education Data System (IPEDS). For our purposes, we are interested in college "quality", which we proxy with whether the college is two- or four-year, and the average PSAT score of students enrolled.[9]

2.3 Matching Resumes to the Truth

We take two approaches to matching administrative College Board records to resumes, resulting in two distinct, though sometimes overlapping, samples. The first method matches records on exact first and last name and high school attended. The second method matches records only on exact first and last name but only relies on unique names across both data sources.

In both samples, we only consider matches who have exactly one NSC-listed college and did not earn a degree, regardless of what education was listed on their resume. Non-completers are the central question of this paper and we focusing on job seekers who only attend one NSC college for two reasons. First, a job-seeker cannot omit a college if he never attended, which accounts of a large share of attrition in matching. Second, this simplifies our thought experiment insofar as we limit to job-seekers who omit the one and only (NSC-listed) college they attended. Further it avoids ambiguity for students who transfer schools, possibly omitting the first or shortest instance, which is substantively different than omitting college attendance wholesale.[10]

[6] For a small subset of students who do not have PSAT scores, we use SAT scores and include an indicator for which test they took, noting that results are not different if we limit only to one test or the other. The PSAT is often thought of as a precursor to the SAT, which is one of two college entrance exams, but it also qualifies students for scholarships and other awards; in many schools it is taken by all students.

[7] College Board serves as the base data and NSC data were merged on prior to our project.

[8] See Dynarski et al. (2015) for a discussion of limitations of NSC data. In robustness checks we check for those who attended training/schooling outside of NSC coverage with no change to results.

[9] See Smith and Stange (2016) for a summary of this measure.

[10] It is unclear how to interpret omitting one school for a job-seeker who attended two. For example, if a job-seeker started transferred schools, omitting the first instance might not be concealment, but rather just reflecting listing the school the student spent the most time at.

The "High School" Sample

The first method matches records on exact first and last name, high school (name and location in case of common school names), and high school graduation year within two years. We would obtain a false positive match only in the case that two individuals both went to the same high school, graduated within two years of one another, and had the exact same first and last name, and only one posted a resume on the jobs board. From administrative records we estimate that the upper bound for this is smaller than 1 percent.[11]

The top panel of Table 1 begins with the full sample of scraped resumes. Column 2 limits to those that listed no employment before 2001. Since our administrative records begin with the high school graduating cohort of 2004, anyone with work experience before 2001 is likely too old to be in our data, and not eligible for a match, though this does not rule out many resumes from workers who could not be in our data, including those who did not take a College Board exam. In column 3 we show characteristics of resumes that listed a high school and graduated in a cohort that could be in our data. These resumes look strikingly similar to those in column 2 that did not list a high school, with exception for the number of schools listed. In column 4 we show the sample of 11,244 resumes that matched to College Board records, which again closely resembles column 3, suggesting that those listing high schools are not observationally different. Columns 5 and 6 then limit to those who attended only one college in NSC records,[12] and among those, the group who did not earn a degree. In sum, comparing our matched and final samples, in columns 4-6 closely resemble the resumes of job-seekers who could potentially match to our data in column 3.

While this procedure minimizes the potential for false matches, which would inflate omitting rates, it restricts the sample to resumes listing a high school and graduation year. Listing a high school on a resume is a decision which be more prevalent among those not listing a college. This is the motivation for our second matching method, which does not rely on listing a high school, described below.

The "Unique Names" Sample

Our second method matches records on exact first and last name, but only for names that are "unique". For this match we first find unique first and last name combinations among the the 12 million males in our administrative educational records that span 11 years. We do the same for the hundreds of thousands of resumes that did not list a college experience prior to 2004. Some of these names are likely not truly unique in the U.S. but they are certainly rare. Column 2 in the bottom panel of Table 1 shows the over 210,000 resumes that have unique first-last name combinations. In column 3 we show the 36,300 of these that matched to unique names in our administrative records.

We implement further restrictions to increase the likelihood that the matched names are accurate matches. First, we eliminate initial matches where a resume's work experience indicates employment

[11] We calculate the percent of students in a high school across the 11 sample years with the same name to be about 1 percent. This is an upper bound because our matching process focuses on students within two years of one another, not across all 11 years.

[12] Approximately 55 percent of the matched sample removed for not having one NSC college (going from column 4 to column 5) are a result of having zero NSC colleges, the remainder being two or more NSC colleges.

prior to age 16 as defined by our administrative data, and only retain those where resume location is in the same state as high school from administrative data, both of which are conservative decisions and rule out some potential matches. After these restrictions, we find 13,895 matches of unique names from both samples (column 4 of Table 1). As before, we focus on the job-seekers who only attended one college (coulmn 5)[13] and further, those who did not complete (column 6).

The final unique names sample consists of 4,384 job-seekers. Relative to all resumes and to the high school sample, the unique names sample lists fewer educational experiences. They also have fewer years worked than typical resumes, since the job-seekers tend to be younger than the general population. We also show that 62 percent of the matched job-seekers who attended one college dropped out of college and remain in our final sample. In contrast, the matched high school sample saw a 86 percent non-completion rate.

2.4 Summary Statistics

Summary statistics describing job-seekers in the high school sample are in the first column of Table 2. About half the job-seekers attend a four-year college (versus a two-year college) for an average of 0.85 years and about one-third are White, one-third are Black, and one-fifth are Hispanic. The average PSAT score for the job-seekers is an 86, which is similar to the average PSAT of the colleges to which they enroll (89).

The right panel of Table 2 displays summary statistics for the 4,384 unique names analytic sample of job-seekers who only attend one college but do not earn a degree. Approximately 62 percent of these job-seekers attend a four-year college, which is 12 percentage points higher than the high school sample. This is mirrored with slightly higher student and college PSAT scores and an average length of study of 1.1 years.

We compare our samples to the larger exam taking population in Table A1, focusing on males who lived in an MSA in high school and attended one college but did not earn a degree, similar to our job-seekers. Our matched samples are much more likely to be Black and less likely to be White than the full educational data. In the high school sample, job-seekers have lower PSAT scores and attend colleges with lower average PSAT scores, frequently two-year colleges, than in the educational data. In the unique names sample, the PSAT scores and college-going is more similar to the educational data sample.

Finally, we note that 597 observations are common to the two samples, which allows us to test the quality of our match in the unique names sample. Let us assume that the high school sample has almost no false positive matches because it has an additional piece of information to match on other than names. We find that of the 597 observations of unique names that also include a high school on their resume, all of those made it into our high school sample too. In other words, after matching on only unique names, the high school matched 100 percent of the time when high school was listed.

[13] Approximately 42 percent of the matched sample removed for not having one NSC college (going from column 4 to column 5) are a result of having zero NSC colleges, the remainder being two or more NSC colleges.

2.5 Additional Measures

We create several key variables for our analysis from the combined set of data. Of primary interest is whether a job seeker omitted post-secondary schooling from his resume. To do this we map each college listed on a resume with each college listed in NSC records by hand. We define *Omit* equal to one if the school exists in NSC records but is omitted from the resume, and zero otherwise.

We also identify colleges or other training listed on resumes that are not found in NSC records. There are a few cases where this arises. The first are colleges not covered by NSC. These are mostly private and for-profit. Second, the majority come from the many job-seekers who list what we call "non-collegiate training", for example Job Corps or highly specialized job specific training offered by companies or third-party vendors. We define any schooling listed on the resume that is not in NSC as a binary indicator called non-NSC "training". Shown in Table 2, just fewer than 30 percent of job-seekers in both sample have some other training on their resumes. In robustness checks we limit to job-seekers with no non-NSC training with little change to results.

Finally, we calculate a measure of non-employment from the number of months not working since exiting schooling. We construct potential experience as the difference between the last date the resume was updated and the approximate date the student left college. We then subtract off work spells listed in work experience, which leaves an estimate of months not working.[14] In both samples, the average job-seeker has about 8.5 months listed of non-employment. We note that just as job-seekers can omit or lie about schooling, job histories are self reported and are similarly subject to fabrication and omission. If job-seekers who omit schooling are also more likely to inflate work histories, and if we expect a negative relationship between omitting schooling, or lying about schooling, and employment, our estimates of this relationship will be biased (upward) toward zero. If we expect a positive relationship, the bias will be opposite signed. Because we lack a resource with which we can verify work histories, we cannot directly address this issue.

3 Omitting Schooling

We begin by outlining a framework for omitting schooling on resumes, and then describe our empirical model that tests predictions.

3.1 Framework

For simplicity, let us assume that college consists of two periods and that those attending either complete one period of college or two, after which they would earn a degree. Assume also that job-seekers make the decision to put or omit schooling on resumes after schooling is completed, knowing at enrollment that the option to do so exists, which lowers the cost of a risky investment.

[14] In some cases job-seekers only list first and last year of employment, and do not put months. In these cases we count years of employment and divide by 12, which will result in measurement error in the dependent variable.

We characterize the *ex post* decision to omit schooling on a resume as:

$$\Omega_{ij} = f(degree_{ij}, qual_j, years_{ij}, exper_i) \tag{1}$$

where $\Omega_{ij} = 1$ indicates if job seeker i omits college j from his resume. We take this to represent when job-seekers believe the likelihood of employment is higher when schooling is omitted, which is a function of the signal of schooling experience j. We assume that all degree earners, as measured by $degree_{ij} = 1$, list their highest degree accurately, though we cannot rule out fear of over-qualification. We in fact find negligible instances of omitting among degree earners, and focus our empirical work on non-completers.

However, non-completers are faced with a choice. They can either signal that they completed one period of schooling, simultaneously signaling that they did not complete, or they can signal that they completed no schooling. By omitting schooling, omitters conceal any negative signal associated with not completing, which comes at the cost of losing any potential value from signaling some completed schooling, and further of incurring the signal of a high school completer who never attended college. This decision will in part depend on whether (job-seekers believe) employers expect lower productivity from non-completers, conditional on school quality, or terminal high school graduates. We describe this relationship as follows:

$$\Omega_{ij} = \mathbb{1}[P_{emp}(hs_i, exper_i) \geq P_{emp}(dropout_{ij}, qual_j, years_{ij}, exper_i)] \tag{2}$$

where P_{emp} is the probability of employment and Ω_{ij} equals 1 if the right hand side of the equation holds and zero otherwise. In this framework, the job-seeker is weighing his perceived probability of employment from omitting compared with his perceived likelihood of employment when he includes partially completed schooling on his resume.

What signal the employer observes depends on whether job-seeker i omits or not. If that individual omits schooling, employers observe $(hs_i, exper_i)$ – that he is a high school completer and his years of listed work experience. If he does not omit and did not graduate, employers observe $(dropout_{ij}, qual_j, years_{ij}, exper_i)$, which includes years of schooling, school quality, and the signal of not completing. This trade-off weighs the human capital and signaling value of non-completed collegiate schooling.

The signaling value reflects costs, for example effort associated with enrolling for one period. We might believe that these are larger at more selective, or higher "quality", schools ($qual_j$) as they may require more effort to attain the same grades or outcome relative to less selective schools. If so, the likelihood of omitting is strictly decreasing in the quality of school j. Similarly, within any school j we expect the likelihood of omitting to be decreasing in years of enrollment ($years_{ij}$) if we assume these costs are cumulative. We note that higher quality schools may also come with larger student support systems, which could imply a stronger negative signal from not completing as quality increases.

Alternatively, we can consider the role of human capital. If students learn valuable labor market skills while at school, human capital increases with the quality and years of education. This implies that the human capital component of schooling, as measured by $qual_j$ and $years_{ij}$, would decrease the likelihood of omitting, just as in the case where we assume no human capital. Hence, regardless of whether schools impart human capital or not, omitting should be unambiguously non-increasing in quality and years enrolled.

Finally, we draw on assumptions in Farber and Gibbons (1996) and Altonji and Pierret (2001) that work histories are observable to the entire labor market, implying that the signal value of schooling declines over time. This implies the likelihood of omitting is decreasing in years of work experience, $exper_i$. However, we note that the negative signal of dropping out may also decrease over time and so the net effect of work experience on omitting may be ambiguous.

Note that our model has no theoretical predictions concerning the relationship between omitting schooling and characteristics specific to person i other than experience. The framework above describes the decision as a function of characteristics of the signal associated with school j and time in the labor force. In our empirical tests below we include person level covariates, such as race and a measure of job-seekers' ability, to test if this is true.

3.2 Empirical Tests

To test these predictions we estimate the following reduced-form model:

$$\Omega_{ij} = \mathbf{\Pi}(4Year_{ij}, \overline{PSAT}_j, Years_{ij}) + \mathbf{\Psi}(Exper_i, Train_{\sim j}) + \mathbf{\Upsilon}(Race_i, PSAT_i) + \tau_t + \epsilon_{ij}. \quad (3)$$

Above, Ω_{ij} is a binary indicator equal to 1 if job-seeker i omitted schooling j from his resume. We measure quality in two ways – $4Year_j$ is an indicator if the school i attended is a four-year as opposed to a two-year school, and $\overline{PSAT}_j$ is the mean PSAT for attendees of school j. While these are not necessarily measures of quality, they describe some measure of difficulty or selectivity to which employers may respond. $Years_{ij}$ are simply years of schooling.[15] Thus the first two coefficients of the vector $\mathbf{\Pi}$ test predictions concerning school quality, while the last tests a prediction about years attended.

The next set of covariates represent the relationship between experience and omitting, captured in the coefficient vector $\mathbf{\Psi}$. Our measure of potential experience is years since graduating high school, less years of post-secondary schooling, as is common in the literature. We also include an indicator $Train_{\sim j}$ if i put any other non-NSC, post-high school training or schooling on his resume that is not recorded in our administrative records.

Finally, in some models, we include indicators for race (Black, Hispanic, Asian, other) and a measure of ability in $PSAT_i$.[16] Similarly as ability, which we proxy with PSAT score, is unobserved by employers at time of hire, we might expect that the decision to omit is uncorrelated with

[15]NSC measures enrollment in days enrolled, which do not necessarily correspond to years of completion. We test robustness to this in subsequent checks.

[16]Note we use a scaled SAT score for a very few number of observations with no PSAT score and include an indicator of which test is used.

ability measures observed by the econometrician conditional on $Qual_j, Years_{ij},$ and $Exper_i$. In some analyses, we include a set of high school graduation year fixed effects, τ_t, to capture secular cohort effects.[17] We estimate the model with ordinary least squares where ϵ_{ij} is an idiosyncratic error term.[18]

4 Omitting Results

4.1 Omitting: Mean Differences

Table 2 shows summary statistics for both samples and by omit status. The bottom of the leftmost panel shows our first key statistic, that 33 percent of resumes in our high school sample omit the one and only NSC college they attended but did not earn a degree in. Despite the substantially different sample construction, the unique names sample in the right panel shows the omit rate is 32 percent. We take these simple statistics as straightforward evidence that a non-trivial share of job-seekers who started schooling but did not complete omit this from their resumes.[19]

We next ask whether the patterns of omission are selective in a manner predicted by our theoretical framework. We start by comparing omitters and non-omitters in the high school sample. We find that omitters are less likely to have attended a four-year school, enrolled for about half the number of years, went to schools with lower average PSAT scores, and were twice as likely to have other training listed on their resumes. Mean differences all correspond to predictions above. Yet, we find omitters are more likely to be non-White and have lower PSAT scores themselves. These run counter to our predictions that individual characteristics beyond experience should not predict omissions. These results are quite similar in the unique names sample. We next consider these factors jointly by estimating Equation 3.

4.2 Omitting: Regression Estimates

In Table 3 we present results which describe omitting as a function of school and student characteristics. Column 1 contains college enrollment and college quality measures – years enrolled, potential experience, whether the school is a four-year college, and the average PSAT of the college's enrollees. The resulting estimates are in line with our theoretical predictions. Each year of completed schooling reduces the likelihood of omitting by nearly 17 percentage points. This suggests that two years of schooling would negate the 0.33 unconditional probability of omitting. We also find that job-seekers attending a four-year school or a school with a higher average PSAT, make them less likely to omit. While the four-year indicator is not statistically significant, this is due to a high correlation with average college PSAT. Omitting average PSAT moves the coefficient on four-year school away

[17] Graduation year fixed effects, potential experience, and years in school are perfectly colinear, so we we frequently exclude the graduation year fixed effect.

[18] Results are unchanged if we use a logistic regression.

[19] In results not shown, almost all matched job-seekers who earned a degree, who we drop from our samples, report their schooling on the resumes. 3.9 percent and 6.9 percent omit in the high school and unique names samples, respectively, the latter of which is inflated by potential mismatches in unique names.

from zero and increases statistical significance. Additionally, other training strongly predicts the likelihood of omitting, suggesting that job-seekers are more comfortable omitting schooling when there is an alternative to highlight.

In the next two columns we turn to our predictions concerning student characteristics. Adding students' PSAT scores results in a precisely estimated zero and does not change the other coefficients. In the last column of the high school sample we find that conditional on schooling, race is unrelated to omitting. These results with student characteristics suggest that school quality, and not observable individual characteristics, explain selectively omitting schooling. Yet, we cannot rule out unobservable individual characteristics predicting omitting that might be correlated with these observables. One example might be savviness or sophistication. If it is the case that omitting is beneficial for job-seekers, more sophisticated applicants will realize this and be more likely to act on it, noting that we rule out ability effects as measured by $PSAT_i$.

The right panel conducts the same analysis with the unique names sample and the main conclusions hold with a few subtle differences. First, job-seekers who dropped out of a four-year college are 8 percentage points less likely to omit than someone who attended a two-year college, even after adding the college average PSAT and a full set of controls, keeping in mind that this sample also has a higher four-year attendance rate. Further, the coefficient on other training is about 0.6 the magnitude as in the high school sample but remains economically and statistically significant.

4.3 Omitting: Robustness

In order to ensure that our results are not reliant on samples or definitions, we explore several robustness tests to our main results. In Tables A2 and A3 we test several sample restrictions for our high school and unique names samples, respectively, showing that our results are robust to a host of alternative rules. We start by replicating our main result in column 1 of each table, and then impose several sample restrictions, including: restricting the sample to those who left high school before 2012 to ensure students had time (6 years) to complete college (column 2); estimating models separately for two- and four-year colleges (columns 3 and 4);[20] limiting to students who were enrolled for at least one or one-half calendar year (columns 5 and 6);[21] and finally, limiting it to those who list no other schooling or training (column 7).[22] We find omitting patterns are consistent across each of these subsamples in both the high school sample and unique names sample.

We also ask whether our linear version of potential experience is correct, since we believe what goes on a resume may change as someone gains experience in the labor market. We find that

[20] We also find that the coefficient on years enrolled is nearly precisely twice as large for two-year schools compared with four-year, which corroborates a story where proportional progress toward degree is the salient mechanism. This corresponds strongly with early work by Kane and Rouse (1995), who find that returns to a credit of two- and four-year schools are nearly identical. We also note that base omit rates are higher for two-year attendees than four-year, shown at the bottom of the table.

[21] Our enrollment is measured in days, hence we do not have measures of enrollment intensity, such as course loads. The measure here limits to students who were enrolled for at least 365 consecutive days according to NSC.

[22] This addresses the concern raised by Dynarski et al. (2015) who show that NSC coverage is not perfect.

including potential experience linearly is a good approximation for how potential experience relates to omitting and all other variables remain stable too. Table A4 shows the results of three regressions where potential experience enters as a set of dummy variables for different amounts of experience, as a binary variable for little experience, and again as a binary variable but interacted with other variables. Results in both samples support our initial analyses but reinforce that omitting is much more common among experienced job-seekers, as expected.

5 Lying

In a final set of analyses we focus on a well known though not deeply explored phenomenon – the decision to lie about schooling on resumes. According to ADP, who conducted 2.6 million background checks in 2001, 41 percent lied about their education, with 23 percent falsifying credentials or licenses (Babcock, 2003; Wood et al., 2007). Further, popular press has unearthed numerous examples of high-profile lies about schooling.[23]

Relative to omitting, where job seekers selectively choose not to disclose information, the decision to lie also includes the probability and cost of detection. If we assume the cost is constant across occupations, for example dismissal and a reputational cost, then theoretical predictions focus on the probability of detection, net of benefits from the positive signal. For example, teaching or medical professions often require certifications, making detection easy and lying worthless. On the other hand, having claimed enrollment, but not a degree, in a general program such as liberal arts or business may be difficult to detect, though may have lower value in the labor market. Alternatively, claiming a computer science degree but not being able to program is easily detectable, though a self-taught computer programmer may have no difficulty passing.

Thus we can model the net benefit in expectation of lying as a trade-off between the value of the signal to employers, less the likelihood that one is found to be lying. We carry this thought process to our analysis by focusing on what job-seekers are most likely to lie about, in particular which degrees. This has two practical applications. First, we find that lies about degrees are more common than lying about simply attending, possibly making a comment on (beliefs about) employer screening. Second, lies about degrees are usually accompanied by information about what the job-seeker claimed to have studied. We also observe lies about attendance and include these in our analysis for completeness. We discuss our sample, definition and methods followed results below.

[23] For example Ronald Zarrella, the CEO of Bausch & Lomb falsely claimed an MBA from New York University, costing him $1 million dollars from his employer. Jack Grubman, a star analyst at Solomon Smith Barney claimed to have attended MIT, which was untrue. Scott Thompson, the CEO of Yahoo! claimed a non-existent degree in computer science. Dave Edmondson, the CEO of Radio Shack made a similar false claim about degrees in Psychology and Theology. Marilee Jones, the MIT Dean of Admissions claimed 3 false degrees. Jeffrey Papows, President IBM claimed a PhD from Pepperdine. Liv Loberg, a Member of Parliament in Norway, falsely claimed several degrees resulting in 14 months in prison.

5.1 Sample, Definition of Lying and Empirical Model

Staring with over 11,000 resumes matched to administrative data on any three pieces of information (full name, any school, including match on college, and attendance year), we focus on the set of students who, according to their resumes, attended zero (to allow for lying about attending) or one (to allow for lying about completion) NSC college. We also restrict the sample to students in high school graduating cohorts prior to 2011 to give students adequate time to complete their degrees. This leaves 4,154 job-seekers. These job-seekers are older than the omitting sample, have fewer colleges listed (by definition), and have more work experience and fewer months not working.

We begin by creating two versions of lies by hand checking resumes with our administrative records to account for differential spellings or abbreviations of school names and degrees. To identify lies about attendance, each college on a resume is matched to an IPEDS code. We then determine whether the school contributes to the NSC database. If it does, we check whether the college on the resume is in the student's official NSC record. If not, it is coded as an attendance related lie.

To identify falsely claimed degrees, we observe whether a job-seeker lists a degree on his resume and then observe whether NSC data verifies it. If not, we categorize the degree as a lie.[24] The online jobs board allows job-seekers to enter fields of study flexibly. While the field is often blank for non-graduates, it is typically populated by those listing degrees. In our procedure for identifying degrees on resumes we first find any reference to a degree or certificate, including abbreviations, though we do not consider "diploma" which might not be covered by NSC. We then classify major or field of study. We allow these to be non-mutually exclusive as many job-seekers list more than one field of study. We categorize fields into the following list: business, education, humanities, social sciences, engineering and computer science, natural sciences, arts, technology (other than computer science), technical or trade degrees, health, communications, criminal justice and general studies.

5.2 Evidence of Lying and Field of Study

Table 4 shows summary statistics of the sample and fields listed in true and falsely claimed degrees. We find that 20% of the sample lies about either attending or graduating from college. 7% of job-seekers never attended the college listed on their resume, and 16% lie about graduating, while a much smaller subset lie about both. Truth-tellers and liars appear quite similar in demographics. We do find that job-seekers who lie about schooling have higher PSAT scores on average.

We focus on what pieces of information on a resume about colleges predict lying, in particular which fields of study. We model this using the following specification:

$$Lie_{ij} = \underbrace{\mathbf{\Pi}(4Year^*_{ij}, PSAT^*_j) + \mathbf{\Theta}(Field^*_{ij})}_{\text{Characteristics of \textit{claimed} schooling}} + \mathbf{\Upsilon}(Race_i, PSAT_i) + \tau_t + \epsilon_{ij} \tag{4}$$

[24] The college enrollment and completion data are truncated at six years after high school graduation. The relatively few students who do graduate after six years will appear as non-graduates. However, for the 2004 cohort, we have seven years of college enrollment and completion data. The small sample does not allow a detailed analysis but the main statistic on the rate of lying is 27 percent, which is larger than in the full sample.

Above, Lie_{ij} is a binary indicator equal to 1 if person i lied about earning a degree from school j. The vector of variables associated with coefficient vector $\mathbf{\Pi}$ describes college-going characteristics listed on resumes, where the asterisk indicates that this information is not necessarily true. The same is true with field of study, $Field^{*}$. Thus, when one of these characteristics is observed, the coefficients reveal the relative likelihood that it is a lie. We estimate a linear probability model and in some specifications control for student characteristics and characteristics of the school the student actually attended.

Table 5 shows results from Equation 4 for field of study among degree claimants. All coefficients should be interpreted in relation to Business and Management, the omitted category. We also plot coefficients and confidence intervals from the full model relative to Business in Figure 1.

We find lies are most likely to be about earning a degree in Business, Social Sciences, and the Humanities. This is consistent with a story where these degrees have few specific and verifiable associated skills, making it easy to list and hard to verify. We also find lies are least likely to be about Health, Technical and Trade, Education, and General Studies degrees. Jobs in Health and Education often require certifications that employers are required to verify by law. Lying about these might be counterproductive.[25] Similarly, Technical and Trade degrees are often in fields that impart specific skills, for example precision machining, which employers could easily verify from observing output. These are often accompanied by certifications and licensures as well. General Studies degrees typically come from two-year colleges, which may have less value in the labor market than a four-year degree. We find little evidence that job-seeker characteristics (race and PSAT) predict lying.

6 Theoretical and Empirical Implications

Our results above demonstrate that a key assumption on the theoretical and empirical work on the returns to schooling – that employers perfectly observe schooling, or that they observe schooling in the same way the researcher does – is violated. In this section, we focus on the implications to employer learning and statistical discrimination models (EL-SD) and a related class of models[26] We also discuss implications for related work on returns to credits and "sheepskin effects."

6.1 Basic Employer Learning Model

We start with the basics of employer learning models as in Farber and Gibbons (1996) and Altonji and Pierret (2001); the intuition for which is as follows. Retrospective looking researchers with panel data observe a measure of ability, schooling, tenure, and wages over time. Employers observe

[25] Few job-seekers in our sample have education degrees (true or not), likely reflecting the fact that positions in schools are not normally secured through online jobs board; hence coefficients here are very noisy.

[26] For example the speed of employer learning (Lange, 2007), employer learning and school quality (Araki et al., 2016), by dimensions of skill (Light and McGee, 2015a), across occupations (Mansour, 2012), and many others, but also to a broader literature on returns to college credits (Jepsen et al., 2014; Kane and Rouse, 1995; Zimmerman, 2014), sheepskin effects (Hungerford and Solon, 1987; Jaeger and Page, 1996), and even resume audits (Bertrand and Mullainathan, 2004).

only schooling at initial hire and then observe (noisily) productivity as workers accrue experience. If ability measures (unobserved by employers at hire, but observed by researchers) are predictive of productivity, they should become more salient in wage determination over time as ability is (slowly) revealed to employers through (noisily observed) productivity. Hence, the estimated relationship between schooling and earnings should weaken over time as employers learn about true productivity. The original model as laid out in Altonji and Pierret (2001) is as follows. Productivity, y, for worker i with t years of labor market experience is:

$$y_{it} = rs_i + \alpha_i q_i + \Lambda z_i + \eta_i + \tilde{H}(t_i). \tag{5}$$

Where s are productive characteristics that are observable to both employers and the researcher; the most common case is schooling. q are characteristics observable only to employers not seen by the researcher, for example reference letters, or in some cases school quality. z are measures of productivity observed by the researcher but not the employer at time of hire; in most cases this is a test score. η are measures observable to neither the researcher nor the employer. $H(t)$ is then the structural relationship between experience and productivity which, by assumption, does not depend on either s or z.

With a few assumptions, including a positive correlation between s and z, Altonji and Pierret (2001) show that r is non-increasing over time, and that Λ is non-decreasing. In other words, that wages become increasingly correlated with the researcher's proxy for productivity (employer learning) and less correlated with schooling (statistical discrimination). The authors, and many others listed above, test this by estimating a variant of Equation 5 where the dependent variable is log wages and the model includes interactions between s and z and a measure of experience, t, along with worker covaraites, X, as below:

$$log(wage_{it}) = \beta_0 + \beta_1 s_i + \beta_2 z_i + \beta_3(s_i \times t) + \beta_4(z_i \times t) + X_{it}\Gamma + \varepsilon_{it}. \tag{6}$$

The model predicts that β_1, the coefficient on schooling at hire (when experience, t, is 0) will be positive (statistical discrimination), and the coefficient on β_2 will be 0, as only the researcher observes the ability measure at hire, though correlations between z and s do not necessitate this assumption. Then also that the coefficient on β_3 will not be positive, and β_4 will be positive as wages increasingly track previously unobserved ability and become more weakly correlated with schooling (employer learning). Broadly, these predictions are borne out in the literature.

6.2 Employer Learning With Omitted or Falsely Claimed Schooling

Our empirical results above demonstrate that job-seekers strategically omit (or lie about) schooling. We now consider what implications follow for the class of EL-SD models as above. To do so, we rewrite the basic EL-SD model from Equation 5 above to allow for omissions (or lies). Here, Ω refers to whether the job-seeker omits schooling, as in Equation 3.

$$y_{it} = r(\Omega_i s_0 + (1 - \Omega_i)s_i^*) + \alpha_i q_i + \Lambda(z_i, \Omega_i(s_i^* - s_0)) + \eta_i + \tilde{H}(t_i). \tag{7}$$

If job-seeker i omits schooling such that $\Omega = 1$, then employers observe schooling level s_0, the base level of schooling (for example high school). Then for those who omit, the difference between true schooling and what the employer observes, $(s_i^* - s_0)$, which is positive, is now a component of z. This results in a weak test of the signaling versus human capital value of (partially completed) college.

If omitted schooling, the difference between s_i^* and s_0, in fact has productive value that employers learn about as workers accrue experience, the schooling gradient with respect to experience, β_3 in the regression model, will be flatter (less negative) for omitters as they listed less employment than the researcher observes. It follows then that the experience gradient on ability, β_4, will be steeper (more positive in this case), assuming schooling and ability are positively correlated. This is because employers are learning about skills that workers did not include on resumes but researchers assumed they did. If, on the other hand, omitted schooling in fact imparted no productive skills, then the experience profile with respect to schooling and ability will be no different for omitters than non-omitters. This also implies that the coefficient on schooling when experience is zero, β_1, should be smaller for omitters, as they should expect to receive a starting wage equal to those who only completed high school. If it is the case that employers expect that many job-seekers omit schooling, then β_1 will reflect a weighted average of expected productivity of high school graduates and the share of college dropouts who omit.

For lies, the pattern is converse. If a job-seeker lies ($\Omega = 1$), then employers observe a schooling level, $\tilde{s}$, that is higher than true schooling, s^*. Now z, the unobservable productivity component at hire, also includes a negative term, which is the difference between observed and true schooling $(\tilde{s} - s_i^*)$.

$$y_{it} = r(\Omega\tilde{s} + (1 - \Omega_i)s_i^*) + \alpha_i q_i + \Lambda(z_i, \Omega_i(\tilde{s} - s_i^*)) + \eta_i + \tilde{H}(t_i). \tag{8}$$

If schooling imparts skill, employers beliefs about true productivity will become increasingly downward biased as they learn over time. More, if there is a reputational component, where employees who are discovered to have lied are fired, many with partially completed schooling who lied about having a degree will have lower earnings than expected, downward biasing estimates of the return to partially-completed college. Because researchers do not know who lied, and who omitted, empirical results will contain weighted averages of these effects.

6.3 Extensions: Sheepskin Effects and Returns to Credits

A related literature studies "sheepskin" effects, or the premium to earning a diploma over and above simply completing (almost) enough credits to graduate. Empirical work here largely revolves around comparing workers with degrees to those with many credits and no degree, or observing discontinuous breaks in returns to education as opposed to linear returns in years (e.g. Hungerford and Solon, 1987 and Jaeger and Page, 1996). Yet, if it is the case that many non-degree earners omit schooling from their resumes, wages in early careers cannot reflect human capital earned in college since employers are unaware of it unless employers are aware that some people omit and the starting wages for high school graduates refelects such information.

The same can be true for related work on returns to credits, which has a strong focus on two-year college students. In this case, a common empirical strategy is to estimate returns to credits using individual fixed effects with administrative educational data, often by broad field of study (Jacobson et al., 2005; Jepsen et al., 2014; Kane and Rouse, 1995). Since the econometrician observes schooling, this means the estimated return to credits is indeed accurate but part of the return includes selectively not revealing the schooling, which is a different interpretation than typically offered. Additionally, the general consensus is that returns are higher for credits in technical and STEM fields, and are low or non-existent in humanities and general studies. If it is the case that only some students inform employers of credits that do not result in a degree (non-completion rates in two-year schools are somewhere in the neighborhood of 70%), then estimates of returns right after school exit will reflect both returns for those who inform employers, and average wages among those who do not. It follows then that if reporting of credits to employers is selective, for example more common among four-year students than two, or more common among students who study technical fields than general, then lower returns among two-year students, or those in humanities courses, for example, will in part reflect employers' ignorance of that schooling.

6.4 Empirical Estimates on Non-Employment

Testing these predictions formally requires wages, in addition to our data, which we do not have. Further, we also lack panel data which would allow for interactions with experience. Lacking these, we test a weak version of these predictions using total non-employment taken from the resumes themselves.

In Table 6 we focus on omitting, where predictions are strongest, and show results from a regression of months of non-employment since college exit on whether a job-seeker omits.[27] Column 1 shows that after controlling for job-seekers' schooling in the high school sample, omitters have 1.4 more months of non-employment than non-omitters. Adding job-seeker race and PSAT score in column 2 do not change the estimate. The estimates are 2.1 additional months of non-employment in the unique names sample.

These results suggest that on average omitters see larger employment gaps. But, we cannot rule out unobservables that may be correlated with omitting and employment (e.g., honesty), nor can we rule out job-seekers only partially reporting employment. Additionally, we cannot know whether the job-seeker has always omitted his schooling, affecting past non-employment spells, or if this is a new occurrence. Thus we cannot rule out endogenous omitting in response to labor market experiences, but point out that observed differences are conditional on similar students with similar educational histories and test scores. A panel dataset with changing resumes is required to answer this more

[27] Results are qualitatively similar if we use the share of potential months of experience that person is non-employed.

fully and represents an avenue for future research.

In Table 7, we perform a similar exercise for lying. Here we create mutually exclusive categories of completed schooling, which can be true or not, and compare employment histories for job-seekers who lie and those who do not. Those with no lie on their resume could either be a high school graduate, attend a two or four year school and not earn a degree, or earn a degree from a two or four year school. This is $(Attend_{2,4}|True)$ and $(Degree_{2,4}|True)$ in Equation 9 below, where high school graduate is the omitted category. Those with lies could either lie about attending a two or four year school, or lie about earning a degree from a two or four year school. This is $(Attend_{2,4}|Lie)$ and $(Degree_{2,4}|Lie)$. Since the mutually exclusive categorization is on highest schooling listed, if a student attended a two year school and lied about completing, that individual's highest schooling is categorized as lying about a two year degree, and the individual is not categorized as having actually attended a two year school. The reference group are high school graduates who did not lie on their resumes. X_i includes race and PSAT score, and τ_t are high school cohort fixed effects. In this case we compare employment histories of degree earners and non-completers with those who lie, relative to high school graduates, in the model below:

$$
\begin{aligned}
NonEmp_i = &\alpha + \Pi_1(Attend_{2,4}|True) + \Pi_2(Degree_{2,4}|True) \\
&+ \Psi_1(Attend_{2,4}|Lie) + \Psi_2(Degree_{2,4}|Lie) + \Phi X_i + \tau_t + \varepsilon_i.
\end{aligned}
\tag{9}
$$

Results in Table 7 suggests that those who lie about attending a two-year school see fewer months of non-employment compared with truth-telling terminal high school graduates. Lying about attending a four-year school or earning a degree from a two-year school have similar signed coefficients, though estimates are sufficiently noisy to rule out zero or a positive relationship. Taken together with a positive relationship between omitting and non-employment suggests that employers may prefer candidates with schooling listed on resumes, potentially highlighting the role of employers' screening of resumes and job-seekers' responses.

7 Conclusion

After decades long efforts to encourage college enrollment, given high non-completion rates, it is natural to ask how the 31 million plus Americans who have enrolled and not completed college over the past 20 years view their education, and how their partially completed schooling is in turn viewed by the labor market.

We make marginal progress in understanding what job-seekers who did not graduate college signal to employers on their resumes, and what we can learn about returns to schooling from their behavior. We find that one-third of college non-completers in our sample of scraped resumes omit their one and only college experience, and that this is predicted by school type and quality, duration, and work experience, and not by race or ability. We also confirm what many already know – that people lie about schooling they did not complete.

Given that the typical U.S. college student does not earn a degree, we believe studying the return to partially completed schooling is an important endeavor. The fact that a non-trivial share of job-seekers would rather employers not know that they ever attended college is telling in itself. Moreover, if wages are in part determined by selective disclosure, then existing estimates may be wrong. Getting these estimates correct is important as we continue to push students into higher education with state and federal subsidies, not to mention costly loans. That students have the ability to omit schooling from their resume is an incentive to take on a risky endeavor. Under reasonable assumptions, having a completion rate less than 100 percent is a desirable general equilibrium outcome, as we should not only expect some riskiness in college investments, but should encourage it.

We also add to an emerging literature demonstrating the power of resumes. While some work already exists (Kuhn and Shen, 2015, 2016; Shen and Kuhn, 2013), to our knowledge we are the first to analyze a large set of resumes in the U.S. context, and the first to match resumes to administrative records. It is our hope that these exercises demonstrate the value of this type of data and its importance in creating a fuller picture of the interaction between workers and employers in the labor market. We acknowledge that in its infancy, this work comes with limitations.

In particular, we are limited by our ability to match resumes. While our sample is generally representative of early career job seekers, a fuller picture of who uses jobs boards, among both employers and job seekers, would be a great step in this research. Data sharing agreements with key matching terms may rectify this shortcoming. Similarly, the ability to match employment records with resumes could give a sense of the degree to which job seekers misrepresent employment histories, as they do education. Employers, who increasingly rely on screening algorithms from electronic applications, should be particularly interested in these findings. Observing wages and resumes would take this research even a step further, offering fresh perspectives on a host of well-known research questions.

References

Altonji, J. G. and Pierret, C. R. (2001). Employer learning and statistical discrimination. *The Quarterly Journal of Economics*, 116(1):313–350.

Araki, S., Kawaguchi, D., and Onozuka, Y. (2016). University prestige, performance evaluation, and promotion: Estimating the employer learning model using personnel datasets. *Labour Economics*, 41:135–148.

Arcidiacono, P., Bayer, P., and Hizmo, A. (2010). Beyond signaling and human capital: Education and the revelation of ability. *American Economic Journal: Applied Economics*, pages 76–104.

Babcock, P. (2003). Spotting lies reference checks alone won't protect you from a mendacious job applicant. *HR MAGAZINE*, 48(10):46–53.

Bauer, T. K. and Haisken-DeNew, J. P. (2001). Employer learning and the returns to schooling. *Labour Economics*, 8(2):161–180.

Bertrand, M. and Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *The American Economic Review*, 94(4):991–1013.

Black, D., Sanders, S., and Taylor, L. (2003). Measurement of higher education in the census and current population survey. *Journal of the American Statistical Association*, 98(463):545–554.

Cellini, S. R. and Turner, N. (2016). Gainfully employed? assessing the employment and earnings of for-profit college students using administrative data. *National Bureau of Economic Research*.

Clemens, J., Khan, L., and Meer, J. (2018). Dropouts need not apply: The minimum wage and skill upgrading.

Deming, D. and Kahn, L. B. (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, 36(S1):S337–S369.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4):1593–1640.

Deming, D. J., Yuchtman, N., Abulafi, A., Goldin, C., and Katz, L. F. (2016). The value of postsecondary credentials in the labor market: An experimental study. *American Economic Review*, 106(3):778–806.

Dynarski, S., Jacob, B., and Kreisman, D. (2017). How important are fixed effects and time trends in estimating returns to schooling? evidence from a replication of jacobson, lalonde, and sullivan, 2005. *Journal of Applied Econometrics*.

Dynarski, S. M., Hemelt, S. W., and Hyman, J. M. (2015). The missing manual: Using national student clearinghouse data to track postsecondary outcomes. *Educational Evaluation and Policy Analysis*, 37(1_suppl):53S–79S.

Farber, H. S. and Gibbons, R. (1996). Learning and wage dynamics. *The Quarterly Journal of Economics*, 111(4):1007–1047.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, 24(3):411–482.

Helleseter, M. D., Kuhn, P., and Shen, K. (2018). The age twist in employers' gender requests: Evidence from four job boards. *Journal of Human Resources*, pages 0416–7836R2.

Hershbein, B. and Kahn, L. B. (2018). Do recessions accelerate routine-biased technological change? evidence from vacancy postings. *American Economic Review*, 108(7):1737–72.

Hungerford, T. and Solon, G. (1987). Sheepskin effects in the returns to education. *The review of economics and statistics*, pages 175–177.

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

Jaeger, D. A. and Page, M. E. (1996). Degrees matter: New evidence on sheepskin effects in the returns to education. *The review of economics and statistics*, pages 733–740.

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

Kahn, L. B. (2013). Asymmetric information between employers. *American Economic Journal: Applied Economics*, 5(4):165–205.

Kahn, L. B. and Lange, F. (2014). Employer learning, productivity, and the earnings distribution: Evidence from performance measures. *The Review of Economic Studies*, 81(4):1575–1613.

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

Kane, T. J., Rouse, C. E., and Staiger, D. (1999). Estimating returns to schooling when schooling is misreported. *National Bureau of Economic Research working paper #w7235*.

Kroft, K., Lange, F., and Notowidigdo, M. J. (2013). Duration dependence and labor market conditions: Evidence from a field experiment. *The Quarterly Journal of Economics*, 128(3):1123–1167.

Kuhn, P. and Shen, K. (2012). Gender discrimination in job ads: Evidence from china. *The Quarterly Journal of Economics*, 128(1):287–336.

Kuhn, P. and Shen, K. (2015). Do employers prefer migrant workers? evidence from a chinese job board. *IZA Journal of Labor Economics*, 4(1):22.

Kuhn, P. and Shen, K. (2016). Gender-targeted job ads in the recruitment process: Evidence from china. *Working paper.*

Lange, F. (2007). The speed of employer learning. *Journal of Labor Economics*, 25(1):1–35.

Light, A. and McGee, A. (2015a). Does employer learning vary by schooling attainment? the answer depends on how career start dates are defined. *Labour Economics*, 32:57–66.

Light, A. and McGee, A. (2015b). Employer learning and the "importance" of skills. *Journal of Human Resources*, 50(1):72–107.

Mansour, H. (2012). Does employer learning vary by occupation? *Journal of Labor Economics*, 30(2):415–444.

Marinescu, I. (2017). The general equilibrium impacts of unemployment insurance: Evidence from a large online job board. *Journal of Public Economics*, 150:14–29.

Oettinger, G. S. (1996). Statistical discrimination and the early career evolution of the black-white wage gap. *Journal of Labor Economics*, 14(1):52–78.

Schönberg, U. (2007). Testing for asymmetric employer learning. *Journal of Labor Economics*, 25(4):651–691.

Schubert, G., Stansbury, A., and Taska, B. (2019). Mitigating monopsony: Occupational mobility and outside options. *Working Paper*.

Shapiro, D., Dundar, A., Yuan, X., Harrell, A. T., Wild, J. C., and Ziskin, M. B. (2014). Some college, no degree: A national view of students with some college enrollment, but no completion (signature report no. 7). *National Student Clearinghouse Research Center*.

Shen, K. and Kuhn, P. (2013). Do chinese employers avoid hiring overqualified workers? evidence from an internet job board. In *Labor Market Issues in China*, pages 1–30. Emerald Group Publishing Limited.

Smith, J. and Stange, K. (2016). A new measure of college quality to study the effects of college sector and peers on degree attainment. *Education Finance and Policy*, 11(4):369–403.

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

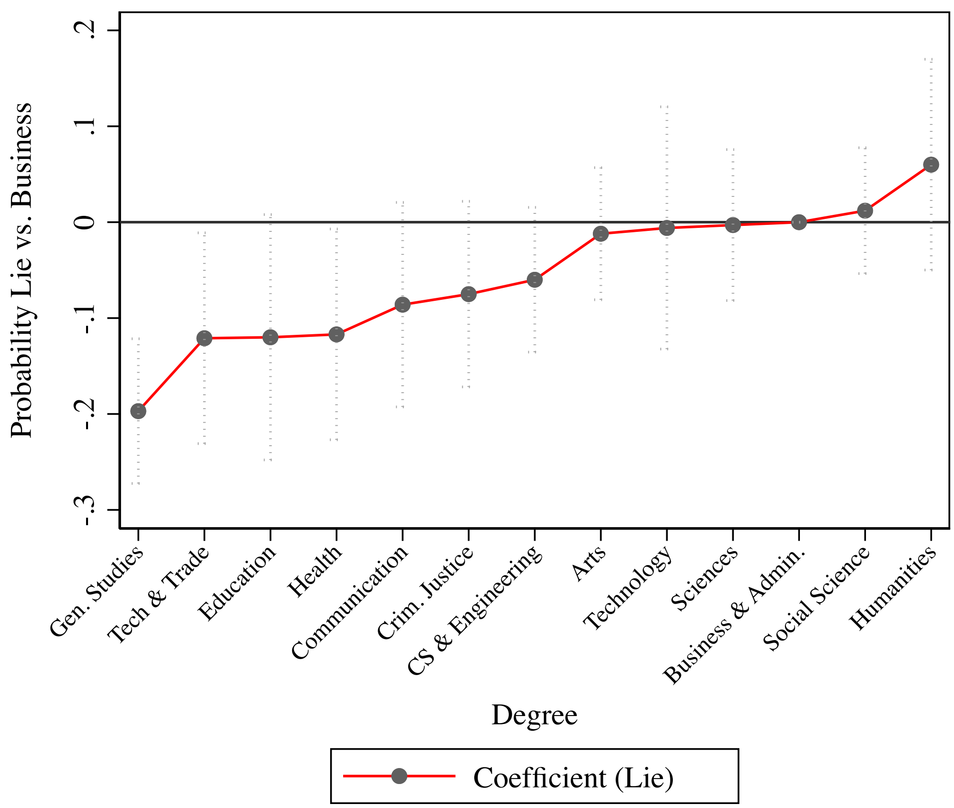
Stiglitz, J. E. (1975). The theory of "screening," education, and the distribution of income. *The American Economic Review*, 65(3):283–300.

Wood, J. L., Schmidtke, J. M., and Decker, D. L. (2007). Lying on job applications: The effects of job relevance, commission, and human resource management experience. *Journal of Business and Psychology*, 22(1):1–9.

Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32(4):711–754.

Tables and Figures

Figure 1: Conditional Probability of Lying (reference group is Business and Admin.)



Notes: Figure plots coefficients and 90% confidence intervals from column 5 of Table 5. Coefficients are likelihood a degree is a lie by field relative to Business, conditional on resume and employee characteristics. Sample includes resumes matched to administrative educational data for male job-seekers who listed no more than one college on their resume and graduated high school between 2004 and 2011. The match is on name and high school or college attended and date of attendance. Field of study was coded by hand from resumes and are not mutually exclusive.

Table 1: Matching statistics.

	High School Sample					
	1. All Resumes	2. No employment Pre-2001	3. Listed HS & Enroll >2003	4. Matched to Truth	5. Attend 1 College	6. Dropped Out Final Sample
Schools listed	1.39	1.36	0.87	1	1.06	1.02
Jobs listed	4.56	4.15	4.1	4.13	4.1	4.02
Years worked	10.85	6.55	5.41	5.74	5.48	5.32
Months not working (2014)	2.4	2.49	2.88	2.56	2.58	2.68
High school year			2009.9	2009.4	2010.0	2010.3
Obs.	556,651	382,953	33,517	11,244	4,516	3,887
% of obs from previous column		0.69	0.09	0.34	0.40	0.86

	Unique Names Sample					
	1. All Resumes	2. Unique Name On Resume	3. Matched & to Truth	4. Same State & No pre-16 work	5. Attend 1 College	6. Dropped Out Final Sample
Schools institution listed	1.39	1.38	1.09	1.03	0.643	0.13
Jobs listed	4.56	4.21	4.18	3.97	3.97	3.74
Years worked	10.85	7.67	6.69	5.00	4.85	4.51
Months not working (2014)	2.4	2.54	2.45	2.52	2.52	2.91
Obs.	556,651	210,038	36,300	13,895	7,038	4,384
% of obs from previous column		0.38	0.17	0.38	0.51	0.62

Notes: Resumes are from a sample of males posting to an online jobs board in fall of 2016 and spring of 2017 from the 100 largest U.S. cities. The top half of the table is the sample that matches on name and high school and the bottom half matches only on unique names. Column 1 is the full sample of scraped resumes and each successive column is a subsample of the previous column, resulting in the last column, which is the final sample. Jobs listed, years worked, and months not working (2014) are all values conditional on listing any jobs (for jobs listed), and working any years (for years worked or months not employed).

Table 2: Summary statistics, by omit status.

	High School Sample			Unique Names Sample		
	All	Non-Omitters	Omitters	All	Non-Omitters	Omitters
College Characteristics						
Attend Four-Year	0.50	0.57	0.36	0.62	0.72	0.40
School Avg. PSAT	88.9	90.8	85.1	91.2	93.8	85.7
Years Enrolled	0.84	1.02	0.50	1.14	1.39	0.64
Potential Experience	4.08	3.68	4.90	4.02	3.46	5.19
Other Training	0.27	0.20	0.42	0.30	0.24	0.43
Job-seeker characteristics						
White	0.35	0.38	0.31	0.38	0.42	0.29
Black	0.32	0.31	0.36	0.31	0.28	0.37
Hispanic	0.20	0.19	0.22	0.14	0.12	0.19
Asian	0.06	0.06	0.04	0.09	0.11	0.07
Other Race	0.07	0.07	0.07	0.07	0.07	0.08
Student PSAT	86.0	88.7	80.4	89.4	93.1	81.2
Employment						
Months Not Working	8.38	7.23	10.42	8.63	6.73	12.36
Omit College	0.33	-	1.00	0.32	-	1.00
Observations	3,887	2,596	1,291	4,384	2,977	1,407

Notes: Sample includes resumes matched to administrative educational data for male job-seekers who graduated high school between 2004 and 2014, only attended one college according to administrative records, and did not complete a degree. The high school sample includes job-seekers who listed a high school on their resume and the unique names sample includes only those within unique first-last name combinations in both initial datasets. Omit is true if a job-seeker left a college experience off a resume. Other training is a non-administrative records post-high school educational entry on a resume. School PSAT is school average PSAT score. Months not working is the number of months not employed from resume job listings since exiting college.

Table 3: The likelihood of omitting schooling

	High School Sample			Unique Names Sample		
Years enrolled	-0.170***	-0.169***	-0.171***	-0.139***	-0.139***	-0.139***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Potential experience	0.020***	0.020***	0.020***	0.030***	0.030***	0.030***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Attend four-year	-0.015	-0.015	-0.012	-0.087***	-0.086***	-0.083***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
School PSAT	-0.004***	-0.004***	-0.004***	-0.003***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Other training	0.182***	0.182***	0.182***	0.110***	0.109***	0.110***
	(0.017)	(0.017)	(0.017)	(0.015)	(0.015)	(0.015)
PSAT		0.000	-0.000		-0.000	-0.000
		(0.000)	(0.000)		(0.000)	(0.000)
Black			-0.020			-0.015
			(0.019)			(0.018)
Hispanic			-0.014			0.053**
			(0.021)			(0.021)
Asian			0.000			0.020
			(0.029)			(0.020)
Other race			-0.021			-0.007
			(0.029)			(0.026)
Mean Dep. Var.	0.33	0.33	0.33	0.32	0.32	0.32
Obs.	3,887	3,887	3,887	4,384	4,384	4,384
R^2	0.181	0.181	0.181	0.231	0.232	0.235

Notes: Results are from a linear probability model. The dependent variable is a binary indicator for omitting college from a resume. Sample includes resumes matched to administrative data for male job-seekers who graduated high school between 2004 and 2014, only attended one college in administrative records, and did not complete a degree. Other training is a non-NSC post-high school educational entry on a resume. School PSAT is school average PSAT score.
(* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Table 4: Summary Statistics, Lying Sample.

	(1) All	(2) No Lie	(3) Lie
Any Lie	0.20	0.00	1.00
Lie About Degree	0.16	0.00	0.79
Lie about Attending	0.07	0.00	0.34
Demographics			
White	0.36	0.36	0.36
Black	0.35	0.35	0.34
Hispanic	0.18	0.18	0.18
Asian	0.05	0.05	0.05
Other Race	0.07	0.06	0.07
Student PSAT	81.87	81.15	84.74
Resume Attributes			
Enrolled in 4-Year College	0.33	0.27	0.56
School Avg. PSAT	90.51	91.30	89.07
Earned Degree	0.29	0.16	0.79
Months Not Working	15.57	16.02	13.79
Share Months Not Working	0.24	0.24	0.22
Resume field of study			
Business	0.11	0.07	0.23
Education	0.01	0.01	0.02
Humanities	0.02	0.02	0.04
Social Science	0.07	0.06	0.13
Engineering/Computer Science	0.04	0.03	0.08
Science	0.05	0.04	0.08
Arts	0.06	0.04	0.13
Technology (not Comp Science)	0.02	0.01	0.03
Technical/Trade	0.02	0.01	0.04
Health	0.02	0.02	0.03
Communications	0.02	0.02	0.03
Criminal Justice	0.02	0.01	0.05
General Studies	0.03	0.03	0.06
N	4154	3309	845

Notes: Sample includes resumes matched to administrative educational data for male job-seekers who listed no more than one college on their resume and graduated high school between 2004 and 2011. The match is on name and high school or college attended and date of attendance. Any lie is set to 1 if the resume contains either a lie about attending, about a degree, or both. Resume attributes are schooling listed on a resume, even if untrue. Field of study is hand coded from resumes and may or may not be true.

Table 5: Lying About Degree Completion by Field of Study

	Outcome = Lied About Degree				
	(1)	(2)	(3)	(4)	(5)
Business			Reference		
Education	-0.096	-0.099	-0.094	-0.092	-0.101
	(0.075)	(0.073)	(0.072)	(0.072)	(0.073)
Technical/Trade	-0.085	-0.121*	-0.124*	-0.131*	-0.130*
	(0.066)	(0.067)	(0.067)	(0.067)	(0.067)
Health	-0.107	-0.118*	-0.116*	-0.119*	-0.116*
	(0.066)	(0.064)	(0.064)	(0.064)	(0.063)
General Studies	-0.188***	-0.212***	-0.205***	-0.205***	-0.207***
	(0.044)	(0.046)	(0.046)	(0.046)	(0.046)
Communications	-0.094	-0.076	-0.074	-0.075	-0.080
	(0.065)	(0.065)	(0.065)	(0.065)	(0.065)
Humanities	-0.016	0.039	0.048	0.047	0.042
	(0.068)	(0.069)	(0.069)	(0.070)	(0.067)
Social Science	-0.021	0.018	0.024	0.024	0.022
	(0.038)	(0.039)	(0.039)	(0.039)	(0.039)
Engineering/Computer Science	-0.069	-0.064	-0.053	-0.056	-0.057
	(0.045)	(0.045)	(0.045)	(0.046)	(0.046)
Science	-0.062	-0.025	-0.014	-0.019	-0.019
	(0.047)	(0.047)	(0.047)	(0.048)	(0.048)
Arts	-0.017	-0.017	-0.016	-0.017	-0.018
	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)
Technology (not Comp Science)	0.010	-0.016	-0.010	-0.008	-0.013
	(0.073)	(0.074)	(0.075)	(0.075)	(0.075)
Criminal Justice	-0.063	-0.075	-0.082	-0.082	-0.088
	(0.059)	(0.059)	(0.058)	(0.059)	(0.059)
4-year school on resume		-0.009	-0.009	-0.007	-0.016
		(0.031)	(0.030)	(0.031)	(0.042)
PSAT of college on resume		-0.004***	-0.003**	-0.003**	0.001
		(0.001)	(0.001)	(0.001)	(0.002)
Student PSAT			-0.001	-0.001	-0.001
			(0.001)	(0.001)	(0.001)
PSAT of actual college attended					-0.005**
					(0.002)
Race				×	×
PSAT				×	×
Cohort	×	×	×	×	×
N	1,643	1,643	1,643	1,643	1,643
R^2	0.036	0.051	0.054	0.056	0.061

Notes: Sample includes resumes matched to administrative educational data for male job-seekers who listed no more than one college on their resume and graduated high school between 2004 and 2011. Match is on exact name and high school or college attended and date of attendance. Field of study was coded by hand from resumes and are not mutually exclusive. Four-year school on resume, and PSAT of college on resume, indicate if the college listed was a four-year, and that school's average PSAT. These may be true or not. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Table 6: Omitting schooling and months of non-employment on resumes

	High School Sample		Unique Names Sample	
	(1)	(2)	(3)	(4)
Omit	1.374**	1.447**	2.133***	2.121***
	(0.571)	(0.575)	(0.586)	(0.582)
Years enrolled	-2.001***	-1.911***	-1.875***	-1.927***
	(0.350)	(0.353)	(0.377)	(0.377)
Attend four-year	1.628***	1.164**	0.065	-0.008
	(0.575)	(0.587)	(0.646)	(0.647)
School PSAT	0.011	0.049	0.065***	0.074***
	(0.027)	(0.032)	(0.024)	(0.028)
Other training	0.718	0.769	0.469	0.529
	(0.575)	(0.575)	(0.517)	(0.517)
Student PSAT		0.001		0.018
		(0.016)		(0.015)
Black		2.456***		1.771***
		(0.640)		(0.556)
Hispanic		0.000		0.838
		(0.635)		(0.717)
Asian		0.795		1.156
		(1.154)		(0.734)
Other race		0.143		0.933
		(0.997)		(0.964)
Cohort FE	X	X	X	X
Observations	2,845	2,845	3,431	3,431
R^2	0.145	0.153	0.204	0.207

Notes: Dependent variable is months of non-employment on the resume, beginning with exit from college. Sample includes scraped resumes matched to administrative educational data for job-seekers who only attended one college, listed their high school on their resume, and graduated high school between 2004 and 2014. Resume with no employment and/or dates are excluded. Omit equals one if job-seeker i omitted his college experience from his resume. Other training is a non-NSC post-high school educational entry on a resume. School PSAT is school average PSAT score. Years enrolled is set equal to zero if a job-seeker earned a degree. These are centered to the mean of the regression sample in column 3 and are interacted with a binary indicator for Omit. Cohort FE are high school cohort.
(* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Table 7: Non-employment (in months) and lying about schooling on resumes

	(1)	(2)
Highest schooling listed is true		
High school	Reference	Reference
Attended 2	-9.178***	-8.813***
	(0.911)	(0.907)
Attended 4	-6.861***	-6.564***
	(0.929)	(0.956)
Degree 2	-10.081***	-9.099***
	(1.794)	(1.807)
Degree 4	-11.489***	-10.166***
	(1.027)	(1.133)
Highest schooling listed is a lie		
Lie, Attended 2	-3.382***	-3.358***
	(1.166)	(1.158)
Lie, Attended 4	-1.660	-1.434
	(1.072)	(1.080)
Lie Degree 2	-4.333	-4.228
	(4.289)	(4.272)
Lie Degree 4	3.427	3.151
	(5.286)	(5.209)
Race, PSAT		×
Cohort FE	×	×
N	3,052	3,052
R^2	0.146	0.155

Notes: Sample includes resumes matched to administrative educational data for male job-seekers who listed no more than one college on their resume and graduated high school between 2004 and 2011. The match is on name and high school or college attended and date of attendance. Attend 2/4 indicates if resume truthfully indicates individual's highest schooling is attending, but not graduating from a 2 or 4 years school, and did not lie about a degree. Degree 2/4 is same for degree. Lie attend 2/4 indicates if individual's resume has a 2 or 4 year schooling listed that he did not attend. Degree is if individual put a degree he did not earn. All categories are mutually exclusive. Dependent variable is months of non-employment on the resume, beginning with exit from college.
(* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Appendix

Table A1: Sample Comparison to Administrative Educational Records.

	Administrative Records		High School Sample		Unique Names Sample	
	Mean	(s.d.)	Mean	(s.d.)	Mean	(s.d.)
White	0.53	(0.50)	0.35	(0.48)	0.38	(0.48)
Black	0.15	(0.35)	0.32	(0.47)	0.31	(0.46)
Hispanic	0.17	(0.38)	0.20	(0.40)	0.14	(0.35)
Asian	0.08	(0.28)	0.06	(0.23)	0.09	(0.29)
Other Race	0.06	(0.24)	0.07	(0.25)	0.07	(0.26)
Student PSAT	93.99	(21.51)	85.97	(20.48)	89.37	(21.75)
College Avg. PSAT	92.68	(12.61)	88.88	(11.26)	91.23	(11.99)
Attend 4-Year College	0.61	(0.49)	0.50	(0.50)	0.62	(0.49)
Observations	3,810,389		3,887		4,384	

Notes: Administrative records are all males who took the any of the PSAT, SAT, or AP in high school graduating cohorts of 2004-2014, live in an MSA, and attended one college but did not earn a degree. The high school and unique names samples are the subset or resumes matched to administrative records.

Table A2: Likelihood of omitting schooling robustness tests - High school sample

	(1) Main result	(2) Left HS pre-2012	(3) Only 2-yr college	(4) Only 4-yr college	(5) Enrolled ≥1 yr	(6) Enrolled ≥0.5 yr	(7) No other training
Years enrolled	-0.171***	-0.191***	-0.282***	-0.131***	-0.067***	-0.095***	-0.159***
	(0.008)	(0.012)	(0.019)	(0.009)	(0.015)	(0.009)	(0.009)
Potential experience	0.020***	0.013***	0.020***	0.020***	0.014***	0.023***	0.021***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.003)
Attend four-year	-0.012	0.007			-0.025	-0.006	-0.026
	(0.018)	(0.023)			(0.031)	(0.022)	(0.021)
School PSAT	-0.004***	-0.005***	-0.002	-0.005***	-0.001	-0.004***	-0.003***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Other training	0.182***	0.176***	0.192***	0.176***	0.144***	0.169***	
	(0.017)	(0.021)	(0.024)	(0.024)	(0.031)	(0.021)	
Student PSAT	-0.000***	-0.326**	-0.043	-0.249*	-0.036	-0.194*	-0.160
	(0.000)	(0.130)	(0.208)	(0.128)	(0.161)	(0.117)	(0.106)
Black	-0.020	-0.027	-0.006	-0.032	-0.007	-0.006	-0.021
	(0.019)	(0.024)	(0.030)	(0.024)	(0.025)	(0.021)	(0.021)
Hispanic	-0.014	-0.021	-0.013	-0.003	0.058*	0.026	-0.035
	(0.021)	(0.027)	(0.030)	(0.030)	(0.032)	(0.025)	(0.024)
Asian	0.000	0.047	0.020	-0.003	0.023	0.013	-0.009
	(0.029)	(0.043)	(0.054)	(0.032)	(0.037)	(0.029)	(0.030)
Other race	-0.021	-0.043	-0.012	-0.024	0.007	0.000	-0.031
	(0.029)	(0.037)	(0.043)	(0.039)	(0.043)	(0.034)	(0.034)
Obs.	3,887	2,354	1,940	1,947	1,084	2,337	2,819
R^2	0.181	0.212	0.128	0.198	0.099	0.135	0.141

Notes: Column 1 replicates column 3 of Table 3, the main regression in the high school sample. Column 2 limits to pre-2012 cohorts. Columns 3 and 4 are limited to two- or four-year attendees. Columns 5 and 6 are limited to those who enrolled for more than 1 or one-half year in college. Columns 7 and 8 limit to those who never graduated and those with no other non-NSC schooling listed. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Table A3: Likelihood of omitting schooling robustness tests - Unique names sample.

	(1) Main result	(2) Left HS pre-2012	(3) Only 2-yr college	(4) Only 4-yr college	(5) Enrolled ≥1 yr	(6) Enrolled ≥0.5 yr	(7) No other training
Years enrolled	-0.139***	-0.150***	-0.242***	-0.112***	-0.028*	-0.077***	-0.154***
	(0.008)	(0.011)	(0.023)	(0.009)	(0.015)	(0.009)	(0.009)
Potential experience	0.030***	0.028***	0.025***	0.035***	0.028***	0.031***	0.027***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.003)
Attend four-year	-0.083***	-0.070***			-0.059*	-0.080***	-0.101***
	(0.018)	(0.023)			(0.032)	(0.023)	(0.022)
School PSAT	-0.003***	-0.003***	0.000	-0.003***	-0.003***	-0.002**	-0.002*
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Other training	0.110***	0.116***	0.096***	0.138***	0.196***	0.146***	
	(0.015)	(0.018)	(0.025)	(0.019)	(0.023)	(0.018)	
Student PSAT	-0.000	-0.000	-0.000	-0.000	0.000	-0.001*	-0.001
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Black	-0.015	-0.019	-0.037	0.001	-0.024	-0.019	-0.003
	(0.018)	(0.021)	(0.033)	(0.020)	(0.021)	(0.019)	(0.020)
Hispanic	0.053**	0.048*	0.060*	0.041*	0.066**	0.077***	0.018
	(0.021)	(0.025)	(0.036)	(0.025)	(0.026)	(0.024)	(0.024)
Asian	0.020	0.035	0.035	0.019	0.059***	0.056***	-0.026
	(0.020)	(0.024)	(0.052)	(0.020)	(0.021)	(0.020)	(0.020)
Other race	-0.007	-0.013	-0.076*	0.038	0.019	0.007	-0.009
	(0.026)	(0.032)	(0.046)	(0.032)	(0.034)	(0.030)	(0.031)
Observations	4,384	2,973	1,679	2,705	1,927	3,002	3,073
R-squared	0.235	0.283	0.095	0.225	0.183	0.181	0.247

Notes: Column 1 replicates column 6 of Table 3, the main regression in the unique names sample. Column 2 limits to pre-2012 cohorts. Columns 3 and 4 are limited to two- or four-year attendees. Columns 5 and 6 are limited to those who enrolled for more than 1 or one-half year in college. Columns 7 and 8 limit to those who never graduated and those with no other non-NSC schooling listed. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Table A4: The likelihood of omitting schooling - Non-linear potential experience.

	High School Sample			Unique Names Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
1-3 years potential experience	-0.006			0.039		
	(0.026)			(0.027)		
3-5 years potential experience	0.086***			0.119***		
	(0.029)			(0.029)		
5+ years potential experience	0.133***			0.216***		
	(0.029)			(0.029)		
< 3 years potential experience		-0.117***	-0.136		-0.136***	-0.078
		(0.015)	(0.086)		(0.014)	(0.079)
<3 years * two-year college			0.059***			0.064***
			(0.017)			(0.016)
<3 years * four-year college			-0.027			-0.067*
			(0.034)			(0.036)
<3 years * school PSAT			-0.000			-0.001
			(0.001)			(0.001)
< 3 year * other training			0.009			-0.034
			(0.034)			(0.030)
Years enrolled	-0.174***	-0.176***	-0.201***	-0.145***	-0.148***	-0.177***
	(0.009)	(0.008)	(0.012)	(0.008)	(0.008)	(0.012)
Attend four-year	-0.010	-0.009	0.004	-0.080***	-0.084***	-0.056**
	(0.018)	(0.018)	(0.023)	(0.018)	(0.018)	(0.024)
College PSAT	-0.004***	-0.004***	-0.004***	-0.003***	-0.003***	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Other training	0.185***	0.188***	0.184***	0.113***	0.121***	0.136***
	(0.017)	(0.017)	(0.022)	(0.015)	(0.015)	(0.021)
School PSAT	-0.238**	-0.251***	-0.255***	-0.209**	-0.233***	-0.218**
	(0.096)	(0.096)	(0.097)	(0.086)	(0.086)	(0.088)
Student PSAT	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Black	-0.019	-0.021	-0.021	-0.014	-0.014	-0.011
	(0.019)	(0.019)	(0.019)	(0.018)	(0.018)	(0.018)
Hispanic	-0.012	-0.015	-0.015	0.050**	0.047**	0.048**
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Asian	0.003	0.000	0.003	0.018	0.012	0.016
	(0.029)	(0.029)	(0.029)	(0.021)	(0.021)	(0.021)
Other race	-0.019	-0.022	-0.023	-0.008	-0.010	-0.012
	(0.029)	(0.029)	(0.030)	(0.027)	(0.027)	(0.027)
Obs.	3,887	3,887	3,887	4,384	4,384	4,384
R^2	0.183	0.181	0.183	0.232	0.227	0.230

Notes: Results are from a linear probability model. The dependent variable is a binary indicator for omitting college from a resume. Sample includes resumes matched to administrative data for male job-seekers who graduated high school between 2004 and 2014, only attended one college in administrative records, and did not complete a degree. Other training is a non-NSC post-high school educational entry on a resume. School PSAT is school average PSAT score.
(* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)