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

Daniel Kreisman*
Georgia State University

Jonathan Smith
Georgia State University

Bondi Arifin
Ministry of Finance,
Republic of Indonesia

March, 2020

Abstract

What do college non-completers put 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 schooling records. We find that one in three strategically omit their only post-secondary schooling. We further show that these are not casual omissions but are strategic decisions systematically related to schooling characteristics, in line with straightforward theoretical predictions. We also find evidence of lying, and show which degrees are most likely untrue. Lastly, we estimate the impacts of omitting and lying on employment. Our results provide evidence of job-seekers' valuation of non-completed degrees, and demonstrates the assumption that employers perfectly observe schooling does not always hold.

Keywords: Signaling; Resume; Employer Learning; Statistical Discrimination; Jobs Board.
JEL: J01, J24

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

*Corresponding author: Kreisman at dkreisman@gsu.edu.

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 simply 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 education. 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 the literature, 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.

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 eliminate ambiguity about transfers and multiple enrollments, we focus on (male) job-seekers who attended one college, though this can be relaxed.[2]

We find that one out of every three non-completers in our matched sample of early career

[1] Shapiro et al. (2014).

[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.

job-seekers omit their only college experience from their resume. We also find that omission is systematically related to schooling characteristics, in line with straightforward theoretical predictions. In particular, students who enroll for fewer than two years, and those who attend lower quality 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.

We note that job-seekers can lie as well. We find that over 20 percent in our sample 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 or humanities. Like strategic omissions, this provides suggestive evidence of which types of schooling impart human capital, and which might be purely 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 low 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, and a literature on increasing degree requirements in response to minimum wage increases (Clemens et al., 2018), increasing skill requirements in response to macroeconomic shocks and technological advancements in the economy (Hershbein and Kahn, 2018), or 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 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 (Jaeger and Page, 1996; Jepsen et al., 2014; Kane and Rouse, 1995). In each of these and other related literatures,

[3] Araki et al. (2016); Arcidiacono et al. (2010); Bauer and Haisken-DeNew (2001); Galindo-Rueda (2003); Kahn (2013); Kahn and Lange (2014); Lange (2007); Light and McGee (2015a,b); Mansour (2012); Oettinger (1996); Schönberg (2007).

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. Further, we show that by extending a standard employer learning and statistical discrimination model to allow for strategic omissions, we can generate a test of the signaling versus human capital value of schooling. While we do not have the data to carry out an exact version of this hypothetical test (i.e. wages), we use a reduced-form estimate of whether omitting is related to non-employment as a proxy. We find that job-seekers who omit schooling also list more months of non-employment than similar job-seekers who do not omit similar schooling experiences.

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; 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.

2 Background - Schooling, Wages, and Resumes

There are several related strands of literature considering how schooling determines wages, particularly in early careers. In each of these, when workers are hired, employers (and sometimes workers) have only limited information about their true productivity. At hire, workers are then paid according to their expected marginal product, and as true productivity is revealed, wages converge on the value of their output. A key feature in the study of this phenomenon is that the econometrician and the employer observe different information.

Seminal work by Farber and Gibbons (1996) and Altonji and Pierret (2001) establish these facts by distinguishing between characteristics observable to employers at time of hire, and those available to the researcher. This distinction allows them to test for what Altonji and Pierret call Employer Learning with Statistical Discrimination. These authors, and the many that follow suit (Araki et al., 2016; Arcidiacono et al., 2010; Bauer and Haisken-DeNew, 2001; Galindo-Rueda, 2003; Kahn, 2013; Kahn and Lange, 2014; Lange, 2007; Light and McGee, 2015a,b; Mansour, 2012; Oettinger, 1996; Schönberg, 2007), assume that the econometrician has novel information that is indicative of productivity – for example test scores – while the employer only observes characteristics such as schooling at time of hire. The model then predicts that as workers gain experience in the labor market, and employers learn about their true ability, the econometrician should observe wages becoming increasingly correlated with test scores over time, and increasingly less related to schooling.

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. On average, this would bias upward estimates of sheepskin effects among young workers. In fact, even if no sheepskin effect exists – that is, if employers do not value a degree beyond the sum of its credits – but non-completers omit schooling from their resumes, then empirical estimates would find evidence of the sheepskin effect simply as a result of employers not observing schooling for non-completers while the econometrician does.

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, often by broad field of study (Dynarski et al., 2018; Jacobson et al., 2005; Jepsen et al., 2014; Kane and Rouse, 1995). 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. But, 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.

It turns out that by testing, and confirming, our hypothesis that many non-completers omit schooling, we are in fact conducting a test of the signaling versus human capital value of partially completed schooling. The intuition for which is simply an extension of the employer learning and statistical discrimination framework, and the related literatures described above. If some college non-completers omit their college attendance from their resumes, then this portion of their schooling is observable only to the econometrician, and not to the employer. In turn, if this schooling imparted human capital, then as employers learn, the econometrician should observe a stronger relationship between schooling and wages over time for omitters than non-omitters. If in fact these educational experiences imparted little or no human capital, they should not register in the researcher's wage equation. The same intuition holds for lying about schooling, with opposite signed implications. While we cannot test this explicitly as we do not have a panel of wages, we do observe employment and can compare whether omitting, or lying about, schooling hurts or helps workers on average.

3 Data

3.1 Resumes

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 (the 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 (and researchers), 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] We scrape only males 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 files. The reason for this is that women are more likely to change their surnames in marriage, which would make matching to college records difficult and limit us to females who never changed their name. Because job-seekers enter all information into uniform fields rather than uploading created resumes, we are able to use the web site's metatags 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

[4] For example, if 77449 (in 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.

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.

The summary statistics for the 556,651 resumes are in column 1 of Table 1. On average, job-seekers listed 4.6 different jobs, had an average of 10.85 years of work experience (for those listing experience), and listed 1.4 educational institution experiences, most of which are college. To take a representative year, we found that these current job-seekers did not list an employment spell for an average of 2.4 months in 2014.

3.2 The Truth from Administrative Educational Records

We match these resumes to 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 ability.[5]

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. The most notable deficiency is for-profit college enrollment, though many of the largest for-profits are included.[6] The data track all spells of enrollment at participating colleges, 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.[7]

3.3 Matching Resumes to the Truth

We match resumes to administrative College Board 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

[5] 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.

[6] 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.

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

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 less than 1 percent. 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. As an alternative approach to relying on high schools for matching, we also match on names that are unique among both the 12 million males in our administrative records and over 550,000 resumes. While we cannot determine whether these names are truly unique and hence the same individual, this secondary matching algorithm allows us test whether our results are driven by matching on resumes that list a high school.

The remaining columns of Table 1 show resume characteristics once we apply a sequential set of additional resume sample restrictions. Our administrative records include only students who graduated high school after 2003. In column 2 we exclude the roughly one-in-three resumes that list employment prior to 2001 (approximately before age 15 for the oldest cohort in our sample). Mechanically, this lowers the average years of worked but the other variables are roughly unchanged. Column 3 then limits to the 11 percent of resumes from column 2 that list both a high school and a high school graduation year.[8] With this information listed on resumes we can then limit to those who attended high school after to 2003 and can be matched to our administrative records in column 4. From this sample of 33,517 eligible resumes, we match 34 percent (11,244) to administrative records. Non-matches can result either from differential name spellings, graduating from a different high school than where they took a high school exam, or job-seekers simply not taking the PSAT, SAT, or and AP exam. Match rates are higher if we limit to states requiring the SAT.

Column 5 shows the 4,506 job-seekers in our main analytic sample for omitting. The sample restricts the matched sample in the previous column to those who only have one NSC college, even if they listed more or less on their resume. We focus 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 for the majority of the attrition between the fifth and sixth columns. Second, this limitation simplifies our thought experiment. It implies we will be observing job-seekers who omit the one and only NSC college they attended, and further avoids ambiguity for students who transfer schools.[9]

In Appendix Table A1 we compare our primary sample to male respondents in the American Community Survey (ACS) who are not in school, live in an MSA, had looked for work in the past year and attended at least some college.[10] We find that the ACS and our final matched sample

[8]This value is relatively low because we cannot filter our resumes of job-seekers too old to be in College Board records.

[9]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.

[10]To create a comparison, we take the 2016 ACS 5-year sample. From all individuals in the survey, we limit to those who are males, not currently enrolled in school, and would have been 18 between 2004 and 2014, to match the sample frame for our College Board data. We then limit to those in the 102 cities identified in the ACS to coincide with our scraping sample frame which used the 100 largest U.S. metro areas. We also limit to ACS respondents who said that they had looked for work in the past calendar year, to gather a sample of job-seekers as one might find on an online jobs board. As a final restriction, we also limit the ACS sample to respondents who attended at least some college.

are similar, though our matched sample is more black and has fewer other non-white racial groups. We also find that those in our matched sample are more likely to have attended some college and are less likely to have graduated from a four-year school. This is in part attributable to the fact that our ACS sample is uniform with respect to age, while our main sample has fewer very young and older individuals. Weighting the ACS to match our sample on age makes the two much more comparable.[11] This also reflects that our sample is comprised of online job-seekers, who may differ from those employed or seeking jobs elsewhere. We also compare our sample to the larger exam taking population in Table A2, focusing on males who lived in an MSA in high school and attended one college, similar to our job seekers. Again, our resume sample is less likely to earn a degree, more likely to be black, and has lower PSAT scores. These differences are not surprising as we cannot limit our administrative data to job-seekers.

3.4 Resume 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". 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.[12] 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.

[11] This reflects the fact that in our omit sample we attempt to limit to those we believe are finished with schooling by limiting to those who have not been enrolled for over a year. In the ACS, we can only limit to those not currently enrolled, many of whom will certainly return. We also note that the ACS is self reported while we use administrative data.

[12] 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.

4 Omitting Schooling

4.1 Framework

For simplicity, we 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. We also assume 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. 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 overqualification.

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 the negative signal of not completing, which comes at the cost of assuming 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:

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

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. What signal the employer observes depends on whether i omits or not. If i omits, employers observe $(hs_i, exper_i)$ – that i is a high school completer and his associated years of experience. If i 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 schooling.

The signaling value reflects costs, for example effort, associated with enrolling for one period. We expect that these are larger at higher quality schools ($qual_j$) as they may require more effort to attain the same grades or outcome relative to lower quality schools. If so, the likelihood of omitting is 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.

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 increasing in years of work experience, $exper_i$.

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.

4.2 Empirical Tests

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

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

Above, $Omit_{ij}$ is a binary indicator equal to 1 if job-seeker i omitted schooling j from his resume. $Degree_{ij}$ is an indicator if i earned a degree from school j. 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 should be interested. $Years_{ij}$ are years of schooling set equal to zero if $Degree_{ij}$ is equal to one.[13] Thus the first three 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 $PSAT_i$.[14] Similarly, since true ability, which we proxy with PSAT score, $PSAT_i$, is unobserved by employers at time of hire, we might expect that the decision to omit is uncorrelated with ability measures observed by the econometrician conditional on $Qual_j, Years_{ij}$, and $Exper_i$. In addition, we include a set of high school graduation year fixed effects, τ_t, to capture secular cohort

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

[14] 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.

effects. We estimate the model with ordinary least squares where ϵ_{ij} is an idiosyncratic error term.[15]

5 Omitting Results

5.1 Basic Statistics

We show summary statistics for the full sample, and then by omit status in Table 2. Column 1 shows means for our 4,506 matched resumes who attend only one school in NSC records. At the bottom of column 1 we show our key statistic – that 29 percent of resumes in our sample omit the one and only NSC college they attended. If we only consider college non-completers, 33 percent omit the only post-secondary schooling they attempted. 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.

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 columns 2 and 3.[16] We find that omitters are less likely to have attended a four-year school, enrolled for a shorter period of time, went to schools with lower average PSAT scores, and were more likely to have other training listed on their resumes. Mean differences all correspond to predictions above. Additionally, we find omitters are more likely to be minority and have lower PSAT scores themselves. These run counter to our predictions that individual characteristics beyond experience should not predict omissions. We next consider these factors jointly by estimating Equation 3.

5.2 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 – degree receipt, 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. Earning a degree negates the likelihood of omitting schooling, reducing the likelihood by about 40 percentage-points for both associate's and bachelor's degrees. Each year of completed schooling reduces the likelihood of omitting by nearly 18 percentage points for those who did not earn a degree. This suggests that two years of schooling would negate the 0.29 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 PSAT. Omitting average PSAT moves the coefficient on four-year school away from zero and increases statistical significance.

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

[16] We note that 2 percent of omitters earn a degree, contrary to our prediction, but that these are largely explained by those who also have other non-NSC training. Those not listing might also be concerned with appearing overqualified, as in Shen and Kuhn (2013).

In column 2 we add an indicator for the presence of other training. While this strongly predicts the likelihood of omitting, suggesting that job-seekers are more comfortable omitting schooling when there is an alternative to highlight, it does not meaningfully change coefficients on our schooling variables, implying that this is not driving results.

In the final two columns we turn to our predictions concerning student characteristics. In column 3 we add students' PSAT scores and find a precisely estimated zero. In column 4 we include indicators for race and omit PSAT to isolate the relationship. We find that conditional on schooling, race is unrelated to omitting. In column 5 we include all student characteristics with coefficients remaining unchanged. In fact, adding student i's PSAT score to the model, which is highly correlated with school average PSAT, years of enrollment, and race, has no impact on those other coefficients. 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 $PSAT_i$. This has implications for our results on employment differentials. If we find that omitters are less likely to be non-employed, we cannot fully separate the effect of omitting on employment from the selection effect where omitters are more sophisticated and would have had lower non-employment rates regardless. This would lead to an overestimation of the true employment gain to omitting. That the main coefficients are unchanged by including student characteristics, and that student characteristics themselves have no predictive power, suggests that the selection effect is likely not entirely driving results.

5.3 Robustness

In Table A3 we test several sample restrictions for our matched sample, showing that our results are robust to a host of alternative rules. These include: 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);[17] limiting to students who were enrolled for at least one or one-half calendar year (columns 5 and 6);[18] limiting to non-degree earners (coulmn 7); and finally, limiting it to those who list no other schooling or training (column 8).[19] We find omitting patterns are consistent across each of these samples.

[17] 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.

[18] 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.

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

5.4 Sample considerations

We next ask whether the way in which we construct our matched sample is driving results, in particular our reliance on resumes that list a high school. One concern is that job-seekers who omit college might be more likely to put their high school on their resume to compensate for a blank schooling section. In the following we test whether this is true and whether it drives results.

To do so we use unique names as an alternative way to match administrative records to resumes that does not rely on listing a high school. While we cannot know with certainty whether a first and last name combination is truly unique without a list of all names, we can observe whether a name is unique in our approximately 9 million male observations in administrative records and roughly 550,000 scraped resumes. We can further refine the likelihood that the unique names in both samples are the same individual by adding other information, such as observing whether work and schooling histories are in conflict, by eliminating names we observe in two different places at the same time. We construct what we call our *unique-unique* match by first eliminating resumes where work experience indicates employment prior to age 16 as defined by our administrative data, and retaining only 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 6,865 names that are unique in both samples. We show summary statistics for both samples, in addition to a combined sample allowing for matching on any rule, in Appendix Table A4.

We then use the overlap between the two samples to assess the quality of the unique names sample. To do this we find those in our original sample (matched on exact name, high school, and year) who also have unique names in both datasets. We then assess what the match rate would have been among the high school sample had we matched only on unique names. In other words, we assume that the name, high school, and year match is the truth, and ask what the error rate is when we match only on unique names. Of those with unique names, 10 percent (689) also list high school and graduation year, and met other requirements for our primary sample. We find that the match rate for this sub-sample is just over 92 percent. That is, we believe our unique to unique match falsely matches resumes to administrative data 8 percent of the time. This would inflate the omit rate of the unique-unique sample by construction, but by less than 8 percentage points, which is an upper bound by definition as some of them would have omitted.

We find that the overall omit rate for the unique names sample is 23 percent, while it is 29 percent for the main (high school) sample. However, in the overlapping sample, the omit rate is 33 percent, slightly larger than the main sample. This is one indication that relying on resumes listing high schools is not driving results.

In Table A5 we show results from our main specification using the high school sample in column 1, we then re-estimate the same model using the unique-unique sample in columns 2-3, and again using the combined sample in columns 4 and 5. This allows us to estimate whether listing a high school, conditional on observed characteristics, increases the likelihood that we observe a job-seeker omitting.

Comparing columns 1 and 2 shows only slight differences in results, with the primary difference being a now statistically significant relationship between omitting and Hispanic and Asian, which may be a function of different match rates based on unique names by race. In column 3 we add an indicator for whether the resume also listed high school. We find no relationship, with a precise point estimate around zero. In other words, among the sample matched on unique names, the sub-sample who also listed a high school were no more likely to omit schooling conditional on other observable factors. In columns 4 and 5 we re-run this exercise on a combined sample and include indicators for whether the resume matched on unique name, and whether a high school was listed. We again find no relationship between omitting and listing a high school, and find a positive relationship for the unique name sample, which reflects the error rate in matching only on name. Taken together, these results suggest that listing high school on a resume does not drive our results.

6 Theoretical and Empirical Implications of Omitting

In this section we consider theoretical and empirical implications of our results for the literature on returns to schooling, in particular for employer learning and statistical discrimination models (EL-SD), and describe how learning about selective omissions can teach us about the signaling versus human capital value of non-completed schooling.

Our results above demonstrate that a key assumption in EL-SD (and many other) models – that employers perfectly observe schooling, or that they observe schooling in the same way the econometrician does – is violated. This extends not only to the class of EL-SD models we focus on here, 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), within firms (Kahn and Lange, 2014), across occupations (Mansour, 2012), and with asymmetric learning (Kahn, 2013), among 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). Since job-seekers choose to omit schooling or not, we show that by augmenting the standard EL-SD model to include selective omission, a test of the signaling versus human capital of partially competed schooling results, offering the potential for insight into whether there is value in schooling omitted from resumes. Lastly, we implement a weak version of this test using job-seekers work histories, asking whether schooling omitted from resumes results in employment differentials.

6.1 Employer Learning and Statistical Discrimination

We start with the basics of employer learning models in Farber and Gibbons (1996) and Altonji and Pierret (2001). The intuition for which is as follows. Econometricians observe ability, schooling, experience and wages over time and retrospectively. Employers observe only schooling at initial hire and then observe noisy measures of productivity over time. If available ability measures, which are unobserved to employers at hire, are predictive of productivity they should become more salient in wage determination over time as ability is revealed to employers through observed productivity. Similarly, the relationship between schooling and earnings should decline with experience.

To be concrete, true productivity for worker i with t years of labor market experience is:

$$y_{it} = rs_i + \alpha_i q_i + \Lambda z_i + \eta_i + \tilde{H}(t_i). \tag{4}$$

Define s as productive characteristics observable to both employers and the econometrician. The most common case is schooling. q are characteristics observable only to employers not seen by the econometrician. z are measures of productivity observed by the econometrician but not the employer at time of hire; in most cases this is a measure of ability such as a test score. η_i are measures observable to neither the econometrician nor the employer. $\tilde{H}(t_i)$ is the structural relationship between experience and productivity which, by assumption, does not depend on either s or z.

Under a few assumptions, key among them a positive correlation between s and z, Altonji and Pierret (2001) and others show that r is non-increasing in t, and that Λ is non-decreasing. In other words, wages become increasingly correlated with the econometrician's proxy for productivity (employer learning) and less correlated with schooling (statistical discrimination). This relationship highlights the firm's hiring dilemma: schooling is a noisy signal of productivity. To test this, the literature uses the following reduced-form model, where w_{it} is log wage, s_i is schooling, z_i is a measure of ability, and t is potential experience measured as years since exiting schooling.

$$w_{it} = \beta_0 + \beta_1 s_i + \beta_2 z_i + \beta_3 (s_i \times t) + \beta_4 (z_i \times t) + \varepsilon_{it}. \tag{5}$$

Then, the model predicts a positive coefficient on β_1 and β_4, a zero coefficient on β_2, and a negative coefficient on β_3.

6.2 Adding Strategic Omitting

Our results above demonstrate that job-seekers strategically omit schooling. We now consider what implications follow for the class of EL-SD models as above. To do so, we rewrite the model from Equation 4 to allow for omitted schooling:

$$y_{it} = r(\Omega_i \bar{s}_0 + (1 - \Omega_i) s_i^*) + \alpha_i q_i + \Lambda(z_i, \Omega_i s_i^*) + \eta_i + \tilde{H}(t_i). \tag{6}$$

Ω_i refers to whether the job-seeker omits schooling, as in Equation 2. 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, true schooling (s_i^*) is now a z variable, something observed by the researcher, but not the employer.

This can be a test of the signaling versus human capital value of (partially completed) college, with the following intuition. If omitted schooling has productive value that employers learn about as workers accrue experience, the schooling gradient with respect to experience, β_3 in the reduced form model above (Equation 5), will be flatter (less negative) for omitters. This is because employers will learn about productive skills that were unobserved to them at hire. It follows then that the experience gradient on ability, β_4, will also be flatter (less positive in this case), assuming schooling and ability are positively correlated. If, on the other hand, omitted schooling in fact imparts 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 non-completers who omit. Above, we learned that omissions are typically about those attending non-selective schools for short durations. Thus, our test is largely about the productive value of attending non-selective schools. Extending the model to lying, which we discuss below, has similar though opposite signed implications.

6.3 Omitting and Employment

Estimating our augmented EL-SD model requires four pieces of information – reported schooling on resumes, true schooling from administrative data, a measure of productivity such as a test score, and a panel of wage histories. Our records contain three of these, but not the last. While we lack wages, we can create self-reported employment from our scraped resumes, allowing us to estimate a simple version of the relationship between omitting and non-employment spells in equilibrium with the following regression:

$$NonEmp_i = \alpha + \beta Omit_i + \Delta' Educ_i + \Pi' School_j + \Gamma' X_i + \tau_t + \epsilon_i. \quad (7)$$

In the model above, $Omit$ is an indicator if job-seeker i omitted his only NSC college from his resume. $Educ_i$ includes years of school if no degree and an indicator for earning a two- or four-year degree if either is the case, while $School_j$ are school quality measures including two- or four-year and the average PSAT score of school j's attendees. X_i includes race and PSAT score, and τ_t indicates high school graduation cohort. Our dependent variable, $NonEmp_i$, is the total number of months not employed since entering the labor market. Thus, we are comparing job histories of observationally similar job-seekers who graduated from the same cohort and attended similar schools for the same amount of time, who did or did not omit schooling.[20] We also consider a version of the model where we interact $Omit_i$ with schooling type and duration to test whether those omitting two- or four-year schools see different employment outcomes than similar peers, or if omitting shorter or longer durations of schooling matter. Results are in Table 4.

The first column, which does not include personal attributes, shows that job-seekers who omit their college experience from their resume have 1.3 more months of non-employment listed on their resumes. If it is the case that those who omit also "pad" their employment histories, then this is

[20] Potential experience is omitted as it is perfectly collinear with years of schooling and cohort.

an underestimate of the relationship. We also show that the relationship between other measures are of the expected signs. Degree earners have fewer years of non-employment, and those attending four-year schools have more since they attend school for longer on average. In column 2 we add personal characteristics with no impact on the relationship between omitting and employment. We do find that black job-seekers list fewer months of employment. Last, in column 3 we interact *Omit* with whether job-seekers attended a two- or four-year school, and for how long. We center years of enrollment such that the main effect of *Omit* is still taken at the mean. Results are unchanged.

These results suggest that on average omitters see larger employment gaps. We note that 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.

That we find employment differences, favoring those who do not omit, is consistent with employers preferring candidates with fewer gaps on their resumes, or those with some college education over none. Disentangling these mechanisms, among others, is difficult with existing data. As such, we present our results as a framework for future research.

7 Lying

In a final set of statistics 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.[21]

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 little 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

[21]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.

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. Second, lies about degrees are usually accompanied by information about what the job-seeker (said he) studied. We also observe lies about attendance and include these in our analysis for completeness. We discuss our sample, definition, methods, and then results below.

7.1 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. The final column of Table 1 shows resume attributes relative to the larger resume sample. 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.[22]

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 untrue 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.[23] 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. Table 5 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

[22] In columns 4 and 5 of Table A1 we compare our lying sample to an ACS analog, which drops the restriction that respondents reported having attended some college and narrows the age range to match those eligible for our lying sample. Again, we find that the racial breakdown is comparable, with our sample having more black students and fewer in other non-white groups. We also find that the ACS has more four-year graduates and fewer with some or no college.

[23] 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.

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

$$Lie_{ij} = \mathbf{\Pi}(HSi^*, 2Year^*_{ij}, 4Year^*_{ij}, PSAT^*_j) + \mathbf{\Theta}(Field^*_i) + \mathbf{\Upsilon}(Race_i, PSAT_i) + \tau_t + \epsilon_{ij} \qquad (8)$$

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 estimates 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.

7.2 Evidence of Lying and Field of Study

Table 6 shows results from Equation 8 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.[24] 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.

7.3 Lying and Employment

We can apply the same intuition to the EL-SD model for lying as for omitting, though with opposite signed implications. Referring to Equation 9 below, If job seekers lie about schooling on resumes, at hire employers observe $s^* > s$, where s^* is the falsely claimed schooling, for example a college degree, and s is true schooling, for example simply attending. If job-seekers lie ($\Psi_i = 1$), then their initial wage should be higher than their expected productivity conditional on schooling observed by the

[24] 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.

econometrician. If the schooling they lied about actually has value, employers will learn about it and wages will decline with respect to schooling, stagnate, or the worker will be fired. Hence, employer learning will be flatter in the empirical model. Further, a fixed cost may be incurred by the worker if he is fired, which may affect future employment. On the other hand, if the falsely claimed degree is a pure signal, and employers do not verify degrees, wage growth will be no different from those who actually earned the degree. If it is the case that more valuable degrees are easier to detect, then the likelihood of detection is directly related to the initial wage gain from lying. For example, while lying about a technical or specialized degree might initially earn one a well paying job, it might also imply skills that are difficult to fake. That we find lies are typically about non-technical degrees, for example Business and Administration, Humanities and Social Sciences, fits this pattern. Then observing lying along with work histories provides an informal test of whether degrees in these fields are useful signals of productivity for firms.

$$y_{it} = r(\Psi_i s^* + (1 - \Psi_i)s_i) + \alpha_i q_i + \Lambda(z_i, \Psi_i s) + \eta_i + \tilde{H}(t_i). \tag{9}$$

To formalize this test 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. We do this as follows.

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 10 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:

$$\begin{aligned} NonEmp_i = &\alpha + \Pi_1(Attend_{2,4}|True) + \Pi_2(Degree_{2,4}|True) \\ &+ \Psi_1(Attend_{2,4}|Lie) + \Psi_2(Degree_{2,4}|Lie) + \Phi X_i + \tau_t + \varepsilon_i. \end{aligned} \tag{10}$$

Table 7 shows results where the dependent variable is the number of months not employed since graduating school. We first show that actual degree earners and actual college attendees have fewer months not employed than terminal high school graduates. Four-year degree earners are not employed for about 11.5 fewer months than non-lying high school graduates. Two-year degree earners are employed about 10 months more. Those who attend two- and four-year schools, and do not lie about that or earning a degree, are employed 9 and 7 more months than truth-telling high school graduates, respectively.

In the bottom panel we show that individuals who lie about having attended but not completed a two-year school see about 3 fewer months of non-employment than high school graduates who did not lie, but about 6 fewer months of employment than actual two year attendees. We find similar signed but noisier results for those who are untruthful about attending a four-year school, and those who lie about a two-year degree. In the second column we add controls for race and student PSAT score with no substantive change to results.

These results suggests that those who lie about attending a two-year school see fewer months of non-employment compared with truth-telling terminal college graduates. Lying about attending a four-year school or earning a degree from a two-year school have similar signed coefficients, 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, keeping in mind that these are early-career workers.

There is much we cannot rule out and interpret these results with caveats. We cannot observe whether workers also lie about employment histories. Certainly many do. Whether this is correlated with lying about degrees we cannot say. Similarly, we cannot rule out heterogeneity in effects of lying, for example about a degree in Health compared with Business as we do not have the statistical power to test for it. The relationships we observe could very well be the result of some workers getting caught in a lie and having worse employment histories as a result, and some not, resulting in gains. We hope these questions motivate future work linking resumes with administrative records and potentially other data, in particular wages and firms.

8 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.

We couch these results in a broader literature on the returns to schooling, in particular focusing on employer learning and statistical discrimination models central to that literature. We augment these models to include the possibility that employers observe different information than researchers, and derive what implications this has for empirical estimates of returns to schooling, as compared with ability and other factors.

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.

Lastly, we 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 post-secondary credentials in the labor market: An experimental study. *American Economic Review*, 106(3):778–806.

Dynarski, S., Jacob, B., and Kreisman, D. (2018). 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*, 33:1098–1108.

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.

Galindo-Rueda, F. (2003). Employer learning and schooling-related statistical discrimination in britain.

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.

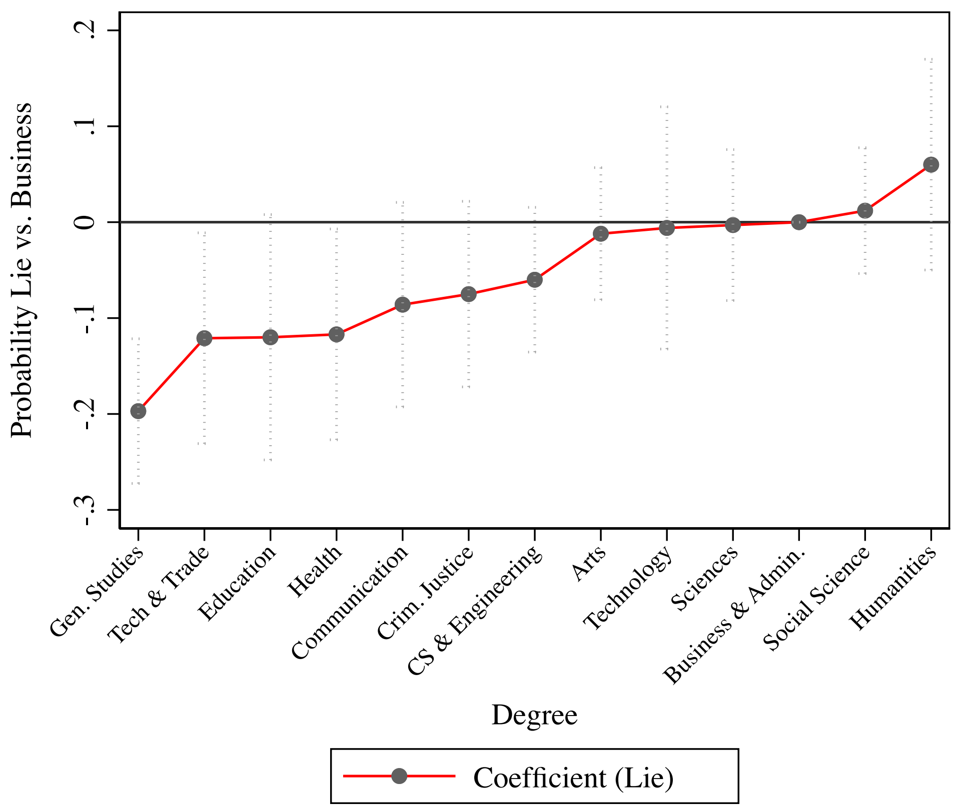
Stiglitz, J. E. (1975). The theory of "screening," education, and the distribution of income. *The American Economic Review*, 65(3):283–300.

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 6. 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: Summary Statistics of Resume Samples and Matched Samples

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All resumes	No emp. pre-2001	Listed HS & year	Enroll post-2003	Matched to Truth	Resume Sample	Lying Sample
Educational institutions listed	1.39	1.36	0.87	0.87	1	1.07	0.88
Jobs listed (if >0)	4.56	4.15	4.18	4.1	4.13	4.1	4.56
Years worked (if >0)	10.85	6.55	6.15	5.41	5.74	5.48	7.1
Months not working (2014)	2.40	2.49	2.72	2.88	2.56	2.58	1.81
High school year			2007.4	2009.9	2009.4	2009.9	2007.6
Listed HS & year	0.07	0.11					
And matched to Truth on HS, year	0.02	0.03	0.27	0.34			
And only 1 (true) college	0.01	0.01	0.11	0.13	0.40	1.00	
And only 1 (NSC) college	0.01	0.01	0.10	0.12	0.37		1.00
Obs.	556,651	382,953	41,559	33,517	11,244	4,506	4,154

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. Top half of table is mean values; bottom half is share of each column that fall in the categories listed. Column 1 is the full sample of scraped resumes. Column 2 is a sub-sample of column 1 who had no work experience listed prior to 2001. Column 3 limits to those who listed a a high school and graduation year. Column 4 is the subset of column 3 that attended high school after 2003. Column 5 are those matched to male PSAT, SAT, and AP exam takers between 2004 and 2014. Column 6 is the sample used for detecting omitting, which restircts to those attending exactly one college in administrative NSC records. Column 7 is the sample used for detecting lying; this limits to resumes from column (5) that list only one college (whether this is true or a lie), and also graduated high school by 2011. 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.

	All	Non-Omitters	Omitters
Degree	0.14	0.19	0.02
Attend four-year	0.55	0.62	0.36
Years Enrolled	0.73	0.83	0.49
Other training	0.27	0.21	0.42
School PSAT	88.65	90.88	83.23
White	0.39	0.42	0.31
Black	0.30	0.28	0.35
Hispanic	0.18	0.17	0.22
Asian	0.06	0.06	0.05
Other race	0.07	0.06	0.07
Student PSAT	84.10	86.82	77.50
Months Not Working	6.93	5.92	9.38
Share Months Not Working	0.09	0.08	0.12
Omit	0.29	0.00	1.00
Obs.	4,506	3,191	1,315

Notes: Sample includes resumes matched to administrative educational data for male job-seekers who only attended one college according to National Student Clearinghouse records, listed a high school on their resume, and graduated high school between 2004 and 2014. See Table 1 for sample definition. Omit is true if a job-seeker left a college experience off a resume. Other training is a non-NSC 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. Share of months is that value as a share of total months since graduating.

Table 3: The likelihood of omitting schooling

	(1)	(2)	(3)	(4)
Degree, 2-Year	-0.398***	-0.400***	-0.398***	-0.402***
	(0.035)	(0.032)	(0.032)	(0.032)
Degree, 4-Year	-0.422***	-0.405***	-0.406***	-0.408***
	(0.017)	(0.017)	(0.017)	(0.017)
Years Enrolled	-0.183***	-0.172***	-0.172***	-0.173***
	(0.008)	(0.008)	(0.008)	(0.008)
Potential Experience	0.023***	0.019***	0.019***	0.019***
	(0.003)	(0.003)	(0.003)	(0.003)
Attend Four-Year	-0.025	-0.023	-0.023	-0.020
	(0.017)	(0.017)	(0.017)	(0.017)
School Avg. PSAT	-0.003***	-0.004***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
Other Training		0.170***	0.170***	0.170***
		(0.015)	(0.015)	(0.015)
PSAT Score			0.000	-0.000
			(0.000)	(0.000)
Black				-0.017
				(0.017)
Hispanic				-0.007
				(0.019)
Asian				0.022
				(0.025)
Other Race				-0.012
				(0.027)
Cohort FE	×	×	×	×
Dep. Mean	0.292	0.292	0.292	0.292
R2	0.188	0.215	0.216	0.216
N	4,506	4,506	4,506	4,506

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 only attended one college in administrative records, a their high school on their resume, and graduated high school between 2004 and 2014. 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. Cohort FE are high school cohort.
(* $p < 0.10$, ** $p<0.05$, *** $p<0.01$)

Table 4: Omitting schooling and months of non-employment on resumes

	(1)	(2)	(3)
Omit	1.351**	1.311**	1.233**
	(0.540)	(0.557)	(0.588)
Degree, 2-Year	-4.210***	-3.956***	-4.632***
	(1.111)	(1.130)	(1.151)
Degree, 4-Year	-5.314***	-5.056***	-4.571***
	(0.940)	(0.942)	(0.898)
Years Enrolled	-2.094***	-1.989***	
	(0.343)	(0.346)	
Attend Four-Year	1.531***	1.144**	
	(0.562)	(0.573)	
School Avg. PSAT	0.020	0.045	0.043
	(0.023)	(0.028)	(0.028)
Other Training		0.569	0.614
		(0.512)	(0.511)
PSAT Score		0.006	0.006
		(0.014)	(0.014)
Black		2.180***	2.168***
		(0.585)	(0.584)
Hispanic		-0.070	-0.086
		(0.587)	(0.586)
Asian		0.256	0.259
		(0.915)	(0.917)
Other Race		0.116	0.169
		(0.897)	(0.900)
Years Enrolled 2-year (centered)			-2.932***
			(0.572)
Years Enrolled 4-year (centered)			-1.630***
			(0.332)
Omit*Years Enrolled 2-year (centered)			0.084
			(1.362)
Omit*Years Enrolled 4-year (centered)			0.014
			(1.375)
Cohort FE	×	×	×
R2	0.139	0.146	0.147
N	3,397	3,397	3,397

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 5: 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 6: 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 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.

	Matched to Truth	Omit Sample	ACS analog	Lying Sample	ACS analog
White	0.36	0.39	0.40	0.36	0.37
Black	0.32	0.30	0.22	0.35	0.27
Hispanic	0.19	0.18	0.22	0.18	0.24
Other race	0.12	0.13	0.15	0.12	0.11
PSAT	83.9	84.1	n/a	82.0	n/a
No college	0.29	0.00	0.00	0.30	0.47
Some college	0.59	0.86	0.53	0.55	0.23
2-year/AA Degree*	0.03	0.02	0.09	0.03	0.05
4-year Degree	0.09	0.11	0.38	0.13	0.25
Obs.	9,536	4,518	3,994	4,154	6,733

Notes: The matched sample are those who listed a high school on their resume and matched to administrative educational data. The Omit sample are those with only one college in the adminitrsative data. Column 3 are ACS respondents in 2016 who graduated high school between 2006 and 2014, lived in a city, were looking for work, and attended some college. The Lying sample have only 1 college on resume (whether this is true or a lie), and graduated high school by 2011. Column 5 are ACS respondents who graduated high school between 2004 and 2010, lived in a city, were looking for work, and graduated high school. Person weights are applied to ACS data. *In the ACS, only AA degrees are identified for two-year schools. In the administrative educational records we can observe degrees that are not Associates.

Table A2: Sample Comparison to Administrative Educational Records.

	Administrative Records		Omit Sample	
No college	0.00	(0.00)	0.00	(0.00)
Some college	0.70	(0.46)	0.86	(0.34)
Two year degree	0.03	(0.16)	0.02	(0.15)
Four year degree	0.28	(0.45)	0.11	(0.32)
White	0.59	(0.49)	0.39	(0.49)
Black	0.12	(0.33)	0.30	(0.46)
Hispanic	0.15	(0.35)	0.18	(0.39)
Asian	0.09	(0.28)	0.06	(0.23)
Other race	0.06	(0.24)	0.07	(0.25)
College avg. PSAT	95.67	(13.46)	90.07	(11.85)
Student PSAT	98.01	(21.88)	88.12	(20.70)
	5,457,728		4,478	

Notes: Administrative records are all males who took the any of the PSAT, SAT, or AP in high school graduating cohorts 2004-2014, live in an MSA, and attended one college. Omit sample are the subset or resumes matched to administrative records.

Table A3: Likelihood of omitting schooling - Robustness tests.

	Main (1)	Pre-2012 (2)	2-Year (3)	4-year (4)	Enroll > 1 (5)	Enroll > 0.5 (6)	No deg. (7)	No othr edu. (8)
Degree, 2-Year	-0.402***	-0.455***	-0.473***		-0.212***	-0.272***		-0.442***
	(0.032)	(0.036)	(0.034)		(0.044)	(0.035)		(0.021)
Degree, 4-Year	-0.408***	-0.480***		-0.361***	-0.225***	-0.282***		-0.361***
	(0.017)	(0.025)		(0.018)	(0.036)	(0.020)		(0.019)
Years Enrolled	-0.173***	-0.197***	-0.282***	-0.136***	-0.072***	-0.096***	-0.170***	-0.160***
	(0.008)	(0.011)	(0.019)	(0.009)	(0.015)	(0.009)	(0.008)	(0.009)
Potential Experience	0.019***	0.012***	0.019***	0.018***	0.009***	0.020***	0.020***	0.019***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Attend Four-Year	-0.020	0.000			-0.032	-0.019	-0.012	-0.032
	(0.017)	(0.022)			(0.029)	(0.021)	(0.018)	(0.020)
School Avg. PSAT	-0.004***	-0.004***	-0.002	-0.004***	-0.001**	-0.003***	-0.004***	-0.003***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Other Training	0.170***	0.157***	0.200***	0.147***	0.124***	0.152***	0.182***	
	(0.015)	(0.018)	(0.023)	(0.019)	(0.021)	(0.017)	(0.017)	
PSAT Score	-0.000	0.001	0.001	-0.001	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Black	-0.017	-0.019	-0.010	-0.022	-0.007	-0.006	-0.020	-0.016
	(0.017)	(0.020)	(0.028)	(0.020)	(0.018)	(0.018)	(0.019)	(0.019)
Hispanic	-0.007	-0.008	-0.010	0.008	0.054**	0.031	-0.013	-0.026
	(0.019)	(0.024)	(0.029)	(0.026)	(0.025)	(0.022)	(0.021)	(0.021)
Asian	0.022	0.074**	0.036	0.020	0.060**	0.037	0.000	-0.000
	(0.025)	(0.033)	(0.051)	(0.027)	(0.029)	(0.025)	(0.029)	(0.025)
Other Race	-0.012	-0.027	-0.014	-0.002	0.023	0.009	-0.021	-0.021
	(0.027)	(0.033)	(0.042)	(0.034)	(0.035)	(0.030)	(0.029)	(0.031)
Dep. Mean	0.292	0.295	0.410	0.193	0.091	0.170	0.332	0.231
R2	0.216	0.268	0.144	0.223	0.118	0.156	0.181	0.175
N	4,506	2,937	2,049	2,457	1,677	2,949	3,887	3,275

Notes: Column 1 replicates column 4 of Table 3. 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: Summary statistics across matching rules.

	H.S. Sample			Names Sample			Combined Sample Sample		
	All	No omit	Omit	All	No omit	Omit	All	No omit	Omit
Omit	0.29	0	1	0.23	0	1	0.25	0	1
	(0.45)			(0.42)			(0.43)		
Attend four-year	0.55	0.62	0.36	0.73	0.82	0.44	0.67	0.75	0.41
	(0.50)	(0.49)	(0.48)	(0.44)	(0.38)	(0.50)	(0.47)	(0.43)	(0.49)
Two-year degree	0.02	0.03	0.01	0.02	0.02	0.02	0.02	0.03	0.01
	(0.15)	(0.17)	(0.10)	(0.15)	(0.15)	(0.13)	(0.15)	(0.16)	(0.11)
Four-year degree	0.11	0.16	0.01	0.34	0.41	0.09	0.26	0.33	0.06
	(0.32)	(0.36)	(0.09)	(0.47)	(0.49)	(0.29)	(0.44)	(0.47)	(0.23)
Years Enrolled*	0.73	0.83	0.49	0.73	0.78	0.57	0.73	0.79	0.54
	(0.74)	(0.81)	(0.44)	(0.90)	(0.96)	(0.64)	(0.84)	(0.91)	(0.56)
School PSAT	88.65	90.88	83.23	93.19	95.69	84.82	91.64	94.08	84.28
	(16.19)	(16.15)	(14.94)	(16.76)	(15.88)	(16.91)	(16.66)	(16.23)	(15.72)
Other schooling	0.27	0.21	0.42	0.31	0.26	0.46	0.30	0.25	0.44
	(0.45)	(0.41)	(0.49)	(0.46)	(0.44)	(0.50)	(0.46)	(0.43)	(0.50)
Black	0.30	0.28	0.35	0.24	0.21	0.34	0.26	0.23	0.35
	(0.46)	(0.45)	(0.48)	(0.43)	(0.41)	(0.48)	(0.44)	(0.42)	(0.48)
Hispanic	0.18	0.17	0.22	0.12	0.10	0.19	0.15	0.13	0.21
	(0.39)	(0.38)	(0.41)	(0.33)	(0.31)	(0.39)	(0.36)	(0.34)	(0.41)
Asian	0.06	0.06	0.05	0.10	0.10	0.08	0.08	0.09	0.07
	(0.23)	(0.24)	(0.21)	(0.30)	(0.30)	(0.27)	(0.27)	(0.28)	(0.25)
Other race	0.07	0.06	0.07	0.07	0.06	0.08	0.07	0.06	0.07
	(0.25)	(0.24)	(0.26)	(0.25)	(0.24)	(0.27)	(0.25)	(0.24)	(0.26)
Student PSAT	84.10	86.82	77.50	90.43	93.27	80.90	88.20	91.09	79.52
	(27.34)	(27.78)	(25.06)	(28.21)	(28.49)	(24.99)	(28.09)	(28.49)	(24.93)
Names Sample	0.15	0.14	0.17	1.00	1.00	1.00	0.64	0.66	0.59
	(0.36)	(0.35)	(0.38)	(0.00)	(0.00)	(0.00)	(0.48)	(0.47)	(0.49)
HS Sample	1.00	1.00	1.00	0.10	0.09	0.14	0.42	0.40	0.49
	(0.00)	(0.00)	(0.00)	(0.30)	(0.28)	(0.35)	(0.49)	(0.49)	(0.50)
	4,506	3,191	1,315	6,865	5,287	1,578	10,682	8,016	2,666

Notes: Table shows means and standard deviations by method of matching resumes to administrative data, and whether the job-seeker omits schooling. Columns 1-3 are records matched on High School, graduation year and exact name. Columns 4-6 are matched on names that are unique in resumes and College Board records. Columns 7-9 are matched on either.

Table A5: Does the sample matter for likelihood of omitting schooling?

	Main	Match on Unique Name		Match on Any	
	(1)	(2)	(3)	(4)	(5)
Degree, 2-Year	-0.402***	-0.425***	-0.425***	-0.417***	-0.426***
	(0.032)	(0.033)	(0.033)	(0.023)	(0.023)
Degree, 4-Year	-0.408***	-0.380***	-0.381***	-0.374***	-0.391***
	(0.017)	(0.017)	(0.017)	(0.012)	(0.013)
Years Enrolled	-0.173***	-0.143***	-0.143***	-0.147***	-0.152***
	(0.008)	(0.008)	(0.008)	(0.006)	(0.006)
Potential Experience	0.019***	0.026***	0.025***	0.023***	0.023***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Attend Four-Year	-0.020	-0.094***	-0.094***	-0.065***	-0.067***
	(0.017)	(0.017)	(0.017)	(0.013)	(0.013)
School Avg. PSAT	-0.004***	-0.002***	-0.002***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Other Training	0.170***	0.105***	0.105***	0.131***	0.129***
	(0.015)	(0.011)	(0.011)	(0.009)	(0.009)
PSAT Score	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Black	-0.017	0.001	0.001	-0.001	-0.001
	(0.017)	(0.013)	(0.013)	(0.011)	(0.011)
Hispanic	-0.007	0.071***	0.071***	0.035***	0.037***
	(0.019)	(0.016)	(0.016)	(0.012)	(0.012)
Asian	0.022	0.046***	0.046***	0.046***	0.042***
	(0.025)	(0.015)	(0.015)	(0.013)	(0.013)
Other Race	-0.012	0.001	0.001	-0.005	-0.007
	(0.027)	(0.019)	(0.019)	(0.016)	(0.016)
Listed High School			-0.005		-0.005
			(0.017)		(0.017)
Unique Name Match					0.046***
					(0.017)
Dep. Mean	0.292	0.230	0.230	0.250	0.250
R2	0.216	0.274	0.274	0.247	0.250
N	4,506	6,865	6,865	10,682	10,682

Notes: Main sample matches on name and high school. Unique name sample matches on names that are unique in each of the administrative educational data and resume scrape (and in same state). Linear probability model with omitting college from resume as dependent variable. Sample includes scraped resumes matched to administrative educational data for job-seekers who only attended one college and graduated high school between 2004 and 2014. (* p < 0.10, ** p<0.05, *** p<0.01)