The Mark of a Woman's Record: Gender and Academic Performance in Hiring

Natasha Quadlin[a]

Abstract

Women earn better grades than men across levels of education—but to what end? This article assesses whether men and women receive equal returns to academic performance in hiring. I conducted an audit study by submitting 2,106 job applications that experimentally manipulated applicants' GPA, gender, and college major. Although GPA matters little for men, women benefit from moderate achievement but *not* high achievement. As a result, high-achieving men are called back significantly more often than high-achieving women—at a rate of nearly 2-to-1. I further find that high-achieving women are most readily penalized when they major in math: high-achieving men math majors are called back three times as often as their women counterparts. A survey experiment conducted with 261 hiring decision-makers suggests that these patterns are due to employers' gendered standards for applicants. Employers value competence and commitment among men applicants, but instead privilege women applicants who are perceived as likeable. This standard helps moderate-achieving women, who are often described as sociable and outgoing, but hurts high-achieving women, whose personalities are viewed with more skepticism. These findings suggest that achievement invokes gendered stereotypes that penalize women for having good grades, creating unequal returns to academic performance at labor market entry.

Keywords

education, gender, employment, audit methods

By most measures, girls and women have better academic performance than do boys and men. Beginning in kindergarten, girls are rated as having better social and behavioral skills than boys, which helps support their early achievement in school (DiPrete and Jennings 2012; Entwisle, Alexander, and Olson 2007; Perkins et al. 2004). The gender gap continues through high school and even college: female students generally earn better grades than their male counterparts, and teachers rate them as having better competencies and skills (DiPrete and Buchmann 2013;

Downey and Vogt Yuan 2005; Dumais 2002). The female advantage in academic performance is so pervasive that some commentators refer to schools as "feminized," because they view schools as promoting and rewarding

[a]The Ohio State University

Corresponding Author:
Natasha Quadlin, The Ohio State University, 238 Townshend Hall, 1885 Neil Avenue Mall, Columbus, OH 43210
Email: quadlin.2@osu.edu

qualities that are more common among female students than male students (Rosin 2012; Tyre 2008).

Extant research shows that women receive some benefits due to their strong academic performance across levels of education.[1] For example, women have higher rates of college completion than do men, largely because they earn better grades in high school. Because girls are more prepared than boys for the rigors of college academics, they are more likely to enroll in higher education and are more likely to complete their degrees once enrolled (Buchmann and DiPrete 2006; Carbonaro, Ellison, and Covay 2011; Riegle-Crumb 2010). Although college completion is an extremely important outcome—especially given the current focus on "college for all" and the necessity of a college degree to achieve a middle-class lifestyle (Goyette 2008; Rosenbaum 2001)—little research has assessed whether academic performance benefits women *outside of schools*. That is, we have yet to understand whether academic performance pays off for women once they leave school, or if the time and effort women spend trying to enhance their academic performance simply does not improve their outcomes in other life domains.

In particular, research has yet to examine whether women's academic performance translates in the labor market. Social scientists often question whether achievement matters to employers, and if the grades students receive in school have any bearing on the way employers evaluate job applicants (Becker 1964; Bills 2003; Bills, Di Stasio, and Gërxhani 2017; Brown 2001; Collins 1971, 1979; Miller 1998; Rosenbaum and Kariya 1991; Stiglitz 1975). These questions are especially relevant for women job applicants, as women face broad disadvantages in the workplace, including assumptions that they are less competent, less committed to their jobs, and less likeable than men (Correll, Benard, and Paik 2007; Fiske et al. 2002; Ridgeway 2011; Rivera and Tilcsik 2016). To what extent can academic performance help women overcome these gendered stereotypes? In other words, does strong academic performance help women demonstrate their employability? Or, do women experience relatively poor labor market outcomes despite evidence of high achievement?

This article is the first to investigate how gender and academic performance affect employment outcomes among recent college graduates. I conducted two interrelated studies to understand *how* men's and women's academic performance affects their chances of advancing to the interview stage with employers, along with *why* employers make the decisions they do. For the first study, a résumé audit study, I submitted over 2,100 fictional applications to entry-level job openings in labor markets across the United States. These applications experimentally manipulated applicants' gender, achievement (as measured by grades), and college major, which allows me to analyze how these factors are related to job callbacks. For the second study, a survey experiment, I replicated the audit study with a sample of more than 250 respondents who work as hiring decision-makers in their companies. I asked them not only to evaluate whether they would recommend interviewing the applicants, but also to provide closed- and open-ended feedback about the applicants' qualifications and personal characteristics.

The audit study shows that men's and women's academic performance have very different consequences in the labor market. Men have approximately the same outcomes regardless of their achievement, such that men with a C+ average are called back at about the same rate as their A-average counterparts. Women's achievement, however, has an inverted U-shaped effect. Women with moderate achievement receive more callbacks than do women with low achievement, but this advantage does not extend to high-achieving women. Consequently, high-achieving men are called back significantly more often than high-achieving women—at a rate of nearly 2-to-1. I further find that the penalty for high-achieving women is most concentrated among women who majored in

a particular field—mathematics. In other words, high-achieving women may be most readily penalized when they demonstrate achievement in STEM fields where they are underrepresented and expected to perform poorly. Using the closed- and open-ended data from the survey experiment, I show that these gendered patterns may be attributable to employers' shifting standards for men and women job applicants. Employers value competence and commitment among men applicants, but they privilege likeability among women applicants, ultimately creating liabilities for high-achieving women.

ACADEMIC PERFORMANCE, GENDER, AND EMPLOYMENT

The Role of Academic Performance in Hiring

The relationship between education and the labor market can most accurately be described as complicated. Particularly now, as many students incur substantial debt to attend college (Dwyer, McCloud, and Hodson 2012; Quadlin and Rudel 2015), and universities are under mounting pressure to show their degrees carry weight with employers (Arum and Roksa 2011, 2014), scholars are revisiting the question of whether grades matter for securing a job. Over the past several decades, sociologists have offered conflicting perspectives on this basic premise. Scholars in the human capital tradition, for example, posit that schools teach students skills that are relevant in the workplace, and grades are a reflection of how well students have developed those relevant skills (see, e.g., Becker 1964). Academic performance is therefore seen as a reflection of students' preparation for the labor market. High grades indicate that students have skills employers value, and low grades indicate that students may pose a risk to the company if they are hired (Miller 1998; Rosenbaum and Kariya 1991). In a similar vein, screening theory contends that employers have relatively little information about job applicants, so they must rely on measures such as academic performance to make comparisons between applicants (Bills 2003; Stiglitz 1975). Numerical measures such as grade-point averages (GPAs) may be especially useful, because they allow employers to easily rank-order an applicant pool or create cutoffs to disqualify applicants below a certain threshold.

Other theories suggest that employers are not concerned with applicants' academic performance as long as they have obtained the minimum credentials needed to do the job. According to Collins's (1971, 1979) theory of credentialism, a college degree acts as a mark of membership in an elite group, rather than an indicator of technical skills. When employers post a job opening, they may list an education requirement (e.g., a bachelor's degree), but only to ensure that applicants possess values and cultural ideals consistent with others in the organization. Once the applicant has surpassed this minimum, details such as academic performance are more or less irrelevant (see also Brown 2001).

Taken together, scholars have historically disagreed about the role of academic performance in hiring—and, thus, whether achievement has an effect on labor market outcomes is an empirical question in and of itself. But conspicuously absent from this literature is a discussion of how gender may complicate the signal of academic performance from an employer's perspective. Because gender carries its own independent assumptions about job performance, employers may interpret—and reward—men's and women's achievement very differently.

Gender, Assumptions about Job Performance, and the Signal of Achievement

In the course of the hiring process, most employers must make assumptions about applicants they have never met in person. Many of these assumptions are made on the basis of status characteristics, such as gender, which affect how people are evaluated in relation to each other (Foschi, Lai, and Sigerson

1994; Pager, Western, and Bonikowski 2009; Ridgeway and Correll 2004; Tilcsik 2011). Gender shapes many perceptions that are routinely made in social interactions, but research shows that gender is particularly salient in the judgment of others' competence, such that women are broadly assumed to be less competent and less capable than men (Fiske et al. 2002; Ridgeway 2011). Employers also subscribe to these perceptions: women tend to be rated as less competent than similarly-situated men in hiring, which can lead to fewer interviews and job offers, lower starting salaries, and fewer promotions to managerial positions for women, among other consequences (Correll et al. 2007; Heilman 2001). Women may thus have worse employment outcomes than men because they are viewed as less equipped to do their jobs well.

Women are also perceived as less committed to their jobs than men, which further disadvantages them in the hiring process. The "ideal worker" has long been depicted as someone who is singularly dedicated to their job and has few outside commitments, such as personal relationships or hobbies (Acker 1990; Blair-Loy 2003; Kelly et al. 2010). The few employees who are able to live up to this norm are disproportionately men, as women are often expected to juggle competing demands on their time (including, but not limited to, housework and childcare) that prevent them from focusing exclusively on their jobs (Correll et al. 2007; Damaske 2011; Williams, Muller, and Kilanski 2012; Williams 2001). Employers' perceptions of women as less committed may prevent women from advancing through the hiring process, even if their résumés are functionally identical to men's (Rivera and Tilcsik 2016).

To what extent can women's academic performance help them overcome these gendered stereotypes? Although many employers hold negative perceptions of women on the basis of their status characteristics, women may be able to offset these perceptions if their résumés provide evidence of superior grades. Existing research provides two contrasting perspectives on whether this is the case. One perspective implies that high-achieving women are highly sought-after in the workplace. As public support for gender equality in the workplace has increased, and more laws that support working women have been passed (Gerson 2010; Pedulla and Thébaud 2015), companies have grown increasingly eager to recruit women who add to the diversity of the workforce—particularly if women meet or exceed the qualifications of the company's existing employees (Correll 2016). This narrative has been especially prominent in the tech industry, where women are extremely underrepresented, and highly-qualified women are thought to be rare (Tam 2016). If this is the case, then high-achieving women should have employment outcomes equal to or greater than high-achieving men, because they allow companies to diversify their workforce while maintaining their usual hiring standards. One potential consequence of high-achieving women being sought-after is that women with strong academic performance may be viewed as overqualified. This would suppress women's chances of being called back for many positions, but perhaps not those with the most competitive pay and benefits.

An alternative perspective suggests that high-achieving women may face additional difficulties that negate the returns they would otherwise receive from their strong academic performance. Although women are disadvantaged by the perception that they are less competent than men, if women present as *too* competent, they tend to be assessed as lacking warmth (Cuddy, Fiske, and Glick 2008; Eagly and Carli 2007; Fiske et al. 2002). In other words, women experience a competence–likeability tradeoff that hurts them if they veer too far in either direction. In the context of hiring, women who are perceived as lacking warmth may be penalized because they are assumed to be less communal and not as nice as other applicants (Rudman 1998). Men do not necessarily experience this tradeoff, and they are able to be perceived as competent and likeable simultaneously.

Similarly, employers may unconsciously assign less value to women's achievement than men's achievement, thus reducing women's returns to academic performance. Research

shows that employers use shifting standards to emphasize the credentials of desired groups (Phelan, Moss-Racusin, and Rudman 2008; Uhlmann and Cohen 2005). When high-achieving men apply for a job, for example, their academic performance may be viewed as an essential criterion for hiring, allowing them to advance in the hiring process. When high-achieving women apply for a job, their academic performance may be viewed as nonessential (or otherwise undesirable) and they may be passed over. If employers subtly shift their definitions of merit, and their ideas about the qualifications needed to do a job, then high-achieving men and women may have very different labor market outcomes.

In summary, although research shows that women face negative perceptions in hiring, it is not known whether women's academic performance can help offset these perceptions—or whether high achievement further penalizes women. Research is thus needed to understand whether achievement differentially affects men's and women's labor market outcomes. Further, majors are broadly understood as gendered, and high achievement in one major may have very different implications than high achievement in another. For these reasons, I consider how academic performance and gender affect employment outcomes for students with majors in diverse fields.

Gendered Majors and Gendered Returns to Achievement

College major may be an additional factor that affects the way employers think about academic performance. Depending on an applicant's field of study, employers may attach different meanings to achievement that affect how applicants are evaluated. A large body of scholarship points to gender differences in the selection of majors, such that men are over-represented in STEM majors and women are over-represented in the arts, humanities, and majors associated with the helping professions (Buchmann, DiPrete, and McDaniel 2008; England and Li 2006; Jacobs 1996). As a result of these differences in sex composition across majors, employers may

regard men and women differently if they major in male- or female-dominated fields of study. Prior research suggests two main ways these patterns could unfold.

One possibility is that students have different returns to achievement in majors where their gender is over- or underrepresented. Women who major in STEM fields are particularly salient in light of recent national attention on increasing women's participation in STEM. Despite widespread programs intended to recruit women to STEM and retain them once they have enrolled, most STEM majors remain male-dominated (Charles and Bradley 2002, 2009; Correll 2001, 2004; Riegle-Crumb et al. 2012). Women who have beaten the odds by not only persisting in STEM, but exhibiting strong academic performance, may be regarded very differently than their male counterparts. As noted earlier, some evidence suggests these women will be sought-after, because companies have incentives to increase gender diversity, especially in tech and other industries adjacent to STEM fields (Tam 2016). But research in organizational psychology offers a different perspective: women may be penalized for excelling in male-dominated majors because "[w]omen, quite simply, are not supposed to excel at jobs and tasks that are designated as male in our culture" (Heilman 2001:667). High-achieving women STEM majors may have worse employment outcomes than high-achieving men STEM majors because they have violated gendered prescriptive norms. Men might experience similar penalties for achievement in female-dominated majors, although these penalties may not be as pronounced, given research on men's structural advantages in female-dominated industries (Williams 1992).

A second possibility is that men and women who major in the same fields have the same returns to achievement. Human capital theory posits that employers evaluate job applicants on the basis of their competencies and skills (Becker 1964). Because students with the same achievement in the same major have theoretically demonstrated the same competencies, it follows that these students

should have the same employment outcomes regardless of their gender. In this case, there could still be differences in employment outcomes across majors, because fields of study signal different types of human capital (Kim, Tamborini, and Sakamoto 2015; Roksa and Levey 2010). Students who majored in STEM or business, for example, may have better outcomes than students who majored in English, because they are perceived as having more job-related skills and can transition seamlessly to an entry-level position at a company. But, these outcomes should be consistent for men and women within majors.[2]

Using Multiple Methods to Study the College-to-Work Transition

Why study the effects of achievement and gender among recent college graduates? This may seem like a counter-intuitive segment of the labor market in which to examine these factors, considering that much of the empirical literature on the school-to-work transition focuses on high school graduates (Griffin, Kalleberg, and Alexander 1981; Johnston and Bachman 1973; Rosenbaum 2001), and the modal highest degree attained among U.S. workers is a high school diploma, not a bachelor's degree (Ryan and Siebens 2012). I chose to assess the college-to-work transition for three main reasons.

First, although the earnings premium associated with a college degree has grown over the past few decades (Brand and Xie 2010; Hout 2012), the public perceives the labor market for college graduates as increasingly competitive, because more people are graduating from college (Arum and Roksa 2014). Some students attempt to differentiate themselves by spending time on academics, whereas others assume "grades don't matter" and direct their efforts toward extracurricular activities, paid work, or other pursuits (Armstrong and Hamilton 2013; Hamilton 2013, 2016; Quadlin 2016). This article assesses the extent to which academic performance benefits (or even penalizes) students applying to their first jobs, and whether the effects of achievement vary by gender and major.

Second, by focusing on college graduates, I minimize selection biases that could arise at other levels of education. If a high-achieving recent high school graduate applied to a full-time job, many employers would question why that student was not going to college, setting off a red flag that would prevent the person from advancing to the interview stage. These red flags would be masquerading as achievement penalties, but in reality they would be unrelated to the employer's assessment of the applicant's abilities. Recent college graduates, in contrast, could reasonably be expected to apply to full-time jobs at all achievement levels, helping to isolate the effect of achievement from other signals.

Third, most college graduates must declare a field of study, whereas this practice is not common among high school graduates. Because majors signal diverse competencies and carry gendered connotations, fields of study may create heterogeneity in returns to achievement among men and women.

To understand how gender, achievement, and major affect employment outcomes among recent college graduates, I first conducted a résumé audit study (Bertrand and Mullainathan 2004; Correll et al. 2007; Gaddis 2015; Pedulla 2016; Rivera and Tilcsik 2016; Tilcsik 2011). Audit studies are becoming increasingly common in the social sciences, because they offer several advantages over other types of data. The randomized experimental design allows for more robust causal inference than observational data, because applicants with different gender, achievement, and college major are randomly assigned to apply to positions. The audit method, therefore, isolates the effects of these signals on callbacks (Pager 2003, 2007) and rules out alternative explanations that can arise with observational data, such as men's and women's self-selection into different job openings. Audit studies also have greater external validity than laboratory experiments, because they produce data about real employers who believe they are making decisions about real job applicants (Rivera and Tilcsik 2016).

As a follow-up to the audit study, I conducted a survey experiment with a sample of

respondents who work as hiring decision-makers in their companies. The audit study provides externally valid data about how employers evaluate achievement, gender, and major in the entry-level labor market, but it provides few clues as to *why* employers make the decisions they do (Pedulla 2016; Rivera and Tilcsik 2016). The survey experiment includes both closed- and open-ended questions that explore the mechanisms behind the audit study results.

STUDY 1: RÉSUMÉ AUDIT STUDY

How does gender shape the relationship between achievement and employment outcomes? And how is this relationship affected by students' fields of study? To answer these questions, I conducted a résumé audit study (also known as a correspondence audit; for a discussion of terminology, see Gaddis 2018) by submitting 2,106 fictional applications to 1,053 job openings using a leading national job-search website. I then tracked responses to these applications to determine the effects of gender, achievement, and major on employment outcomes for recent college graduates.

The applications experimentally manipulated three variables of interest that were displayed on applicants' résumés. *Gender* was signaled using first names. I used regionally specific names that were among the top-five baby names in the region in the mid-1990s, when traditional-aged students in the college class of 2016 were born: Ashley and Michael (Midwest); Samantha and Nicholas (Northeast); Sarah and Daniel (Southwest); Jessica and Christopher (South); and Stephanie and David (West).[3] I also randomly assigned one of two common surnames to men and women within regions. Unlike other studies that have used racialized names (Bertrand and Mullainathan 2004; Gaddis 2015), the names I used were not intended to signal race or ethnicity.[4] Based on typical demographics for these names, however, employers may have interpreted the applicants as white. This is a key scope condition that is considered further in the Discussion section.

Achievement was signaled using grade-point averages (GPAs).[5] Because there is little consensus as to what constitutes a high or low GPA, I did not choose specific values of GPA to signal different levels of achievement. Rather, I used a random number generator to assign GPAs that ranged between 2.50 (C+ to B– average) and 3.95 (A average). I also included a major GPA that was .05 points higher than the cumulative GPA, which I used to reiterate both the GPA signal and the field of study signal. I ultimately created categories of GPA to simplify the presentation of results, as discussed below, but GPAs were entered on résumés as a continuous variable. Although recent college graduates often list their GPA on their résumés (and the résumés I created would not have looked atypical to hiring managers on this basis), guidebooks for first-time jobseekers offer different advice about this practice. Some contend GPAs should always be listed on résumés for recent college graduates, while others suggest students only include their GPA if it exceeds a certain threshold, such as 3.0 or 3.5. Consequently, low-achieving applicants in this study may have been penalized not only for having low achievement, but also for violating employers' expectations by exposing their low GPA. I consider this possibility in the Results section, including the extent to which it may differentially affect men's and women's outcomes.[6]

Field of study was signaled using majors: English, mathematics, or business. These majors are female-dominated, male-dominated, and sex-neutral in their respective sex compositions, and they represent a range of topic areas. I considered multiple alternatives when selecting majors, but these majors satisfied two main criteria that made them ideal for this study. First, these majors offer relatively broad academic training that translates across many employment sectors. Majors that are more applied in their orientation, such as nursing or computer science, would be less likely to be seen as appropriate training for the entry-level office positions I applied to, and employers might have dismissed applicants who majored in these relatively narrow fields. Second, I conducted pretests with

undergraduates to confirm that these majors' *perceived* sex compositions are in line with their *actual* sex compositions, which is important for audit studies and other experiments that rely on signaling. Psychology, for example, skews female in its sex composition (England and Li 2006), but many people believe it is sex-neutral, so psychology is a messy signal of sex composition.

Because each employer received two résumés, I designed two sets of application materials that were equivalent aside from the main experimental manipulations. I used two aesthetically different résumé templates that were presented as exemplars in guidebooks for first-time jobseekers. Within regions, both applicants indicated they attended the same large, public, moderately selective university. These universities were ranked between 50 and 100 by *U.S. News and World Report,* and they were either in the same state as the target city or in a nearby state.[7] I chose moderately selective institutions, because the most- and least-selective institutions carry strong signals of high school achievement that could obscure the experimental manipulations. The résumés reported addresses that were near each other in the target city; the addresses were real, but the apartment numbers were made up. The résumés also included phone numbers linked to voice mailboxes with either a male or female greeting, where employers could leave voicemail messages. Both applicants reported membership in three campus groups, including one leadership position; I pretested these to ensure they were perceived as similarly gendered and prestigious.[8] Both applicants reported a summer internship, a part-time job during the school year, and a short list of skills (e.g., computer skills, language proficiency). Finally, both applicants submitted cover letters that matched the qualifications on the résumés and were tailored with the name of the employer and the position being applied to.

I constructed a sampling frame using a leading national job-search website. College career centers routinely advise students to use these websites to facilitate their transition from college to work, so many recent graduates rely on these websites to search and apply for jobs. Although many studies emphasize social networks as a primary mechanism of job placement, as well as inequality in employment outcomes (DiMaggio and Garip 2012; Lin 1999), networks may be less relevant for recent college graduates, who often lack referral channels (Arum and Roksa 2014). The applicants in this study were also presented as moving to a new city where they presumably had few contacts, so they would be especially likely to use these websites in their job search.

To maximize the overall sample size and enhance the geographic diversity of the sample, applications were submitted to positions in major metropolitan areas corresponding to five regions of the United States: Midwest, Northeast, Southwest, South, and West. For each city, I compiled a list of job openings that were full-time; located within a 30-mile radius; posted within the past 30 days; in the job categories "entry level" or "general," which helped screen out jobs with overly specific criteria such as an engineering degree; and could be applied to directly through the job-search website.[9] The education requirements of the positions varied but, as Pedulla (2016) notes, many jobs that do not require a college education are not posted on national recruitment websites. Approximately 70 percent of the jobs explicitly required a college degree, and most of the remaining jobs listed a college degree as preferred or did not specify an education requirement.[10] If employers had posted more than one opening, I retained only the most recent opening to ensure employers did not review multiple sets of the same résumés. Although I did not target any particular type of job in this study, instead opting to keep the industries and positions as broad as possible, I recognize that graduates with certain types of degrees are likely to seek jobs in certain industries. Accordingly, I conducted robustness checks to examine outcomes among students applying to jobs in industries most relevant to their major, as discussed below.[11]

Once the list of job openings was finalized for a given city, I randomly assigned a matched pair of applicants to each employer.

Table 1. Proportion of Applicant Callbacks by Gender, Achievement, and Major

	Callbacks/Applications	% Callbacks
Gender		
Men	147 / 1,053	14.0
Women	125 / 1,053	11.9
Achievement		
C+/B– average (2.50 to 2.83 GPA)	50 / 519	9.6[a]
B–/B average (2.84 to 3.20 GPA)	68 / 531	12.8
B/B+ average (3.21 to 3.59 GPA)	86 / 527	16.3[a]
A–/A average (3.60 to 3.95 GPA)	68 / 529	12.9
Major		
English	76 / 699	10.9[b]
Business	106 / 708	15.0[b]
Math	90 / 699	12.9
Overall Callbacks	272 / 2,106	12.9

[a,b]Callback rates for these categories are significantly different from each other at $p < .05$ (two-tailed tests).

The applicants differed in terms of gender and major, and GPA was an additional randomly assigned component. Applications were submitted two days apart to reduce employer suspicion. Résumé format, content, and order of submission were randomized and counterbalanced, so these aspects of the applications were uncorrelated with the experimental manipulations. Applicants' characteristics, as well as the characteristics of the job openings applied to, are summarized in Appendix Table A1. If starting salaries or salary ranges were listed (about 40 percent of positions), I recorded them in order to conduct supplementary analyses related to job quality; these findings are reported in a later section.

I timed data collection to correspond with graduation dates for the universities included in the study, to remove the possibility of red flags that could occur if applications were submitted long before or long after graduation. If applications were submitted too early, employers could dismiss the applicant for being unavailable to interview or start work. But if applications were submitted too late, employers could question the applicant's work ethic or assume the applicant had already been passed over for many other jobs.

The submission period in the spring and early summer of 2016 represented a reasonable timeframe for recent college graduates to be applying to jobs.[12]

The primary outcome for the audit study is whether the employer responded to the applicant via phone or email (a callback). Responses were coded as callbacks if the employer invited the applicant to an in-person interview or requested that the applicant contact the employer to discuss the position further. Form emails and acknowledgments of applications were not coded as callbacks.

Résumé Audit Study Results

Table 1 shows the overall results of the audit study across all experimental conditions. In this table and subsequent analyses, GPA is separated into four approximately equal-sized groups, which I found was the most parsimonious way to capture the full complexity of the observed patterns.[13] The overall callback rate was 12.9 percent—slightly higher than audit studies of professional workers (Correll et al. 2007; Rivera and Tilcsik 2016), but similar to what others have found in audit studies of recent college graduates in the
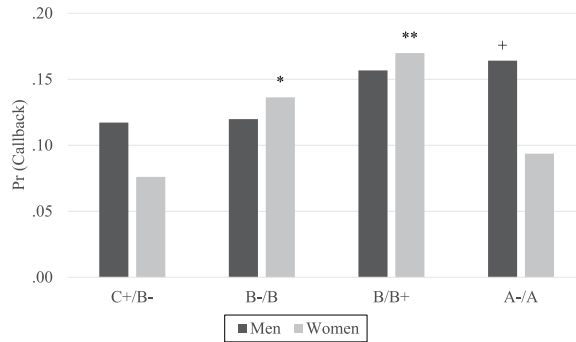
Figure 1. Callback Rates by Gender and Achievement, $N = 2,106$
*$p < .05$; **$p < .01$ (significantly different from baseline within gender).
+$p < .05$ (significantly different between gender).

entry-level labor market (Gaddis 2015). This overall callback rate does not differ by gender. Men were called back for 14.0 percent of job openings, and women were called back for 11.9 percent of job openings—not significantly different from each other ($p = .15$).

Despite the gender equality observed in the overall sample, Figure 1 shows that men's and women's callbacks were highly differentiated by achievement. Men were called back at approximately the same rate regardless of their GPA. In other words, the callback rate for men with the lowest GPAs was not statistically different than that for men with moderate or high GPAs. This is partly because the callback rate for men with the lowest GPAs was quite high. Although low-achieving men were called back only marginally more often than low-achieving women (11.7 versus 7.6 percent, $p < .10$, not shown), low-achieving men's callback rate was so high that the achievement effect for men was rendered non-significant. As mentioned earlier, some guidebooks caution applicants to not report their GPA if it falls below a certain threshold, but this pattern suggests men's outcomes may not be harmed if they violate this advice.

For women, the effect of achievement takes a different form. As women's achievement increases, they initially have higher callback rates than those at baseline. Women in the B–/B range were called back more often than those in the C+/B– range ($p < .05$), and women in the B/B+ range also had more

callbacks than those at baseline ($p < .01$). Yet, this achievement effect does not extend to women with the highest GPAs. In fact, the dropoff between women with B/B+ grades and A–/A grades is so steep that there is a significant difference in callbacks between these two groups ($p < .01$, not shown).

As a result of this inverted U-shaped effect of achievement for women, men with the highest grades were called back substantially more often than women with the highest grades—approximately 16 percent of the time for men, versus only 9 percent for women ($p < .05$). This gender gap is statistically significant and substantial: high-achieving men were called back nearly twice as often as their female counterparts. In sheer percentage terms, the highest-achieving women were called back even less often than the *lowest*-achieving men, although the point estimates for these groups are not statistically different.[14]

The audit study also points to meaningful patterns within majors. Because majors signal various competencies and gendered connotations, it may be reasonable to expect that men and women will have different employment outcomes depending on their achievement in different fields of study. The top panel in Figure 2 shows results for English majors only. When the sample is restricted to applicants who reported a major in English, neither gender nor achievement significantly affects callbacks. The general shape of results for English majors mimics that of the overall sample, but
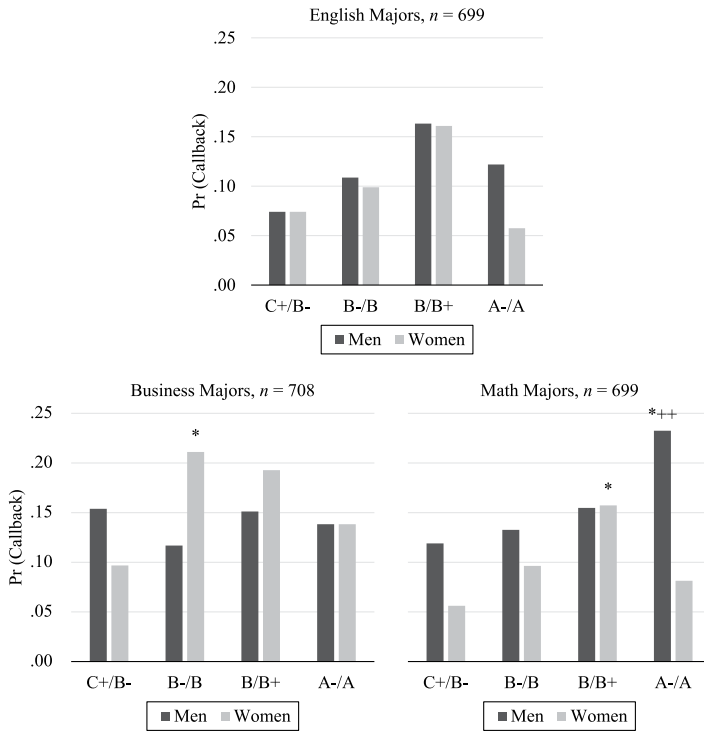
English Majors, *n* = 699

Business Majors, *n* = 708

Math Majors, *n* = 699

Figure 2. Callback Rates by Gender and Achievement, within Majors
Note: Gender does not predict callbacks for English and business majors; men math majors were called back significantly more often than women math majors (16 versus 10 percent, *p* < .05).
*p < .05 (significantly different from baseline within gender).
++p < .01 (significantly different between gender).

the sample size is considerably smaller, and thus the effects of applicant characteristics are not significant. The overall callback rate for English majors, 10.9 percent, is significantly lower than that for business majors and marginally lower than that for math majors (see Table 1). This pattern suggests that English majors are perceived as having few skills applicable to the entry-level jobs I submitted applications to. Employers may believe that English majors have less to contribute to their companies than do business or math majors, and applicants' gender and achievement may not shift that perception in any discernable direction.

The bottom-left panel in Figure 2 shows results for business majors only. Within each level of GPA, the callback rates for men and women do not differ. High-achieving men and women have an identical callback rate (13.8

percent), suggesting that high grades in business—and perhaps gender-neutral majors more broadly—are evaluated similarly for applicants of both genders. Men, moreover, were called back at approximately the same rate regardless of their grades. But when women had moderate grades—in the B–/B range—they were called back significantly more often than when they had low grades (*p* < .05). The fact that this achievement advantage does not extend to women business majors with higher grades suggests that employers gravitated toward women business majors with moderate grades but penalized women with higher achievement. This pattern contributes to the inverted U-shaped effect of achievement for women in the overall sample.

Finally, the bottom-right panel in Figure 2 shows results for math majors only. Men with the highest grades received more callbacks

than did those with the lowest grades ($p <$ .05). Math, therefore, is the only major that facilitates a positive effect of achievement (or any statistically significant effect of achievement, for that matter) for men. This may be because men who excel in math are viewed as exceedingly competent, whereas men who excel in English or business are not as well-regarded. Conversely, women with grades in the B/B+ range had more callbacks than did those at baseline ($p < .05$), but this advantage does not extend to the highest-achieving women math majors. As a result, high-achieving men math majors were called back significantly more often than high-achieving women math majors—only 8 percent for women, but three times as often, 24 percent, for men ($p < .01$). This gender gap among high-achieving math majors reinforces the gender gap among high achievers in the overall sample, and it suggests that high-achieving women were most readily penalized when they reported a major in math. In other words, when women demonstrate achievement in the precise field where they are expected to be least competent, they may be particularly likely to be penalized in hiring.

As a robustness check, I conducted analyses to assess how gender and achievement affect applicants' outcomes when they apply to jobs in industries that are most relevant to their major. A possible limitation of the audit study is that applications were submitted to positions across a range of industries—but in reality, students may only apply to jobs in industries that are closely aligned with their coursework. Results may thus be most realistic when they are limited to industries that are most relevant to applicants' majors.

Accordingly, I replicated the main analyses with the following combinations of majors and industries: math majors applying in the financial or information sectors, and business majors applying in the professional and business services sector. Although the sample sizes are relatively small ($n = 182$ and 149, respectively), results from these analyses mimic those of the overall sample. The financial and information sectors, in particular, are

male-dominated and involve tasks that tend to be viewed as masculine. The fact that high-achieving men (but not women) were routinely called back for positions in these industries suggests that the demand for skilled workers reinforces the industries' current demographics, rather than a shift toward women employees. For English majors, no one industry may be most relevant to students' coursework—and because there are no effects of gender or achievement within the sample of English majors, there are no empirical patterns to verify per se. However, I did not find effects of gender or achievement among English majors applying to any of the 10 industries examined in the audit study. These robustness checks confirm that achievement has very different effects for men and women—even for jobs that are closely matched to applicants' college majors.[15]

STUDY 2: SURVEY EXPERIMENT

The audit study demonstrates that college achievement has very different consequences for men and women in the entry-level labor market. For men, achievement has little to no effect on employment outcomes (with the exception of those who major in math). For women, in contrast, applicants with GPAs near the middle of the distribution receive more callbacks than both lower- and higher-achieving women. As a result of the steep penalty against high-achieving women, high-achieving men receive significantly more callbacks than do their female counterparts.

Despite this evidence of gendered returns to achievement, audit studies provide few clues as to *why* employers evaluate applicants the way they do. As discussed earlier, employers may subscribe to several perceptions that affect the way they evaluate gender and academic performance, including perceptions of men and women as more or less competent, likeable, and committed to their jobs. To understand how employers perceive applicants—and, in turn, how these perceptions shape hiring decisions—I conducted an online

survey experiment that replicated and expanded upon the audit study using a sample of 261 individuals who work as hiring decision-makers in their companies.

The research design for the survey experiment repeated many of the same components as the audit study. The résumés experimentally manipulated applicants' gender, achievement, and major in a 2 (male, female) × 3 (high, moderate, low) × 3 (English, mathematics, business) factorial design. The signals for gender and major were the same as those used in the audit study, but it was not feasible to use a continuous GPA variable in the survey experiment due to the relatively small sample size. Accordingly, I took the GPA categories from the audit study and chose round numbers near the medians of those categories to approximate high, moderate, and low GPAs for the survey experiment. These were 3.80 (mostly A's), 3.25 (near the median of the two moderate GPA categories; mostly B's), and 2.70 (mostly B–'s and C+'s).

Each respondent in the survey experiment was randomly assigned to evaluate two résumés from recent college graduates who were said to be seeking entry-level positions in their companies.[16] Unlike the audit study, which manipulated gender within subjects (i.e., each employer received one man's résumé and one woman's résumé), the gender manipulation in the survey experiment was between subjects (i.e., each respondent evaluated two men's or two women's résumés). I did this to make the gender manipulation less obvious, as respondents were less likely to intuit that gender was a main variable of interest if they did not encounter a gender manipulation themselves. This practice may also minimize social desirability bias related to the evaluation of women job applicants. Résumé format and résumé order were randomized and counterbalanced.

I worked with a professional survey firm, Qualtrics, to recruit and pay respondents. Qualtrics maintains several panels of respondents who represent different segments of the population and are regularly called on to take surveys for pay. The panel I used was composed of people with business expertise whose professional identities were verified through their LinkedIn profiles and calls to their employers. To ensure respondents were well-suited to answer questions from the perspective of a hiring decision-maker, I included two screening items designed to disqualify respondents if they did not meet certain criteria (adapted from Pedulla 2016): as part of their job, respondents must routinely make hiring decisions; and respondents must be either a human resources manager, human resources assistant/associate, business executive, mid-level manager, or business owner. Given that the fictional applicants were college graduates and most of the jobs in the audit study required a college degree, the sample was restricted to respondents with college degrees.[17,18]

Table 2 shows descriptive statistics for the 261 respondents in the survey experiment. Because respondents were not drawn from a random probability sample of human resources professionals (nor does such a sampling frame exist), these data cannot be generalized to a broader population. However, they can be used to produce internally valid, causal estimates of the effects of academic performance and gender on hiring decisions (Pedulla 2016).[19]

After reviewing each résumé, respondents were asked to evaluate the applicant and rate him or her on several dimensions. To provide a comparison between the audit study and the survey experiment, respondents were asked: "How likely would you be to recommend that your company interview (name)?" Responses were entered on an 11-point scale, from 0 ("not at all likely") to 10 ("very likely"). I transformed these responses into a binary variable representing those who were "very likely" to recommend the applicant for an interview. This outcome closely mirrors the outcome from the audit study, because applicants would only have received a callback if they achieved this "very likely" rating. The percentage of applicants who were rated as "very likely" to receive an interview in the survey experiment (9.6 percent) is comparable to the percentage of applicants who received callbacks in the audit study (12.9 percent).

Table 2. Descriptive Statistics for Survey Experiment Respondents and Firms, *N* = 261

	Sample Percent/ Mean
Respondents	
Gender	
Woman	57.1
Man	42.9
Race/Ethnicity	
White	78.5
Non-white	21.5
Income	$107,503.30
Age	40.0
Firms	
Number of employees	
Fewer than 10	10.7
Between 10 and 99	30.3
Between 100 and 499	22.6
500 or more	36.4
Industry	
Agriculture, mining, or construction	5.4
Education or health	21.8
Financial or information	11.9
Leisure or hospitality	7.7
Manufacturing	10.3
Professional or business services	19.9
Public administration	5.8
Retail	12.3
Transportation, utilities, or wholesale	4.2
Other	.8

Note: Percentages may not add to 100 due to rounding. Income variable was top-coded at the 99th percentile to minimize skew. Industry categories are adapted from Pedulla (2016).

Next, I asked respondents to provide their impressions of the applicant's personal traits. This portion of the survey is summarized in Table 3. The items were collapsed into five scales representing perceptions that have been shown to affect hiring decisions, including perceptions of the applicant's competence, likeability, commitment, social skills, and whether the applicant is a hard worker (Correll et al. 2007; Heilman 2001; Pedulla 2016; Rivera and Tilcsik 2016). Because

respondents had access only to applicants' résumés, it may seem as if they did not have enough information to make judgments about such things as applicants' social skills or commitment to their job. Other studies have shown, however, that résumés alone are sufficient to elicit strong assumptions about what an applicant is like in person, and these assumptions are tied to hiring decisions (Pedulla 2016; Rivera and Tilcsik 2016).

Finally, respondents were asked to report their overall impressions of the applicant in an open-ended format. The prompt asked, "What is your overall assessment regarding whether to interview (name) for an entry-level position? In your own words, please write a few sentences explaining why you feel this way." As discussed below, I coded these responses and compared them across experimental conditions to understand how respondents explained their hiring recommendations. Results generally show that when discussing applicants with different combinations of gender and achievement, respondents use very different rationales, and reference very different aspects of applicants' résumés, to justify their decisions.

Throughout these analyses, I collapse across major categories to yield six main experimental conditions (i.e., high-achieving woman, moderate-achieving woman, low-achieving woman, high-achieving man, moderate-achieving man, low-achieving man). Although I include major as an experimental condition in the survey design, cell sizes were typically insufficient to detect differences across major categories. Six respondents were dropped from the sample due to item non-response, but results are consistent when these respondents are included and mean-imputed. Results are also consistent without the control variables shown in Table 4.

Interview Recommendations and Ratings of Applicant Characteristics

I begin by examining factors that predict applicants' chances of being rated as "very likely" to be recommended for an interview. Table 4 shows these results, with separate

Table 3. Items Used to Assess Employer Perceptions of Applicants

Scale	Items
Competence (α = .96)	How do you think most people would view (name)? (1 = Not at all; 7 = Extremely) - Competent - Capable - Efficient - Organized - Skilled - Self-confident - Independent
Likeability (α = .94)	How do you think most people would view (name)? (1 = Not at all; 7 = Extremely) - Likeable - Good-natured - Sincere - Pleasant - Trustworthy - Warm - Aware of others' feelings - Aggressive (reverse-coded)
Hard work (α = .91)	How do you think most people would view (name)? (1 = Not at all; 7 = Extremely) - Productive - Hardworking
Commitment (α = .88)	How do you think most people would view (name)? (1 = Not at all; 7 = Extremely) - Committed Compared to similar employees who already work at your company, how committed do you think (name) would be to (her/his) job if (s/he) were hired? (1 = Not committed at all; 7 = Very committed) If your company needed to ask (name) to work extra hours, how likely is it that (name) would meet that request? (1 = Extremely unlikely; 7 = Extremely likely) If (name) were hired at your company, how long do you think (s/he) would stay? (1 = Less than 1 year; 5 = More than 4 years)
Social skills (α = .78)	How do you think most people would view (name)? (1 = Not at all; 7 = Extremely) - Has excellent social skills How well do you think (name) would get along socially with other employees who already work at your company? (1 = Not well at all; 7 = Extremely well)

Note: Items assessing competence and likeability were adapted from Fiske and colleagues (2002). Items assessing perceived commitment were adapted from Pedulla (2016).

models for men and women applicants. The base models assess the extent to which the audit study findings replicate in the survey experiment. The second set of models adds respondents' ratings of applicant characteristics to understand how their broader perceptions of applicants are related to the callbacks they assigned to men and women.

Model 1 examines the effects of achievement and major for men. Only one applicant characteristic is significant for men: men with high achievement were more likely than those with low achievement to be rated as "very likely" to

be recommended for an interview ($b = 1.302$, $p < .01$). This result diverges somewhat from the audit study. The audit study showed that achievement generally did not predict men's chances of receiving a callback, and only one group of men—math majors—benefited from having high achievement. Model 2 focuses exclusively on women applicants. Moderate-achieving women were more likely than low-achieving women to be recommended for an interview ($b = 1.101$, $p < .05$), but high achievers were statistically indistinguishable from low achievers. The survey experiment, therefore, replicates

Table 4. Effects of Achievement, Major, and Ratings of Applicant Characteristics on Chances of Being "Very Likely" to Be Recommended for an Interview

	Base Models		With Ratings of Applicant Characteristics	
	1. Men Applicants	2. Women Applicants	3. Men Applicants	4. Women Applicants
Achievement				
Moderate	.770	1.101*	−.382	1.383
	(.484)	(.480)	(1.284)	(.707)
High	1.302**	.809	−.528	.152
	(.464)	(.593)	(1.012)	(.771)
Major				
Business	.319	−.035	−.350	−.237
	(.421)	(.453)	(1.033)	(.786)
Math	.018	−.130	−.546	.268
	(.425)	(.559)	(.690)	(.666)
Applicant Characteristics				
Competence			7.553***	.861
			(2.042)	(.975)
Likeability			1.054	1.107*
			(1.252)	(.559)
Works hard			−.997	.473
			(2.210)	(.785)
Commitment			4.798*	−.270
			(1.896)	(1.413)
Social skills			−.482	1.676
			(.827)	(1.191)
Résumé presented second	.798*	.317	1.299	.683
	(.349)	(.334)	(1.276)	(.528)
Woman respondent	−1.017*	−.333	−.236	.130
	(.504)	(.605)	(.848)	(.667)
Non-white respondent	−.280	.498	−.566	.629
	(.651)	(.584)	(.989)	(.665)
Respondent income	.001	.005	−.010**	.007
	(.001)	(.004)	(.004)	(.004)
Respondent age	−.038	−.026	−.077	−.068*
	(.022)	(.017)	(.041)	(.033)
Constant	−1.388	−2.701*	−11.504**	−5.331*
	(1.762)	(1.074)	(3.778)	(2.293)
n (clusters)	137	124	137	124
n (observations)	274	248	274	248

Note: Logistic regressions; coefficients reported. Clustered standard errors are in parentheses. Omitted categories are low achievement, English major, résumé presented first, man respondent, and white respondent.
*$p < .05$; **$p < .01$ (two-tailed tests).

the inverted U-shaped effect of achievement for women found in the audit study.

Table 4 does not include tests for gender differences within levels of achievement. In the audit study, I found that high-achieving men were called back significantly more often than high-achieving women. In the survey experiment, however, I found no gender differences within achievement levels. One possible reason for this is that employers are less likely to differentiate between applicants in a survey, where they are asked about their

preferences directly, versus in an audit study, where their behaviors are observed. Survey participants might be hesitant to distinguish between applicants on the basis of gender or other status characteristics, even if they routinely do so in their jobs (Heerwig and McCabe 2009; Pager and Quillian 2005; Pedulla 2016). Another possibility is that gender is more salient in the audit study because employers are under more time pressure, and thus they use gender as a "shortcut" to determine who is most qualified (Fiske 1998). Most estimates indicate that employers spend only seconds reviewing a typical job application, because they must process dozens or even hundreds of résumés (Lahey and Beasley 2009). In the survey experiment, however, respondents were asked to review only two résumés, and they could familiarize themselves with the applicants' qualifications at their leisure, allowing gender to fade in importance. Additionally, the survey experiment had considerably less statistical power than the audit study. Overall, however, the shape of results for men and women suggests that achievement had similar effects across studies.

Models 3 and 4 incorporate respondents' ratings of applicant characteristics. These models assess how respondents' perceptions of applicants—including their perceptions of competence, likeability, commitment, social skills, and whether the applicant is a hard worker—account for the relationship between achievement and interview recommendations in the survey experiment. Model 3 shows results for men. This model reveals that two characteristics give men an advantage—perceptions of competence ($b = 7.553$, $p < .001$) and perceptions of commitment ($b = 4.798$, $p < .05$). When men are rated as highly competent or highly committed to their jobs, they are more often rated as "very likely" to be recommended for an interview. The effect of high achievement is no longer significant in this model, implying that perceptions of competence and commitment account for the relationship between achievement and callbacks.

Model 4 shows results for women. Here we see that women have better outcomes only when they are perceived as likeable ($b =$

1.107, $p < .05$). Other perceptions, including competence, commitment (both of which have a positive effect for men), working hard, and social skills, do not significantly affect women's outcomes. The positive effect of moderate achievement is no longer significant in this model, although the coefficient is larger than in the base model. This suggests that the effect of moderate achievement may have been crowded out, rather than accounted for with the addition of applicant characteristics.

This analysis provides key insight into why achievement has different effects among men and women applicants. For men, competence and commitment are the primary attributes that employers tend to reward. Because high-achieving men are perceived as highly competent and highly committed to their jobs, high-achieving men may experience advantages in the labor market that lower-achieving men cannot access. For women, however, employers tend to think about achievement and perceive applicants quite differently. Women's employment outcomes are not tied to perceptions of competence, working hard, or commitment to their jobs—the exact perceptions that are important for men's outcomes (in the case of competence and commitment) and that many people think of as important for workplace success. Rather, women's likeability is the primary perception that drives employment outcomes. This shifting standard for men and women helps explain why returns to academic performance are so distinctly gendered. High-achieving men may be sought-after because they possess the attributes that are typically ascribed to the ideal worker; but when women present the same qualifications, they may be passed over because they are perceived as not warm enough or not sincere enough—considerations that are virtually irrelevant when men's résumés are being evaluated.

Open-Ended Narratives

As a final component of the survey experiment, I examined respondents' open-ended narratives that summarized why they would or would not recommend applicants for an

interview. To code these narratives, I first created a spreadsheet for each combination of achievement and gender, yielding six total spreadsheets—often referred to as data matrices for qualitative data analysis (see Calarco 2018; Miles and Huberman 1994). Then, I copied the narratives in the leftmost column, and I recorded each time a respondent mentioned one of the attributes of interest in the survey experiment: competence, likeability, hard work, commitment, and social skills. Each instance was coded as positive, negative, or neutral, depending on how the respondent assessed the applicant's traits. For example, if a respondent made a comment about how the applicant appeared highly capable, this was coded as "competence–positive," but if a respondent made a disparaging comment about the applicant's ability to work in an efficient and independent manner, this was coded as "competence–negative." Neutral comments made reference to a trait without stating whether the applicant possessed that trait or not, such as mentioning they would need more information about the applicant's skills before deciding whether they would be a good fit for the company ("competence–neutral"). These neutral comments initially seemed unimportant, but I found they were used disproportionately in certain experimental conditions, and respondents may have used them to express negative sentiments about applicants in an indirect manner, as discussed below.

The narratives uncovered several patterns that help explain why respondents were more or less willing to interview particular applicants. Here, I focus on the three groups that provide the most insight into the patterns in the audit study and survey experiment: high-achieving women, moderate-achieving women, and low-achieving men. The quotes presented in the text are representative of the most salient themes that emerged for each of these groups, in that they capture how these groups were viewed in contrast to their relevant comparators.

When asked to comment on high-achieving women, respondents placed an overriding emphasis on likeability. Nearly a quarter (23 percent) of these respondents mentioned likeability, versus only 9 percent of respondents who were asked about high-achieving men. These comments about high-achieving women's likeability were not uniform in their directionality. For example, one respondent had positive things to say: "When I look at her résumé, I feel she may be a better fit than [the other applicant]. She has a passion for people." Another had a more negative take:

> Stephanie seems over-confident and very smart. She would be overqualified for any position in my company. Also, she doesn't quite seem socially warm. Not sure why, there's nothing wrong with being confident, but I get the feeling she's arrogant.

In other words, even though the applicant was seen as overqualified for every position in the respondent's company, these perceptions of arrogance and lack of warmth—based solely on the applicant's résumé—would prevent the applicant from being invited to an interview. This assessment of high-achieving women as overqualified is inherently positive, as it implies that the applicant is being passed over for being perceived as exceedingly capable and skilled. To further explore this perception of high-achieving women as overqualified—and whether this perception contributes to hiring penalties for high-achieving women—I present a set of supplementary analyses in the next section using both the audit study and survey experiment data.

Another group of respondents made neutral statements about how likeability was important, and said they would need more information about the high-achieving woman to make a judgment call about her personality. One respondent said, "Stephanie seems like a solid student, but I would need to see the personal side in the interview." Another said, "I would like to see how she communicates in person." And another remarked, "She certainly seems qualified, although her work history is somewhat limited. Based on our business model, a lot would be determined by

her personal skills." Although these comments are not explicitly gendered, the frequency with which they were mentioned for high-achieving women suggests they are somehow activated by gender. Perhaps employers are skeptical about high-achieving women's likeability, or they see these women as unlikeable but are unwilling to state it directly. Importantly, every respondent who made a neutral comment about likeability also indicated they would not recommend the applicant for an interview. These comments, therefore, may be most accurately described as negative assessments couched in neutral language.

When asked to describe moderate-achieving women, respondents had many positive things to say about their likeability and social skills. In both the audit study and the survey experiment, moderate-achieving women had some of the best employment outcomes across all experimental conditions. The open-ended narratives suggest this may be due to employers' belief that moderate-achieving women will be able to fill a particular social niche within their organizations. One respondent said, "She has excellent customer service skills, so that will be a plus in our business model." Notably, this respondent highlighted the applicant's experience in a customer-oriented role even though this aspect of the résumé was held constant across conditions. Another said, "She does seem to enjoy life, so she may be someone who enjoys other people and is curious, adventuresome, and seeks challenges." Another respondent made it clear that moderate-achieving women could play a distinct role on a team: "A real worker bee. Involved in projects where something gets done. She fits in and is a team player—not a wallflower."

Approximately 32 percent of respondents who were asked about moderate-achieving women mentioned likeability or social skills. Their comments suggest that when women have moderate achievement, they are perceived as competent *enough,* but not the *most* competent, which allows them to be perceived as more likeable than women with higher grades. Although moderate-achieving women receive a premium in hiring, these women's long-term career prospects are less promising. Because moderate-achieving women benefit as a result of their personalities—and not their ability—they may not achieve the same level of pay, responsibility, and general esteem as other workers, allowing subtle forms of gender inequality to persist.

Finally, when asked to comment on low-achieving men, respondents made a variety of excuses for men's poor grades. The audit study demonstrates that achievement does not have a significant effect on men's employment outcomes, in part because the callback rate for low-achieving men is quite high. The explanations respondents provided in the survey experiment suggest this may be due to employers' tendency to explain away men's low achievement. As one respondent said, "He is literally your average guy. He probably does his job just as he is told. Sometimes you have to give people the benefit of the doubt and go from there." Another said, "David's grades suck, but his experience seems ok. I would follow up with his past employer." Another respondent acknowledged the applicant's low grades, but decided to focus on other parts of the résumé when deciding whether to interview him:

> Could have done better academically, but was involved in school, and led a project at his internship. Could be motivated and become a very good employee. It appears that his current skill set is more in the human interaction area than his major of math.

By focusing on the applicant's apparent strengths instead of his weaknesses, this respondent effectively shifted their standards in order to rationalize a more favorable outcome for the applicant—a tendency that past research has highlighted (Uhlmann and Cohen 2005). About 21 percent of respondents made comments that acknowledged but mitigated men's low grades, versus 12 percent of respondents who were presented with a low-achieving

woman. This pattern suggests that low-achieving men may have better than expected labor market outcomes because employers search for other positive traits when men have poor academic performance.

ARE HIGH-ACHIEVING WOMEN PENALIZED BECAUSE THEY ARE VIEWED AS OVERQUALIFIED?

The previous analyses demonstrate that high-achieving women are penalized in the entry-level labor market, and that this penalty is based, at least in part, on employers' belief that these women do not meet the prevailing standard for likeability. But could these penalties also be rooted in more positive assessments of high-achieving women? One possibility is that high-achieving women are passed over for jobs because employers perceive them as overqualified. Employers may believe that high-achieving women are so sought-after that they will have many competing job offers, and they should not waste time on an applicant who would not reasonably be expected to accept their offer. High-achieving women who are perceived as overqualified would still be penalized in terms of callbacks, but these penalties would be tied to employers' perceptions of them as extremely desirable employees.

Both the audit study and the survey experiment data provide insight into whether high-achieving women are viewed as overqualified. I used the audit study data to conduct analyses presented by Deming and colleagues (2016) in their audit study of returns to for-profit and non-profit credentials. These analyses examine the relationships between achievement, gender, and job quality, using potential salary as a proxy for job quality. If high-achieving women are viewed as overqualified, then they might receive disproportionately *fewer* callbacks for lower-quality jobs, because these employers could not reasonably compete for them. But, high-achieving women might also receive *more* callbacks for higher-quality jobs that

offer the most competitive pay and benefits. Table 5 shows the effects of achievement and gender on applicants' chances of receiving a callback for jobs sorted by salary. The first two columns separate jobs into salary ranges: less than \$42,500 (the median for the sample), and \$42,500 or more. The third column uses the full sample and includes interactions between applicant characteristics and salary.[20]

Table 5 provides strong evidence against the notion that high-achieving women are viewed as overqualified. In Model 1, we see that high-achieving women are neither advantaged nor disadvantaged when it comes to lower-quality jobs. High-achieving women's chances of being called back for these positions are statistically indistinguishable from that of low-achieving men. Interestingly, both moderate-achieving men ($b = .704$, $p < .05$) and women ($b = .690$, $p < .05$) received a disproportionately high number of callbacks in this category, suggesting that the premiums associated with moderate achievement may be concentrated among lower-quality jobs.

Model 2 in Table 5 shows that high-achieving women are disadvantaged in the competition for lucrative positions. High-achieving women are less likely to receive callbacks than low-achieving men, even though their GPAs are upward of 1.5 grade-points higher on a 4.0 scale ($b = -.987$, $p < .05$). The only other group to be similarly penalized is low-achieving women ($b = -1.072$, $p < .05$). Finally, Model 3 indicates that high-achieving women experience a negative gradient in job quality in the full sample. The significant interaction term suggests that high-achieving women's callback rate decreases relative to low-achieving men as job quality increases ($b = -.313$, $p < .05$). These three models collectively show that high-achieving women are not viewed as overqualified. If anything, high-achieving women are even more penalized in hiring for higher-quality jobs than for lower-quality jobs, revealing a double disadvantage.

The survey experiment data reiterate the idea that high-achieving women are not passed over for jobs because they are viewed

Table 5. Logistic Regression Estimates for Effects of Gender and Achievement on Callbacks, by Job Quality

Achievement/Gender	1. Less than $42,500	2. $42,500 and Greater	3. All Jobs
Moderate Man	.704*	−.412	1.012*
	(.343)	(.338)	(.495)
× Salary (in $10,000s)			−.203
			(.104)
High Man	.713	.094	.515
	(.392)	(.358)	(.552)
× Salary (in $10,000s)			−.033
			(.111)
Low Woman	−.003	−1.072*	.312
	(.412)	(.454)	(.633)
× Salary (in $10,000s)			−.209
			(.136)
Moderate Woman	.690*	−.072	.746
	(.311)	(.292)	(.411)
× Salary (in $10,000s)			−.105
			(.083)
High Woman	.416	−.987*	1.042
	(.389)	(.456)	(.609)
× Salary (in $10,000s)			−.313*
			(.138)
Application fixed effects?	Yes	Yes	Yes
n (clusters)	505	548	1,053
n (observations)	1,010	1,096	2,106

Note: Logistic regressions; coefficients reported. Clustered standard errors are in parentheses. Achievement was sorted into three groups: low (2.50 to 2.83 GPA), moderate (2.84 to 3.59 GPA), and high (3.60 to 3.95 GPA). Omitted category is low-achieving man. Models include controls for major, an indicator variable for whether the expected salary was imputed, order of résumé submission, résumé template, and work and education template. Model 3 also includes a control for the main effect of salary. "× Salary" is an interaction of the variable above that line times the expected salary for the job (based on the median salary for the job). The first two columns separate the sample into expected salary ranges: less than $42,500 (the median for the sample), and $42,500 or more. The third column includes the entire expected salary range.
*$p < .05$ (two-tailed tests).

as overqualified. For example, if an applicant were perceived as overqualified, then employers might be wary of that applicant trying to leave the company as soon as a better position opened up elsewhere. But when asked how long they expected applicants to stay at the company if hired, respondents assigned high-achieving women the highest mean score of more than two to three years—significantly higher than low-achieving women ($p < .01$, not shown) but not statistically different from other applicants. Further, the open-ended narratives contained only two instances of high-achieving women being described as overqualified, one of which was mentioned in the earlier results. As another respondent explained, "I think Stephanie is more of a person who would want to be her own boss, and it appears she's headed in that direction." This respondent indicated they would not recommend the applicant for an interview because they viewed her as too ambitious.[21] Yet, this line of reasoning was relatively uncommon, as far more respondents relied on their perceptions of likeability and social skills to pass over high-achieving women.

In summary, there are many reasons why high-achieving women could be penalized in the labor market. Some explanations—such as holding applicants to gendered standards of competence and likeability—imply that high-achieving women are passed over for jobs because they do not meet employers' expectations. Other explanations are considerably more positive. It may be, for example, that high-achieving women are perceived as overqualified, implying that these women are passed over for jobs because they exceed employers' expectations to the extent that they are viewed as unattainable. Although high-achieving women are often described as in-demand, particularly in sectors where women are underrepresented (Correll 2016; Tam 2016), these data suggest that most employers who pass over high-achieving women do not view those women as overqualified. The more plausible explanations are less complimentary toward women's achievement.

DISCUSSION

Using data from an audit study conducted with entry-level employers, I show that men's and women's academic performance have very different consequences in the labor market. I find that employers penalize women—but not men—for signaling strong academic performance on their résumés. Achievement bears little relation to men's employment outcomes, but women experience an inverted U-shaped effect of achievement, such that women with the highest grades are disproportionately penalized. The callback rate for high-achieving men, as a result, is nearly double that of high-achieving women.

Yet this penalty for high achievement does not apply equally to women in all fields of study. Of the majors I examined, only women in math were penalized, whereas high-achieving women in business and English did not experience a significant penalty. The callback rate for high-achieving men math majors was triple that of high-achieving women math majors. This finding highlights the barriers women face in STEM fields. Previous

research shows that STEM classrooms can be inhospitable to women (Jacobs 1996), and graduating from college with a STEM degree is an unlikely outcome in and of itself. But when women exhibit high achievement in the precise fields where they are expected to perform poorly, they may be particularly unlikely to be rewarded. Although universities have designed many programs to increase women's enrollment in STEM, these findings imply that STEM achievement is unlikely to help women advance in the labor market as long as employers continue to penalize this group. Over time, we might expect these penalties to diminish as more women enter and succeed in STEM majors. These penalties reflect an expectation that only men are capable of excelling in STEM fields, and that women who receive consistently high grades in math classes are violating gendered prescriptive norms. But as these fields become more integrated, and the association between gender and success in STEM breaks down, it follows that the labor market outcomes associated with STEM achievement should reach parity for men and women college graduates. This is by no means an easy task, given the durability of gendered expectations and enrollment patterns in STEM (Charles and Bradley 2002; Correll 2001, 2004; Riegle-Crumb et al. 2012), but it is extremely important if women are to receive the same returns as their male counterparts.

The survey experiment highlights some potential mechanisms that explain why employers think about achievement and gender the way they do. The quantitative portion of the survey experiment shows that employers shift their standards to reward different perceptions among men and women applicants. Men are more likely to be called back if they are perceived as competent and committed to their jobs—traits that are typically ascribed to the "ideal worker." Women, however, are more likely to be called back if they are perceived as likeable—an assessment that is more or less irrelevant to men's employment outcomes. The qualitative data reveal that while moderate-achieving women are

often viewed as likeable and socially skilled, employers are more skeptical about high-achieving women's personalities. These negative perceptions of high achievers contribute to the inverted U-shaped effect of achievement for women. Further, the audit study and survey experiment together rule out the explanation that high-achieving women are passed over for jobs because they are viewed as overqualified. This positive take on women's penalty for high achievement, which one might expect given employers' incentives to promote gender equality, is countered by findings that suggest employers view women with high grades in a negative light.

The concept of gendered employment penalties is not new, as scholars have highlighted perceptions that disproportionately penalize women in hiring and the workplace more generally (Acker 1990; Blair-Loy 2003; Correll et al. 2007; Rivera and Tilcsik 2016; Williams et al. 2012). What is new here, however, is the notion that academic performance elicits gendered labor market penalties. Many college students and even scholars assume that "grades don't matter"—or, if anything, that academic performance has a uniformly positive effect on employment outcomes. But this study demonstrates that high achievement activates gendered stereotypes that hurt women's chances of advancing to the interview stage with employers. Two implications of these findings are particularly notable, given the demography of gender and attainment in the United States today.

First, because many women earn high grades in college, a large number of women are potentially affected by the labor market penalties documented in this article. This is likely attenuated by the fact that women's grades tend to be lower in STEM fields (Conger and Long 2010), and much of the observed penalty for high achievement is concentrated among high-achieving women math majors. It may also be the case that high-achieving women have better outcomes than this study would predict because they do not often compete against high-achieving men for jobs. Because most jobs (and, presumably, most

applicant pools) are sex-segregated, high-achieving women are most likely to compete against other women for entry-level positions. Yet, an important finding of this study is that high-achieving women are penalized relative to both cross- *and* same-sex comparison groups. High-achieving women, in other words, are penalized not just relative to high-achieving men, but also relative to moderate-achieving women. The sheer breadth of these disadvantages suggests that high-achieving women may be penalized in many applicant pools in one way or another.

Second, these penalties are observed at labor market entry—the point where gender gaps in employment are at their smallest (Bobbitt-Zeher 2007; Marini and Fan 1997). Consistent with extant research, I find no significant differences in callback rates for men and women overall; but once academic performance is taken into account, large and consequential gender penalties emerge. Prior research on entry-level workers has thus overlooked an important dimension of gendered labor market stratification by neglecting to consider prior achievement. There is some evidence to suggest that these penalties may be mitigated as job tenure increases. Most guidebooks advise applicants to not report their GPA after they have accumulated work experience, so academic performance may not play a role for applicants who have already held a full-time job. But because achievement has such profound effects among entry-level employers, as this article shows, workers' trajectories may be shaped by their college records. Research also indicates that achievement is enhanced by certain noncognitive skills, such as being organized, meeting deadlines, and following rules (DiPrete and Buchmann 2013)—skills that have the capacity to cast women as unlikeable because they are perceived as uptight or "bitchy" (Williams et al. 2012). To the extent that high-achieving women exhibit these traits in the workplace, they may be penalized throughout their careers for possessing noncognitive skills that tend to be correlated with academic performance.

Future research can extend these findings to investigate other contexts where gender and achievement jointly affect student outcomes. Although I find that women are penalized for high achievement in entry-level employment, academic performance may certainly pay off for women applying to graduate and professional school, where achievement is a primary criterion for admission (Mullen, Goyette, and Soares 2003). College achievement, in this sense, may provide women with additional avenues into elite and high-wage jobs. Achievement may also have different effects for students whose highest level of education is a high school diploma, although research shows that women have fewer opportunities for blue-collar work than do men, which affects the relationship between gender and self-selection into college versus work (Sutton, Bosky, and Muller 2016).

Research can also look more explicitly at how men's and women's achievement affects hiring outcomes in heavily gendered occupations, such as nursing or computer programming. Because the demand for men and women workers varies widely depending on the job (Yavorsky 2017), employers might attach special premiums to achievement when there is alignment between the applicant's gender, the field of study, and the specific job opening. Although I applied to a wide range of jobs to understand how achievement and gender are evaluated in the labor market broadly, studies can hone in on certain types of jobs to consider an additional layer of gendered expectations.

Further, because the applicants in this study may have been perceived as white, research is needed to determine whether employers evaluate academic performance differently for men and women who belong to other racial groups. A long line of research shows, for example, that white and black applicants are called back at different rates, and that moderating variables (e.g., criminal record, elite college degree) have very different effects depending on the applicant's race (Bertrand and Mullainathan 2004; Gaddis 2015; Pager 2003). Studies have also demonstrated that white and black individuals are held to different standards for likeability (Doan 2016; van Ryn and Burke 2000). Black women may be held to an even stricter standard than white women when it comes to demonstrating warmth, implying that black women may experience a steep penalty for high achievement as they make the transition from college to work.

Women made great strides over the course of the twentieth century, demonstrating strong achievement across levels of education and entering the workforce in record numbers. But despite this progress, employers may penalize women who signal high achievement on their résumés. This article demonstrates the many and varied penalties high-achieving women face in the entry-level labor market, as well as the gendered stereotypes that allow these penalties to persist. Although women have made many advances in higher education, further change is needed for women to make comparable advances at work.

APPENDIX

Table A1. Summary of Applications Submitted in Audit Study, *N* = 2,106

	Frequency	Percent
Achievement		
C+/B– average (2.50 to 2.83 GPA)	519	24.6
B–/B average (2.84 to 3.20 GPA)	531	25.2
B/B+ average (3.21 to 3.59 GPA)	527	25.0
A–/A average (3.60 to 3.95 GPA)	529	25.1
Gender		
Woman	1,053	50.0
Man	1,053	50.0
Major		
Business	708	33.6
English	699	33.2
Mathematics	699	33.2
Region		
South	376	17.9
Midwest	650	30.9
Southwest	382	18.1
West	368	17.5
Northeast	330	15.7
Industry		
Agriculture, mining, or construction	154	7.3
Education	36	1.7
Financial or information	570	27.1
Health	244	11.6
Leisure or hospitality	70	3.3
Manufacturing	140	6.7
Professional or business services	456	21.7
Public administration	42	2.0
Retail	112	5.3
Transportation, utilities, or wholesale	282	13.4
Job type		
Administrative assistant	320	15.2
Analyst	448	21.3
Clerical	98	4.7
Customer service	212	10.1
Human resources	154	7.3
Managerial	134	6.4
Other—kids	20	1.0
Other—physical	134	6.4
Other	52	2.5
Sales	534	25.4

Note: Percentages may not add to 100 due to rounding. Industry categories are adapted from Pedulla (2016). Job type categories are adapted from Gaddis (2015).

Acknowledgments

Notes

1. Throughout this article, I use "academic performance" and "achievement" to refer to grades, and "attainment" to refer to a person's highest degree received or level of education. This article focuses on differences in academic performance among job applicants with the same attainment.

2. Both of these perspectives imply that men and women will have equal returns to achievement in sex-neutral majors—a possibility I explore in this study.

3. The names differed across regions because names vary in popularity in different parts of the country, and I wanted names to be ubiquitous to employers in the area. This also served a practical purpose, as it was more efficient to match callbacks to job applications when they were funneled into 10 smaller accounts rather than two large accounts.

4. Some studies have used names to signal social class (e.g., Gaddis 2015; Rivera and Tilcsik 2016). Although I did not purposefully invoke class, the names I used were so common that employers may have interpreted the applicants as middle-class by default.

5. I considered including additional signals of achievement, such as awards or Latin honors (e.g., cum laude). Although students with high grades would normally have received these awards and listed them on their résumés, it would be atypical for students with low grades to have received any honors in college (even lukewarm awards like "perfect attendance" or "most improved," which do not necessarily signal achievement but do signal some amount of diligence). If I were to include awards on some résumés but not others, the résumés would not be equivalent in length and style across experimental conditions and thus would not be as comparable. One potential consequence of this choice is that employers penalized high-achieving applicants because they had fewer awards than high achievers typically do, but this cannot be tested with the data.

6. An additional consideration is that cumulative GPAs are likely to be more salient, and more often reported, among students who were continually enrolled in higher education and transferred schools infrequently or not at all. The students who meet these criteria are disproportionately middle-class (Goldrick-Rab 2006; Goldrick-Rab and Pfeffer 2009). Thus, applicants may be viewed favorably for reporting a cumulative GPA on their résumé, even if that GPA is low, because reporting a GPA operates as a subtle signal of social class.

7. These universities are large enough that it would not be unusual for employers to receive applications from two students within a few days of each other. I chose to have both applicants attend the same school to ensure college selectivity was held constant within regions.

8. To conduct the pretests, I asked undergraduates to rate a list of activities in terms of masculinity, femininity, and prestige on a 0 to 10 scale. I chose activities that fell in the middle on all three dimensions, suggesting that the activities were perceived as gender-neutral and moderately prestigious. I used a similar procedure to select majors.

9. Some openings linked to the employer's website, where the applicant was expected to provide additional information, such as essay questions covering the applicant's career goals and reasons for applying to the position. Scholars have noted difficulty in standardizing answers to these questions so as to not obscure the experimental manipulations (Gaddis 2015; Pedulla 2016). To avoid this issue, these jobs were screened out of the sampling frame.

10. When the sample is limited to jobs that explicitly require a college degree, results are consistent with those shown here. I present results with the full sample because it maximizes the number of cases available for analysis and facilitates comparisons across majors and industries.

11. In pilot testing, I identified some spam openings that were being used as phishing scams to collect applicant information. These generally consisted of commission-only jobs in marketing firms (for discussions, see Deming et al. 2016; Mishel 2016). These jobs were screened out of the sampling frame because all applicants were called back regardless of their qualifications. During data collection, I dropped companies from the sample if they spammed applicants or if their emails were flagged as having been sent from bots.

12. Because some applications were submitted before the applicant's supposed graduation date, and some were submitted after, I varied the text of the cover letter slightly to reflect the applicant's enrollment status. Before graduation, I presented the applicant as "preparing to graduate" from the university; after graduation, I presented the applicant as a "recent graduate" of the university. Significance tests revealed no differences in callbacks between applicants who had and had not graduated.

13. Results are consistent when GPA is analyzed as a continuous variable with an additional squared term.

14. I found few regional differences in the sample, with one exception: the overall callback rate was higher in the Southwest than in the Midwest or Northeast (but not significantly different from the South or West). Employers in the Southwest were particularly likely to call back low-achieving men.

15. I also conducted analyses that matched applicants on the basis of majors and job types—for example, math majors applying to analyst positions, business majors applying to sales positions. Due to the small cell sizes for some job types (see Appendix Table A1), many of these models had convergence issues, but results that could be generated are consistent with those presented here. I consider this issue further in the Discussion section.

16. The cover story for the survey experiment was intentionally broad so it would be applicable to as many respondents as possible. I did not want to limit the industries in the survey experiment, because the industries in the audit study were broad; similarly, I did not want to have respondents evaluate résumés for a specific position (e.g., salesperson), because respondents would not necessarily be able to provide expertise related to that position. Thus, I chose to provide a general cover story.

17. I collected a separate set of responses from respondents with less education to assess how education level shapes perceptions of college achievement. Generally, non-college-educated respondents responded positively to achievement signals and were less sensitive to gender when making hiring recommendations; these findings will be presented in a follow-up article.

18. Respondents were compensated at different rates depending on their job description and how they were recruited into the panel. I paid $15 per respondent for an approximately eight-minute survey, and a portion of that amount was used to compensate respondents.

19. See Pedulla (2016) for estimated demographics for the population of hiring decision-makers (note that my sample differs because it is limited to college degree-holders and has a different composition of occupations). Generally speaking, my respondents had higher incomes and worked at larger companies than the average hiring decision-maker.

20. About 40 percent of job advertisements listed a starting salary or salary range. For jobs that posted a salary range, I assigned a salary equal to the median of that range. For jobs with missing salary data, I imputed the median salary for the job type (see Appendix Table A1) and region associated with that job. Ten salaries could not be imputed using the job type and region, so I imputed the median salary for the job type across all five regions.

21. Both respondents who described high-achieving women as overqualified found fault with her personality, consistent with research showing that women are often perceived as bossy or "bitchy" in the workplace (Williams et al. 2012). This sug-

gests that the perception of high-achieving women as overqualified is ultimately a perception of high-achieving women as unlikeable.

References

Acker, Joan. 1990. "Hierarchies, Jobs, Bodies: A Theory of Gendered Organizations." *Gender & Society* 4(2):139–58.

Armstrong, Elizabeth A., and Laura T. Hamilton. 2013. *Paying for the Party: How College Maintains Inequality*. Cambridge, MA: Harvard University Press.

Arum, Richard, and Josipa Roksa. 2011. *Academically Adrift: Limited Learning on College Campuses*. Chicago: University of Chicago Press.

Arum, Richard, and Josipa Roksa. 2014. *Aspiring Adults Adrift: Tentative Transitions of College Graduates*. Chicago: University of Chicago Press.

Becker, Gary S. 1964. *Human Capital*. New York: National Bureau of Economic Research.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94(4):991–1013.

Bills, David B. 2003. "Credentials, Signals, and Screens: Explaining the Relationship between Schooling and Job Assignment." *Review of Educational Research* 73(4):441–69.

Bills, David B., Valentina Di Stasio, and Klarita Gërxhani. 2017. "The Demand Side of Hiring: Employers in the Labor Market." *Annual Review of Sociology* 43:291–310.

Blair-Loy, Mary. 2003. *Competing Devotions: Career and Family among Women Executives*. Cambridge, MA: Harvard University Press.

Bobbitt-Zeher, Donna. 2007. "The Gender Income Gap and the Role of Education." *Sociology of Education* 80(1):1–22.

Brand, Jennie E., and Yu Xie. 2010. "Who Benefits Most from College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education." *American Sociological Review* 75(2):273–302.

Brown, David K. 2001. "The Social Sources of Educational Credentialism: Status Cultures, Labor Markets, and Organizations." *Sociology of Education* 74(1):19–34.

Buchmann, Claudia, and Thomas A. DiPrete. 2006. "The Growing Female Advantage in College Completion: The Role of Family Background and Academic Achievement." *American Sociological Review* 71(4):515–41.

Buchmann, Claudia, Thomas A. DiPrete, and Anne McDaniel. 2008. "Gender Inequalities in Education." *Annual Review of Sociology* 34:319–37.

Calarco, Jessica McCrory. 2018. *Negotiating Opportunities: How the Middle Class Secures Advantages in School*. New York: Oxford University Press.

Carbonaro, William, Brandy J. Ellison, and Elizabeth Covay. 2011. "Gender Inequalities in the College Pipeline." *Social Science Research* 40(1):120–35.

Charles, Maria, and Karen Bradley. 2002. "Equal but Separate? A Cross-National Study of Sex Segregation in Higher Education." *American Sociological Review* 67(4):573–99.

Charles, Maria, and Karen Bradley. 2009. "Indulging Our Gendered Selves? Sex Segregation by Field of Study in 44 Countries." *American Journal of Sociology* 114(4):924–76.

Collins, Randall. 1971. "Functional and Conflict Theories of Educational Stratification." *American Sociological Review* 36(6):1002–19.

Collins, Randall. 1979. *The Credential Society: An Historical Sociology of Education and Stratification*. New York: Academic Press.

Conger, Dylan, and Mark C. Long. 2010. "Why Are Men Falling Behind? Gender Gaps in College Performance and Persistence." *ANNALS of the American Academy of Political and Social Science* 627(1):184–214.

Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessments." *American Journal of Sociology* 106(6):1691–1730.

Correll, Shelley J. 2004. "Constraints into Preferences: Gender, Status, and Emerging Career Aspirations." *American Sociological Review* 69(1):93–113.

Correll, Shelley J. 2016. "Inside the Black Box of Organization Life: The Gendered Language of Performance Assessment." Presented at Indiana University-Bloomington.

Correll, Shelley J., Stephen Benard, and In Paik. 2007. "Getting a Job: Is There a Motherhood Penalty?" *American Journal of Sociology* 112(5):1297–1339.

Cuddy, Amy J. C., Susan T. Fiske, and Peter Glick. 2008. "Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map." *Advances in Experimental Social Psychology* 40:61–149.

Damaske, Sarah. 2011. *For the Family? How Class and Gender Shape Women's Work*. New York: Oxford University Press.

Deming, David J., Noam Yuchtman, Amira Abulafi, Claudia Goldin, and Lawrence F. Katz. 2016. "The Value of Postsecondary Credentials in the Labor Market: An Experimental Study." *American Economic Review* 106(3):778–806.

DiMaggio, Paul, and Filiz Garip. 2012. "Network Effects and Social Inequality." *Annual Review of Sociology* 38:93–118.

DiPrete, Thomas A., and Claudia Buchmann. 2013. *The Rise of Women: The Growing Gender Gap in Education and What It Means for American Schools*. New York: Russell Sage.

DiPrete, Thomas A., and Jennifer L. Jennings. 2012. "Social and Behavioral Skills and the Gender Gap in Early Educational Achievement." *Social Science Research* 41(1):1–15.

Doan, Long. 2016. "How Do Racial and Gender Beliefs Affect Emotions?" PhD dissertation, Indiana University, Bloomington, IN.

Downey, Douglas B., and Anastasia S. Vogt Yuan. 2005. "Sex Differences in School Performance during High School: Puzzling Patterns and Possible Explanations." *Sociological Quarterly* 46(2):299–321.

Dumais, Susan A. 2002. "Cultural Capital, Gender, and School Success: The Role of Habitus." *Sociology of Education* 75(1):44–68.

Dwyer, Rachel E., Laura McCloud, and Randy Hodson. 2012. "Debt and Graduation from American Universities." *Social Forces* 90(4):1133–55.

Eagly, Alice H., and Linda L. Carli. 2007. *Through the Labyrinth: The Truth about How Women Become Leaders*. Boston: Harvard Business School Press.

England, Paula, and Su Li. 2006. "Desegregation Stalled: The Changing Gender Composition of College Majors, 1971–2002." *Gender & Society* 20(5):657–77.

Entwisle, Doris R., Karl L. Alexander, and Linda S. Olson. 2007. "Early Schooling: The Handicap of Being Poor and Male." *Sociology of Education* 80(2):114–38.

Fiske, Susan T. 1998. "Stereotyping, Prejudice, and Discrimination." Pp. 357–411 in *Handbook of Social Psychology,* edited by D. Gilbert, S. T. Fiske, and G. Lindzey. Boston: McGraw-Hill.

Fiske, Susan T., Amy J. C. Cuddy, Peter Glick, and Jun Xu. 2002. "A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow from Perceived Status and Competition." *Journal of Personality and Social Psychology* 82(6):878–902.

Foschi, Martha, Larissa Lai, and Kirsten Sigerson. 1994. "Gender and Double Standards in the Assessment of Job Applicants." *Social Psychology Quarterly* 57(4):326–39.

Gaddis, S. Michael. 2015. "Discrimination in the Credential Society: An Audit Study of Race and College Selectivity in the Labor Market." *Social Forces* 93(4):1451–79.

Gaddis, S. Michael, ed. 2018. *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*. New York: Springer.

Gerson, Kathleen. 2010. *The Unfinished Revolution: Coming of Age in a New Era of Gender, Work, and Family*. Oxford, UK: Oxford University Press.

Goldrick-Rab, Sara. 2006. "Following Their Every Move: An Investigation of Social-Class Differences in College Pathways." *Sociology of Education* 79(1):67–79.

Goldrick-Rab, Sara, and Fabian T. Pfeffer. 2009. "Beyond Access: Explaining Socioeconomic Differences in College Transfer." *Sociology of Education* 82(2):101–125.

Goyette, Kimberly A. 2008. "College for Some to College for All: Social Background, Occupational Expectations, and Educational Expectations Over Time." *Social Science Research* 37(2):461–84.

Griffin, Larry J., Arne L. Kalleberg, and Karl L. Alexander. 1981. "Determinants of Early Labor Market Entry and Attainment: A Study of Labor Market Segmentation." *Sociology of Education* 54(3):206–221.

Hamilton, Laura T. 2013. "More Is More or More Is Less? Parental Financial Investments during College." *American Sociological Review* 78(1):70–95.

Hamilton, Laura T. 2016. *Parenting to a Degree: How Family Matters for College Women's Success*. Chicago: University of Chicago Press.

Heerwig, Jennifer A., and Brian J. McCabe. 2009. "Education and Social Desirability Bias: The Case of a Black Presidential Candidate." *Social Science Quarterly* 90(3):674–86.

Heilman, Madeline E. 2001. "Description and Prescription: How Gender Stereotypes Prevent Women's Ascent Up the Organizational Ladder." *Journal of Social Issues* 57(4):657–74.

Hout, Michael. 2012. "Social and Economic Returns to College Education in the United States." *Annual Review of Sociology* 38:379–400.

Jacobs, Jerry A. 1996. "Gender Inequality and Higher Education." *Annual Review of Sociology* 22:153–85.

Johnston, Jerome, and Jerald G. Bachman. 1973. *The Transition from High School to Work*. Ann Arbor, MI: Institute for Social Research.

Kelly, Erin L., Samantha K. Ammons, Kelly Chermack, and Phyllis Moen. 2010. "Gendered Challenge, Gendered Response: Confronting the Ideal Worker Norm in a White-Collar Organization." *Gender & Society* 24(3):281–303.

Kim, ChangHwan, Christopher R. Tamborini, and Arthur Sakamoto. 2015. "Field of Study in College and Lifetime Earnings in the United States." *Sociology of Education* 88(4):320–39.

Lahey, Joanna N., and Ryan A. Beasley. 2009. "Computerizing Audit Studies." *Journal of Economic Behavior & Organization* 70(3):508–514.

Lin, Nan. 1999. "Social Networks and Status Attainment." *Annual Review of Sociology* 25:467–87.

Marini, Margaret Mooney, and Pi-Ling Fan. 1997. "The Gender Gap in Earnings at Career Entry." *American Sociological Review* 62(4):588–604.

Miles, Matthew B., and A. Michael Huberman. 1994. *Qualitative Data Analysis,* 2nd ed. Thousand Oaks, CA: Sage.

Miller, Shazia Rafiullah. 1998. "Shortcut: High School Grades as a Signal of Human Capital." *Educational Evaluation and Policy Analysis* 20(4):299–311.

Mishel, Emma. 2016. "Discrimination against Queer Women in the U.S. Workforce: A Résumé Audit Study." *Socius* 2.

Mullen, Ann L., Kimberly A. Goyette, and Joseph A. Soares. 2003. "Who Goes to Graduate School? Social and Academic Correlates of Educational Continuation after College." *Sociology of Education* 76(2):143–69.

Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108(5):937–75.

Pager, Devah. 2007. "The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future." *ANNALS of the American Academy of Political and Social Science* 609(1):104–133.

Pager, Devah, and Lincoln Quillian. 2005. "Walking the Talk? What Employers Say Versus What They Do." *American Sociological Review* 70(3):355–80.

Pager, Devah, Bruce Western, and Bart Bonikowski. 2009. "Discrimination in a Low-Wage Labor Market: A Field Experiment." *American Sociological Review* 74(5):777–99.

Pedulla, David S. 2016. "Penalized or Protected? Gender and the Consequences of Nonstandard and Mismatched Employment Histories." *American Sociological Review* 81(2):262–89.

Pedulla, David S., and Sarah Thébaud. 2015. "Can We Finish the Revolution? Gender, Work-Family Ideals, and Institutional Constraint." *American Sociological Review* 80(1):116–39.

Perkins, Robert, Brian Kleiner, Stephen Roey, and Janis Brown. 2004. *The High School Transcript Study*. Washington, DC: National Center for Education Statistics.

Phelan, Julie E., Corinne A. Moss-Racusin, and Laurie A. Rudman. 2008. "Competent Yet Out in the Cold: Shifting Criteria for Hiring Reflect Backlash toward Agentic Women." *Psychology of Women Quarterly* 32(4):406–413.

Quadlin, Natasha Yurk. 2016. "Gender and Time Use in College: Converging or Diverging Pathways?" *Gender & Society* 30(2):361–85.

Quadlin, Natasha Yurk, and Daniel Rudel. 2015. "Responsibility or Liability? Student Loan Debt and Time Use in College." *Social Forces* 94(2):589–614.

Ridgeway, Cecilia L. 2011. *Framed by Gender: How Gender Inequality Persists in the Modern World*. New York: Oxford University Press.

Ridgeway, Cecilia L., and Shelley J. Correll. 2004. "Motherhood as a Status Characteristic." *Journal of Social Issues* 60(4):683–700.

Riegle-Crumb, Catherine. 2010. "More Girls Go to College: Exploring the Social and Academic Factors Behind the Female Postsecondary Advantage among Hispanic and White Students." *Research in Higher Education* 51(6):573–93.

Riegle-Crumb, Catherine, Barbara King, Eric Grodsky, and Chandra Muller. 2012. "The More Things Change, the More They Stay the Same? Prior Achievement Fails to Explain Gender Inequality in Entry Into STEM College Majors Over Time." *American Educational Research Journal* 49(6):1048–73.

Rivera, Lauren A., and András Tilcsik. 2016. "Class Advantage, Commitment Penalty: The Gendered Effect of Social Class Signals in an Elite Labor Market." *American Sociological Review* 81(6):1097–1131.

Roksa, Josipa, and Tania Levey. 2010. "What Can You Do with That Degree? College Major and Occupational

Status of College Graduates over Time." *Social Forces* 89(2):389–415.

Rosenbaum, James E. 2001. *Beyond College for All: Career Paths for the Forgotten Half*. New York: Russell Sage.

Rosenbaum, James E., and Takehiko Kariya. 1991. "Do School Achievements Affect the Early Jobs of High School Graduates in the United States and Japan?" *Sociology of Education* 64(2):78–95.

Rosin, Hanna. 2012. *The End of Men: And the Rise of Women*. New York: Penguin.

Rudman, Laurie A. 1998. "Self-Promotion as a Risk Factor for Women: The Costs and Benefits of Counterstereotypical Impression Management." *Journal of Personality and Social Psychology* 74(3):629–45.

Ryan, Camille L., and Julie Siebens. 2012. *Educational Attainment in the United States: 2009*. Washington, DC: United States Census Bureau.

Stiglitz, Joseph E. 1975. "The Theory of 'Screening,' Education, and the Distribution of Income." *American Economic Review* 65(3):283–300.

Sutton, April, Amanda Bosky, and Chandra Muller. 2016. "Manufacturing Gender Inequality in the New Economy." *American Sociological Review* 81(4):720–48.

Tam, Pui-Wing. 2016. "A Talent War for Women." *The New York Times*. Retrieved (https://www.nytimes.com/2016/12/30/technology/join-our-board-companies-hotly-pursue-new-wave-of-women-in-tech.html).

Tilcsik, András. 2011. "Pride and Prejudice: Employment Discrimination against Openly Gay Men in the United States." *American Journal of Sociology* 117(2):586–626.

Tyre, Peg. 2008. *The Trouble with Boys: A Surprising Report Card on Our Sons, Their Problems at School, and What Parents and Educators Must Do*. New York: Three Rivers Press.

Uhlmann, Eric Luis, and Geoffrey L. Cohen. 2005. "Constructed Criteria: Redefining Merit to Justify Discrimination." *Psychological Science* 16(6):474–80.

van Ryn, Michelle, and Jane Burke. 2000. "The Effect of Patient Race and Socio-Economic Status on Physicians' Perceptions of Patients." *Social Science & Medicine* 50(6):813–28.

Williams, Christine L. 1992. "The Glass Escalator: Hidden Advantages for Men in the 'Female' Professions." *Social Problems* 39(3):253–67.

Williams, Christine L., Chandra Muller, and Kristine Kilanski. 2012. "Gendered Organizations in the New Economy." *Gender & Society* 26(4):549–73.

Williams, Joan. 2001. *Unbending Gender: Why Family and Work Conflict and What to Do about It*. New York: Oxford University Press.

Yavorsky, Jill. 2017. "Inequality in Hiring: Gendered and Classed Discrimination in the Labor Market." PhD dissertation, The Ohio State University, Columbus, OH.

Natasha Quadlin is Assistant Professor of Sociology at The Ohio State University. Her research focuses on social inequality in access and returns to education. Current projects examine perceptions of responsibility for college costs, income inequality among college graduates, and public attitudes toward gender and sexuality.