Cornell University
ILR School

Cornell University ILR School
DigitalCommons@ILR

Compensation Research Initiative

6-2009

The Signaling Role of Promotions: Further Theory and Empirical Evidence (CRI 2009-008)

Jed DeVaro
California State University

Michael Waldman
Cornell University, mw46@cornell.edu

Follow this and additional works at: http://digitalcommons.ilr.cornell.edu/cri

This Article is brought to you for free and open access by DigitalCommons@ILR. It has been accepted for inclusion in Compensation Research Initiative by an authorized administrator of DigitalCommons@ILR. For more information, please contact hlmdigital@cornell.edu.

The Signaling Role of Promotions: Further Theory and Empirical Evidence (CRI 2009-008)

Abstract
An extensive theoretical literature investigates the role of promotions as a signal of worker ability. In this paper we extend the theory by focusing on how the signaling role of promotion varies with a worker's education level, and then investigate the resulting predictions using a longitudinal data set that contains detailed information concerning the internal-labor-market history of a medium-sized firm in the financial-services industry. Our results support signaling being both a statistically significant and economically significant factor in promotion decisions. The paper also contributes to the extensive literature on the role of education as a labor-market signal.

Keywords
promotion, internal labor markets, signaling

Comments

Suggested Citation
DeVaro, J., & Waldman, M. (2009). *The signaling role of promotions: Further theory and empirical evidence* (CRI 2009-008). Retrieved [insert date] from Cornell University, ILR School, Compensation Research Initiative site: http://digitalcommons.ilr.cornell.edu/cri/8

This article is available at DigitalCommons@ILR: http://digitalcommons.ilr.cornell.edu/cri/8

THE SIGNALING ROLE OF PROMOTIONS:

FURTHER THEORY AND EMPIRICAL EVIDENCE

by

Jed DeVaro
Department of Management
and Department of Economics
College of Business and Economics
California State University, East Bay
Hayward, CA 94542
(510) 885-3289, jed.devaro@csueastbay.edu

and

Michael Waldman
Johnson Graduate School of Management
Cornell University
Sage Hall
Ithaca, NY 14853
(607) 255-8631, mw46@cornell.edu

June 2009

* We would like to thank Dan Bernhardt, Bob Gibbons, Michael Gibbs, Heather Haveman, James Malcomson, Joshua Pinkston, Bryan Ricchetti, Eric Van den Steen, Jan Zabojnik, two anonymous referees, the editor, Bentley MacLeod, and seminar and conference participants at Colgate University, Columbia University, Cornell University, University of Rochester, University of Washington, St. Louis, the Harvard-MIT Organizational Economics Workshop, and the 2005 Society of Labor Economists Conference for helpful comments, Fidan Ana Kurtulus and David Rosenblum for research assistance, and George Baker and Michael Gibbs for providing us with the data used in the study.

ABSTRACT

An extensive theoretical literature investigates the role of promotions as a signal of worker ability. In this paper we extend the theory by focusing on how the signaling role of promotion varies with a worker's education level, and then investigate the resulting predictions using a longitudinal data set that contains detailed information concerning the internal-labor-market history of a medium-sized firm in the financial-services industry. Our results support signaling being both a statistically significant and economically significant factor in promotion decisions. The paper also contributes to the extensive literature on the role of education as a labor-market signal.

I. INTRODUCTION

An extensive theoretical literature investigates the signaling role of promotions.[1] In these analyses, when a worker is promoted other potential employers observe the promotion and infer the worker has high ability. In this paper we first extend the theory by focusing on how the signaling role of promotion varies with a worker's education level, and then empirically investigate the resulting predictions. Our results support the idea that promotions serve as a signal of worker ability.

Most of the papers in this literature consider a model similar to the one originally investigated in Waldman (1984a). In that two-period analysis a firm's job ladder consists of two jobs, all young workers are assigned to the low-level job, and at the end of period 1 each worker's period-1 employer privately observes the worker's ability. Then in period 2 each firm chooses which of its period-1 employees to promote, where this promotion decision is publicly observed. There are four main results. First, when a worker is promoted other potential employers infer the worker is of high ability and thus increase their wage offers. Second, anticipating this, the period-1 employer offers a large wage increase with the promotion in order to prevent the worker from being bid away. Third, because a large wage increase is necessary, firms promote fewer workers than is first-best optimal. Fourth, this distortion decreases with the importance of firm-specific human capital.

In the first part of our analysis we enrich this theoretical approach to make it more realistic. Rather than taking the standard theoretical approach which assumes all workers are observationally equivalent when they enter the labor market, we build on Bernhardt (1995) and assume workers vary in terms of publicly observed schooling levels when they enter the labor market. Workers with more education have higher expected ability, although as in the standard approach a worker's actual ability is initially unobserved.

We first show that the model exhibits the basic signaling results found in the previous literature. That is, when a worker is promoted the worker receives a large wage increase in order to prevent the worker from being bid away. In turn, because of this wage increase, firms promote fewer workers than in the first best. We then show that these basic signaling results vary with a worker's schooling level.

---

[1] Papers in this literature include Waldman (1984a,1990), Ricart i Costa (1988), MacLeod and Malcomson (1988), Bernhardt (1995), Zabojnik and Bernhardt (2001), Owan (2004), and Golan (2005). See also the related analysis of Milgrom and Oster (1987). Other papers that allow for asymmetric information of the sort considered in this literature but do not consider the possibility of promotions serving as signals include Greenwald (1986), Lazear (1986), and Chang and Wang (1996).

Because a worker with a higher level of schooling has higher expected ability upon entering the labor market, the signal associated with promotion improves beliefs about worker ability less when the education level is high. This, in turn, yields the following three testable implications. First, because the wage associated with not being promoted is higher for workers with higher education levels, firms distort the promotion decision less for these workers. Second, the wage increase associated with promotion decreases with education because the signal is smaller for more highly educated workers. Third, since a first promotion signals a worker is high ability, the first two predictions hold more strongly for first promotions than for later promotions. The first prediction implies that, holding performance fixed, the probability of promotion is higher for more highly educated workers, as we explain in Section III.

To understand the logic, consider a firm that hires two workers into the same job, where the workers are similar except that one has an MBA while the other has only an undergraduate degree. Because employers believe MBAs are more productive on average, other firms learn little about the MBA from a promotion. As a result, if the MBA is promoted, the firm does not offer a large wage increase since there is not a big increase in other firms' wage offers. Further, since a promotion is not associated with a large wage increase, the firm does not distort the MBA's promotion decision in a significant fashion.

In contrast, suppose the firm promotes the worker who has only an undergraduate degree. Because such workers are, on average, not as highly thought of as MBAs, when the worker is promoted other firms positively update their beliefs concerning the worker's ability by a significant amount and, in turn, significantly increase their wage offers. The current employer must then provide a large wage increase upon promotion to prevent the worker from leaving and, because of the large wage increase, promotions only occur when the worker is significantly more productive at the higher-level job than at the lower-level job. In other words, the wage increase from a first promotion is larger for the less educated worker and the promotion decision is more biased against this worker.

In the second part of our analysis we investigate these predictions empirically, using panel data on the personnel records for managerial workers in a single medium-sized firm in the financial-services industry over a twenty-year period (this data set was first investigated in Baker, Gibbs, and Holmstrom's (1994a,b) classic empirical study of internal labor markets).[2] Two aspects of this data set make it ideal

[2] Other studies that investigate the operation of internal labor markets by analyzing the personnel records of a single firm include Lazear (1992), Seltzer and Merrett (2000), and Seltzer and Simons (2001).

for investigating our predictions. First, the personnel records contain annual supervisor ratings of each worker's job performance. Second, the data set includes detailed information on the firm's job ladder constructed by Baker, Gibbs, and Holmstrom from the raw data on job titles and typical promotion paths. Together, these features allow us to test how education affects both the probability of promotion and the size of wage increases received upon promotion.

The results of our analysis support the signaling theory of promotion. We find that increasing a worker's education level increases the probability of promotion, and other than for high school graduates we similarly find that the wage increase associated with promotion decreases with education. For example, after controlling for a variety of worker attributes such as initial job level, job performance, and firm tenure, decreasing the education level from masters degree to bachelors degree decreases by about twenty percent the probability a worker is promoted in the following year. In terms of education and wage growth, we find that after controlling for a number of worker attributes, decreasing the education level from masters degree to bachelors degree increases the average percentage wage increase due to promotion by over seventy percent. Consistent with the theory, we also find that these relationships are stronger for first promotions than for subsequent promotions. We also consider other potential explanations for our results and find that none match the evidence as well as the promotion-as-signal hypothesis.

Our analysis is related to Bernhardt (1995). That paper also considers an asymmetric-learning setting where upon labor-market entry there are observationally identifiable groups such as workers of different schooling levels that vary in terms of expected ability. Similar to what we show, in Bernhardt's analysis even after controlling for performance higher-ability groups are favored in the promotion process. But Bernhardt does not derive our other testable predictions, i.e., that wage increases due to promotion are smaller for higher-ability groups and that these relationships are stronger for first promotions. Also, although Bernhardt provides extensive discussion of the empirical literature, he does not conduct empirical tests.

This paper contributes to a small but growing empirical literature on asymmetric learning in labor markets (see Waldman (2008) for a discussion of this literature). The classic paper in this literature is Gibbons and Katz (1991) which focuses on the idea that a layoff sends a more damaging signal of worker

ability than being fired in a plant closing.[3] They develop a number of theoretical predictions and then empirically test the predictions and find supporting evidence. Doiron (1995), Acemoglu and Pischke (1998), and Grund (1999) also find supporting evidence for the Gibbons and Katz predictions, while Krashinsky (2002) and Song (2007) argue that the Gibbons and Katz results are due to factors other than asymmetric learning. More recently, Schonberg (2007), Kahn (2009), and Pinkston (2009) develop further tests of asymmetric learning in labor markets and in general find supporting evidence. For example, Schonberg finds evidence consistent with asymmetric learning for university graduates but not for high school graduates and dropouts.[4]

This paper also contributes to the extensive literature on education as a labor-market signal (see Riley (2001) for a survey). Much of the work on that topic focuses on the return to education as a signal in terms of the initial wage a worker receives when entering the labor market, or how the higher wage dissipates over time as firms learn true ability. A major point of our theoretical analysis, however, is that in a world of asymmetric learning, part of the return to education as a signal is improved promotion prospects that can lead to better outcomes even later in careers. Further, our empirical analysis shows clear support for the idea that a higher level of education results in improved promotion prospects.

II. MODEL AND THEORETICAL ANALYSIS

In this section we first present and analyze a two-period asymmetric-learning model with two jobs and multiple schooling levels. We then consider a three-period extension characterized by three job levels. The production technology employed throughout is closely related to that employed in Gibbons and Waldman (2006) in a recent study that focuses on symmetric learning.

A) A Two-Period Model

There is free entry into production, where all firms are identical and the only input is labor. A

---

[3] The Gibbons and Katz predictions can be interpreted either as a layoff serving as a negative signal, or in terms of the adverse-selection theory of labor-market turnover first put forth in Greenwald (1986).

[4] Another related paper is Belzil and Bognanno (2005). They employ an eight-year panel of promotion histories of 30,000 American executives to test between two explanations for fast tracks within firms – a symmetric-learning explanation developed in Gibbons and Waldman (1999a) and an asymmetric-learning/signaling explanation developed in Bernhardt (1995). Although their main result is that fast tracks are mostly explained by the Gibbons and Waldman explanation, they also find a result concerning schooling's effect on the probability of promotion that they interpret as the signaling role of promotion varying negatively with education.

worker's career lasts two periods, where in each period labor supply is fixed at one unit for each worker. We call workers in their first period in the labor market young and those in their second period old. Worker i enters the labor market with a schooling level, S_i, that can take on any integer value between 1 and N. We assume a positive number of workers at each value of S. Note that given much of our focus is on the information transmitted by a worker's schooling level, a simple interpretation is that the schooling level represents the highest degree earned by the individual.

Let η_{it} denote worker i's "on-the-job human capital" in period t, where

(1) $$\eta_{it}=\theta_i f(x_{it}).$$

In equation (1), θ_i is the worker's ability to learn on the job and x_{it} is the worker's labor-market experience prior to period t, i.e., $x_{it}=0$ for young workers and 1 for old workers. Also, $f(1)>f(0)>0$.[5] We assume that worker i with schooling level S_i has an ability to learn on-the-job given by $\theta_i=\varphi_i+B(S_i)$, where $B(S)>B(S-1)$ for $S=2,3,\ldots,N$. φ_i is a random draw from the probability density function $g(\varphi)$, where $g(\varphi)>0$ for all $\varphi_L<\varphi<\varphi_H$ and $g(\varphi)=0$ for all φ outside of this interval. Also, let $\theta^E(S)$ denote the expected value of θ for workers with schooling level S. Note that in this specification schooling is positively correlated with a worker's ability to learn on the job. This can be because schooling enhances human capital and thus increases a worker's ability to learn on the job, or because there is a positive relationship between schooling and innate ability to learn on the job and schooling serves as a signal. In Section V we discuss these two ways of interpreting/extending the model.

A firm consists of two jobs, denoted 1 and 2. If worker i is assigned to job j in period t, then the worker produces

(2) $$y_{ijt}=(1+k_{it})[d_j+c_j\eta_{it}]+G(S_i),$$

where d_j and c_j are constants known to all labor-market participants, $G'>0$ and $G''<0$, and k_{it} equals k, $k>0$, if the worker was employed at the firm in the previous period and zero otherwise (this means all young workers in any period t are characterized by $k_{it}=0$). $G(S_i)$ represents productivity due to general-purpose human capital accumulated prior to a worker entering the labor force, while k represents the importance of firm-specific human capital in this economy.

---

[5] We assume $f(0)>0$ rather than $f(0)=0$. One interpretation is that this model is a simplified version of a continuous-time model where production in the first period represents production in the early part of the worker's career. Since, on average, during this early part of the career the worker has a positive amount of labor-market experience, it is natural to assume $f(0)>0$.

Let η' denote the amount of on-the-job human capital at which a worker is equally productive at jobs 1 and 2. That is, η' solves $d_1+c_1\eta'=d_2+c_2\eta'$. We assume $c_2>c_1>0$ and $0<d_2<d_1$, i.e., as in Rosen (1982) and Waldman (1984b) output increases more quickly with ability in the high-level job. Thus, given full information about worker abilities, the efficient assignment rule for period t is to assign worker i to job 1 if $\eta_{it}<\eta'$ and to job 2 if $\eta_{it}>\eta'$.

We assume each worker's schooling level is known to all labor-market participants when the worker enters the labor market. In contrast, each worker's value for θ_i is not known by either the firms or the worker, although the density function g(.) and the function B(S) are common knowledge. Learning about θ_i takes place at the end of the worker's first period in the labor market when the first-period employer privately observes the worker's output. In addition, we assume the job assignment offered an old worker by the worker's first-period employer is public information. As discussed earlier, the result is that a promotion at the beginning of a worker's second period in the labor force serves as a signal of the worker's ability.

Workers and firms are risk neutral and have a zero rate of discount, while there is no cost to workers of changing firms or to firms from hiring or firing workers. To make the model consistent with standard wage determination at most firms, we assume wages are determined by spot-market contracting. In addition, since each worker's output is privately observed rather than publicly observed and verifiable, the wage specified in such a contract consists of a wage determined prior to production rather than a wage determined by a piece-rate contract where compensation depends on the realization of output.

The wage-setting process and timing of events is similar to that found in Zabojnik and Bernhardt (2001). At the beginning of each period, each firm offers each old worker it employed in the previous period a job assignment or fires the worker, where this decision is publicly observed. We assume a firm does not retain any worker it anticipates leaving during the wage-determination process. This assumption is consistent with the existence of a small cost of retaining a worker who then chooses to leave. Following Greenwald (1986), Lazear (1986), and Milgrom and Oster (1987), we also assume that the wage-determination process is characterized by counteroffers. That is, after the stage just described, each firm has the opportunity to make a wage offer to all workers other than those it employed in the previous period and then each firm makes a wage counteroffer to each of those previously employed workers.[6]

[6] One exception is that a firm does not make a counteroffer to a worker it fired at the beginning of the period. Also, when a firm makes an offer to an old worker it did not employ in the previous period, we assume the firm observes

Each worker then chooses the firm that offers the highest wage. If there are multiple firms tied at the highest wage, the worker chooses randomly among these firms unless one was the worker's employer in the previous period, in which case the worker remains with that firm. This tie-breaking rule is equivalent to assuming an infinitesimally small moving cost. Finally, at the end of each period each firm privately observes the output of each of its workers.

To reduce the number of cases, we restrict the analysis to parameterizations that satisfy the following conditions. First, $\theta^E(N)f(0)<\eta'$. That is, it is efficient for young workers to be assigned to job 1. Second, $[\varphi_H+B(S)]f(1)>\eta'>[\varphi_L+B(S)]f(1)$ for all S. This condition states that for each schooling level it is efficient for some old workers to be assigned to job 2 and others to be assigned to job 1.

Finally, we limit attention to Perfect Bayesian Equilibria, where we also impose an assumption similar to Trembling-Hand Perfection (see Selten (1975) for a discussion of Trembling-Hand Perfection). Specifically, in the second period there is a small probability the first-period employer mistakenly does not make a counteroffer when the firm has the smallest cost of choosing that action. To be more precise, suppose for example that in equilibrium a firm retains but does not promote old workers with schooling level S whose abilities to learn on the job fall in the interval $[\theta', \theta'']$. Then the firm sometimes mistakenly fails to make a counteroffer when ability to learn on the job falls in the interval $[\theta', \theta'+\varepsilon]$, where our focus is on what happens in the limit as ε approaches zero.[7] Restricting attention in this way means our analysis is characterized by a winner's-curse result similar to that found in Milgrom and Oster (1987) (see also Greenwald (1986) and Lazear (1986)), i.e., in bidding on another firm's retained old worker a firm is only willing to offer a wage equal to the productivity at the bidding firm of the lowest productivity worker with the same labor-market signal (meaning schooling level and job assignment).[8]

---

whether the worker was retained or fired and the job assignment if the worker was retained. The firm does not observe how many old workers the original employer fired, retained, or assigned to each job in the current period.

[7] Note we are not formally assuming Trembling-Hand Perfection since we assume the mistake of not making a counteroffer only occurs for some realizations of θ rather than there being a positive probability of this for all cases (further Trembling-Hand Perfection is not defined when the strategy space is continuous). Further, our equilibrium refinement is similar to the notion of a Proper equilibrium first discussed in Myerson (1978). We also assume that, if a worker's first-period employer makes an off-the-equilibrium path promotion decision at the beginning of the second period, then other firms infer the worker's ability is of the type that makes this decision least costly for the first-period employer.

[8] In our model this is not an immediate implication of restricting the analysis to Perfect Bayesian Equilibria both because of the presence of firm-specific human capital which Milgrom and Oster did not incorporate into their analysis, and because in our model a promoted worker is signaled to have high productivity while in theirs a worker's productivity becomes public knowledge.

B) Analysis and Testable Implications

We begin with a benchmark analysis. Suppose output is publicly observable so all firms learn each worker's ability to learn on the job after the worker's first period in the labor market (but compensation is still spot-market wages). Given our parameter restriction $\theta^E(N)f(0)<\eta'$, every young worker is assigned to job 1. Let $w_Y^*(S)$ denote the wage for young workers with schooling level S. We have that $w_Y^*(S)>d_1+c_1\theta^E(S)f(0)+G(S)$ for all S, i.e., young workers are paid more than their expected output. This occurs because old workers are paid less than expected output – see below – and the zero-profit condition associated with competition thus means young workers must be paid more than expected output.

Now consider old workers. There are three conditions that define equilibrium behavior when workers are old. First, there is no turnover. Second, old worker i is assigned to job 1 if $\eta_{it}<\eta'$ and to job 2 if $\eta_{it}\geq\eta'$.[9] Third, $w_O^*(\eta_{it})=\max\{d_1+c_1\eta_{it}+G(S_i),d_2+c_2\eta_{it}+G(S_i)\}$, where $w_O^*(\eta_{it})$ is the wage paid to old worker i as a function of the worker's current on-the-job human capital.

The logic for these results is as follows. Given output is publicly observable, there is no asymmetric information in this benchmark case, i.e., at any date all firms (and the worker) are equally informed about a worker's on-the-job human capital. The result is that, given the information available, workers are assigned to jobs in the efficient fashion and switch employers in the efficient fashion. In turn, given there is firm-specific human capital, this last condition means there is no turnover. Finally, each old worker is paid the wage that the market, i.e., other firms, offers the worker which is the worker's expected productivity given he or she switches employers (note that a worker who switches employers has zero firm-specific human capital which explains the expression for $w_O^*(\eta_{it})$).

The main point of the benchmark analysis is that, if output is publicly observable, then job assignments as well as turnover decisions are efficient. As we show below, in contrast, once a worker's output is privately observed by the worker's employer, then job assignments are no longer efficient. Rather, firms assign too few old workers to job 2 in order to avoid sending the positive signal about productivity associated with this assignment.

---

[9] To simplify descriptions of behavior, throughout the theoretical analysis we assume that an old worker is promoted whenever the employer is indifferent between promoting and not promoting the worker. Also, in descriptions of equilibrium behavior we ignore what happens when in the second period the first-period employer mistakenly fails to make a counteroffer.

Suppose a worker's output each period is privately observed by the worker's employer. We start with some preliminary results. Equilibrium behavior concerning young workers is similar to the benchmark case. That is, as in the benchmark, our parameter restriction $\theta^E(N)f(0)<\eta'$ yields that all young workers are assigned to job 1. Also similar to what was true in the benchmark, the wage paid to young worker i, $w_Y(S_i)$, is above expected output and is such that a firm hiring a young worker earns zero expected profits from the hire. In the benchmark case this occurred because of profits earned in the following period due to the presence of firm-specific human capital. In contrast, now this occurs because of both future profits due to the presence of firm-specific human capital and future profits due to the presence of asymmetric information about worker productivity.

We now formally state what happens in this case. Below $w_O(S_i,\eta_{it})$ is the wage paid to old worker i as a function of the worker's schooling level and on-the-job human capital, while j_{it} is the firm that individual i works at in period t.[10] All proofs are in the Appendix.

<u>Proposition 1</u>: If a worker's output is privately observed by the worker's employer, then there exists a function $\eta^+(S)$, $\eta'<\eta^+(S)\leq[\varphi_H+B(S)]f(1)$ for all S, such that i) through iii) describe equilibrium behavior in each period t.[11]

. i) Each young worker i is assigned to job 1 and is paid $w_Y(S_i)>d_1+c_1(\theta^E(S_i)f(0))+G(S_i)$.

ii) If old worker i is such that $\eta_{it}\geq\eta^+(S_i)$, then the worker is assigned to job 2, remains at firm j_{it-1}, and is paid $w_O(S_i,\eta_{it})=d_2+c_2\eta^+(S_i)+G(S_i)$.

iii) If old worker i is such that $\eta_{it}<\eta^+(S_i)$, then the worker is assigned to job 1, remains at firm j_{it-1}, and is paid $w_O(S_i,\eta_{it})=d_1+c_1[\varphi_L+B(S)]f(1)+G(S_i)$.

Proposition 1 tells us that, if a worker's output is privately observed by the worker's employer, then for each schooling group there is a critical value for η_{it}, $\eta^+(S_i)$, that determines what happens when the worker is old. If η_{it} is below the critical value, then the worker is not promoted and stays with the initial employer. If η_{it} is above the critical value, then the worker is promoted and again stays with the initial employer. Further, in each case the worker's wage equals the productivity at another potential

---

[10] We focus on the unique equilibrium characterized by no workers being fired. This is the only equilibrium if the two jobs are sufficiently similar or k is sufficiently large.

[11] When no old workers of schooling level S are promoted, we say $\eta^+(S)$ equals $[\varphi_H+B(S)]f(1)$.

employer of the lowest productivity worker with the same labor-market signal. For example, the wage paid to a worker with schooling level 1 who is not promoted equals the productivity at an alternative employer of an old worker with on-the-job human capital equal to $[\varphi_L+B(1)]f(1)$. The logic is that, given our Trembling-Hand type assumption concerning counteroffers, there is a winner's-curse problem in which other potential employers are only willing to pay the lowest possible productivity of a worker with the same labor-market signal. Hence, in order to retain a worker, this is all an old worker's previous employer needs to offer.

Another interesting aspect of the proposition is that promotion decisions are not efficient. That is, $\eta^+(S)>\eta'$ for all S means each firm assigns fewer old workers to job 2 than is efficient given the firm's knowledge concerning its workers' productivities. The logic here is the same as in Waldman (1984a). Because assigning an old worker to job 2 signals that the worker has high productivity, firms give promoted workers large wage increases in order to stop them from being bid away. In turn, because of the need to pay this high wage, an old worker's initial employer only assigns the worker to job 2 if his or her productivity in job 2 significantly exceeds productivity in job 1.

We now translate these results into two testable implications. The model's first testable implication concerns how a young worker's output translates into the following period's promotion decision. Below, $y^P(S)$ denotes the minimum output level required for a young worker with schooling level S to be promoted when he or she becomes old.

<u>Corollary 1</u>: Suppose there is a strictly positive number of promotions for workers of schooling levels S_1 and S_2, $S_2>S_1$.[12] Then $\eta^+(S_2)<\eta^+(S_1)$ and, if k is sufficiently small, $y^P(S_2)<y^P(S_1)$.

Corollary 1 states that, if k is sufficiently small, then the performance level required to achieve promotion falls with education. This is closely related to the idea that the incentive to distort the promotion decision is decreasing with education, i.e., $\eta^+(S_2)<\eta^+(S_1)$ for $S_2>S_1$. There are two steps to the argument. First, as discussed earlier, a firm promotes fewer workers than is efficient because of the wage increase associated with promotion. But since the wage paid to workers not promoted, i.e., $[d_1+c_1(\varphi_L+$

[12] It is possible that the incentive for firms to distort the promotion decision results in no promotions for workers of specific schooling levels.

$B(S))]f(1)+G(S)$, increases with schooling, the incentive to distort the promotion decision is smaller for workers with more education. Hence, firms distort the promotion decision less for workers with more schooling or, in other words, the critical value $\eta^+(S)$ is closer to η' for higher values for S.

The second step translates this first result concerning how $\eta^+(S)$ varies with S into a statement concerning how $y^P(S)$ varies with S. By definition, $y^P(S)=c_1+d_1\eta^+(S)+G(S)$. Thus, there are two countervailing effects as S rises from S_1 to S_2. First, G(S) rises. Second, as just discussed, $\eta^+(S)$ falls. When k is small, there is a large difference between education levels in terms of how much the promotion decision is distorted, i.e, $\eta^+(S_1)-\eta^+(S_2)$ is large. The logic is that the return to efficiently assigning a worker is lower the lower is k. As a result, when k falls both $\eta^+(S_1)$ and $\eta^+(S_2)$ move further from η' and $\eta^+(S_1)-\eta^+(S_2)$ gets larger. Finally, because when k is small $\eta^+(S_1)-\eta^+(S_2)$ is large, with small enough k the second effect dominates with the result that $y^P(S_2)<y^P(S_1)$.[13]

The second testable implication concerns the size of wage increases due to promotion, i.e., the wage increase a worker receives upon promotion minus the wage increase the same worker would have received if there had been no promotion. Below $\Delta w^P(S)$ denotes the wage increase due to promotion as a function of the education level, while $\Delta_{\%}w^P(S)$ denotes the percentage wage increase due to promotion as a function of the education level.

<u>Corollary 2</u>: Suppose there is a strictly positive number of promotions for workers of schooling levels S_1 and S_2, $S_2>S_1$. Then $\Delta w^P(S_2)<\Delta w^P(S_1)$ and $\Delta_{\%}w^P(S_2)<\Delta_{\%}w^P(S_1)$.

Corollary 2 states that both the absolute and percentage wage increases due to promotion are decreasing in the schooling level. The logic follows. As discussed earlier, the wage of a promoted worker is the expected productivity at an alternative employer of the lowest productivity worker with the same schooling level who is promoted. Similarly, the wage of an old worker who is not promoted is the expected productivity at an alternative employer of the lowest productivity worker with the same schooling level who is not promoted. Given no one is fired in equilibrium, the absolute wage increase due to promotion thus equals $[[d_2+c_2\eta^+(S)]-[d_1+c_1[\varphi_L+B(S)]]]f(1)$. Given this, there are two reasons why

[13] In a less realistic version of the model that omits the general human capital term, G(S), $y^P(S_2)<y^P(S_1)$ holds regardless of the level of k. In this case, however, the model would yield a prediction that is at odds with the standard finding in the empirical literature that, even after controlling for job assignment and experience, a worker's wage is positively related to the worker's education level.

the absolute wage increase due to promotion is decreasing in the schooling level. First, as discussed earlier, $\eta^{+}(S)$ falls with the education level because the incentive to distort the promotion decision falls. Second, B(S) is increasing in the education level or, in other words, the expected productivity of the worst overall worker rises with education. Further, since the wage for old workers not promoted increases with education, the fact that the absolute wage increase due to promotion falls with education means the percentage wage increase due to promotion also falls with education.

As a final point, note that although our first two testable implications are derived from a specific model of the promotion-as-signal hypothesis, the two predictions are in fact robust predictions of this hypothesis, i.e., various alternative models also yield these predictions. The reason is that the basic logic of the signaling argument leads to these two results. This logic is that a promotion serves as a positive signal of ability, so firms limit the number of promotions in order to avoid paying the higher wage associated with a promotion which is necessitated by the signal. Now add to this basic logic workers who vary in terms of schooling, where higher schooling levels are correlated with higher ability. Since workers with higher schooling levels are already thought of as being of higher ability, the signal associated with promotion and thus the wage increase associated with promotion is smaller for such workers, i.e., our second testable implication. Further, since the wage increase associated with promotion is smaller for workers with more education, the incentive to distort the promotion decision is smaller for such workers. In turn, this frequently translates into the performance level required to achieve promotion being smaller for more highly educated workers, i.e., our first testable implication.[14]

C) A Three-Period Extension

In this subsection we consider a three-period extension of the model analyzed above. Because of the added complexity of the three-period specification and space limitations, we provide an informal discussion of the three-period extension rather than a formal analysis. The main point is that in the three-

[14] A specific alternative model of interest is one in which everything is the same as in the model we consider but there is a stochastic term in the production function. We have considered this alternative specification given two periods and two jobs and most of our results are unchanged. First, the basic description of equilibrium captured in Proposition 1 is qualitatively unchanged, where cutoff values for on-the-job human capital are defined in terms of expected values rather than actual values. Second, our second testable implication that the wage increase due to promotion is a decreasing function of worker education is unchanged. Third, our first testable implication that, given k sufficiently small, the performance level required to achieve promotion falls with education does not hold generally, but does hold given the variance on the production function's stochastic term is not too large, i.e., as long as a worker's output is sufficiently informative of the worker's ability to learn on the job.

period setting the two testable implications derived above hold more strongly for first promotions than for subsequent promotions.[15]

Most of the set-up is the same as in the model analyzed above except that workers are in the labor market for three periods and there are three jobs. We call workers in their first period in the labor market young, those in their second period middle-aged, and those in their third period old. Equation (1) still describes the evolution of on-the-job human capital where $f(2)>f(1)>f(0)>0$. Equation (2) still describes productivity where, consistent with the earlier definition, $k_{it}=k$, $k>0$, if the worker was at the same firm in the previous period and zero otherwise. That is, if middle-aged worker i moves from firm A to firm B and then back to firm A in the following period, the worker's productivity in the last period is not increased by any firm-specific human capital.[16] Also, we assume $c_3>c_2>c_1>0$ and $0<d_3<d_2<d_1$. As before, η' solves $d_1+c_1\eta'=d_2+c_2\eta'$, while η'' solves $d_2+c_2\eta''=d_3+c_3\eta''$. So worker i in period t is efficiently assigned to job 1 when $\eta_{it}<\eta'$, efficiently assigned to job 2 when $\eta'<\eta_{it}<\eta''$, and efficiently assigned to job 3 when $\eta_{it}>\eta''$. One other new assumption is that we now allow discounting (introducing discounting into the two-period model has no effect on the qualitative results).

To reduce the number of cases, we restrict the discussion to parameterizations that satisfy the following conditions. First, $\theta^E(N)f(0)<\eta'$. That is, as before, it is efficient for all young workers to be assigned to job 1. Second, $[\varphi_L+B(S)]f(1)<\eta'<[\varphi_H+B(S)]f(1)<\eta''$ for all S. This condition states that for each schooling group it is efficient for some middle-aged workers to be assigned to job 1, some to job 2, while none are efficiently assigned to job 3. Third, $[\varphi_L+B(S)]f(2)<\eta'<\eta''<[\varphi_H+B(S)]f(2)$ for all S. This condition tells us that for each schooling group it is efficient for some old workers to be assigned to job 1, some to job 2, and some to job 3.

The equilibrium in this model has a number of intuitive features. First, as workers age, more workers earn a first promotion. That is, there are values $\eta_M^+(S)$ and $\eta_O^+(S)$, $\eta_O^+(S)<\eta_M^+(S)$, such that middle-aged worker i with schooling level S earns a promotion from job 1 to job 2 in period t when $\eta_{it}\geq\eta_M^+(S)$, while old worker i with schooling level S earns a promotion from job 1 to job 2 in period t when $\eta_O^+(S)\leq\eta_{it}<\eta_M^+(S)$. Second, a subset of the workers promoted to job 2 when middle-aged earn a

[15] We would like to thank one of the anonymous referees for suggesting this result to us.

[16] Following Bernhardt (1995), we also assume that if middle-aged worker i moves from firm A to firm B, then by the beginning of the subsequent period firm A has "forgotten" worker i's value for θ_i. Assuming otherwise greatly complicates the analysis because then a worker has an incentive to move when middle-aged in order to increase the number of firms who know the worker's ability (see Ricchetti (2007) for an analysis along these lines).

subsequent promotion to job 3 when old. That is, there exists a value $\eta_O^{++}(S)$, $\eta_O^{++}(S)>\eta_M^{+}(S)$, such that old worker i with schooling level S earns a promotion from job 2 to job 3 in period t when $\eta_{it}\geq\eta_O^{++}(S)$. Third, because of signaling, these promotion cutoff values are such that an inefficiently small number of workers earn promotions. That is, $\eta'<\eta_O^{+}(S)<\eta_M^{+}(S)$ and $\eta''<\eta_O^{++}(S)$ for all S.

We now consider the extent to which this model is consistent with the two testable implications of the previous subsection. We begin by considering the promotion decision concerning old workers who were not promoted when middle-aged. This promotion decision is mathematically identical to the promotion decision of old workers in the two-period model, so both testable implications hold for this case. First, because the wage associated with not promoting such a worker is increasing in the worker's schooling level, the incentive to distort the promotion decision falls with the schooling level. The result is that, if k is sufficiently small, the minimum middle-aged output level required for promotion falls with the schooling level. Second, this translates into both absolute and percentage wage increases due to promotion falling with the schooling level.

Now consider the promotion decision for middle-aged workers. Analysis of this promotion decision is more complicated than the promotion decision of old workers who were not previously promoted. This is because the analysis needs to take into account that whether or not the worker is promoted this period affects what happens when the worker becomes old. There are now two factors influencing how extra education affects the promotion decision. First, similar to what was true before, more education increases the wage of middle-aged workers who are not promoted. This suggests the incentive to distort the promotion decision should fall with education. Second, the marginal middle-aged worker not promoted will be promoted when old and, because as above the incentive to distort the promotion decision for old workers falls with education, part of the wage of workers first promoted when old falls with education. This suggests the incentive to distort the promotion decision for middle-aged workers should increase with education. If there is sufficient discounting, the first effect dominates, the distortion of the promotion decision for middle-aged workers falls with education, and our two testable implications hold.

Now consider old workers who were promoted when middle-aged. Based on the above, if there is sufficient discounting, the distortion of the promotion decision for middle-aged workers falls with schooling. When this is the case the minimum value for on-the-job human capital for this set of workers falls rather than rises with schooling. The result is that, instead of the distortion of the promotion decision

falling with education as in the previous two cases, the distortion of the promotion decision actually rises with education. And this, in turn, means that our two testable implications do not hold for old workers who were promoted while middle-aged.

As a final point, the above discussion indicates that in our three-period extension, given sufficient discounting, our two testable implications hold for first promotions but not for subsequent promotions. But we believe this result is not fully robust. In a richer model where ability is multi-dimensional and different jobs put different weights on the various dimensions of ability, it is quite possible that even for subsequent promotions more education would serve to decrease the distortion of the promotion decision. But to the extent there is positive correlation between the realizations for ability across these different dimensions, using logic like that above suggests that education should have a smaller effect on the distortion of the promotion decision for subsequent promotions. So we will take as our third testable implication that our first two testable implications should hold more strongly for first promotions than for subsequent promotions, but that these two predictions might hold to some extent even for subsequent promotions.

III. TESTING THE PREDICTIONS

This section first describes the data and then presents our basic tests of the theoretical predictions developed in the previous section.

A) Data

Our data consist of the complete set of annual personnel records during the period 1969 to 1988 for all white-male-managerial employees of a medium-sized US firm in the financial-services industry. The data were originally constructed by George Baker, Michael Gibbs, and Bengt Holmstrom from the raw data contained in the firm's personnel records, and then analyzed in their classic empirical study of internal labor markets found in Baker, Gibbs, and Holmstrom (1994a,b) (the first of these papers contains a detailed description of the data). Their analyses used the full sample of managerial employees, including females and nonwhite males, for a total of 68,437 employee-years of data. The sample of white males that Baker, Gibbs, and Holmstrom shared with us has 50,556 employee-years. The key variables for our analysis are promotions, salaries, education, and supervisor subjective performance ratings measured annually on a five-point scale where 1 denotes the highest performance level and 5 the lowest.

As control variables we also employ demographic characteristics, firm tenure, job title, and level in the job hierarchy.

All variables are measured on December 31 for each employee in each year and pertain to that year. We do not observe the exact date of changes in job title or pay, so if an individual is promoted in, for example, 1979, we do not know whether the promotion occurred early in the year or late. Thus, the meaning of the worker's 1979 performance rating is unclear. If the promotion occurred early in the year, then the rating likely reflects performance in the post-promotion job. However, if the promotion occurred late in the year, then the rating likely pertains to the pre-promotion job. To avoid this ambiguity, we define pre-promotion performance as performance in the year prior to the promotion and post-promotion performance as performance in the year after the promotion.

Most of the variables are observed for each employee for each of the sample years in which the individual worked as a managerial employee at the firm.[17] One exception is that job titles were not recorded for some new hires in the last years of the data set, though other variables were. This means that we lose some observations since we include job title dummies as controls. Another source of missing observations results from our use of subjective performance ratings, since some workers were not rated in some years. Yet another source of missing observations concerns firm tenure, since we do not observe the year in which workers observed in 1969 entered the firm. Since all of our tests control for tenure, all workers observed in 1969 are dropped.

To define a promotion, we begin with the job ladder constructed by Baker, Gibbs, and Holmstrom from information on job titles. There are eight job levels in the firm, where level 8 is the highest level job held by the CEO. We define a promotion as a transition to a higher-level job, so, for example, a worker who moves from level 4 to level 5 in a given year is counted as receiving a promotion, as is a worker who moves from level 4 to levels 6 or 7. In other words, we do not distinguish between one-step and multiple-step promotions, but we do not think this is a concern since roughly ninety-eight percent of promotions in the sample are one-step promotions. In addition, non-promoted workers include a very small number of workers who were demoted (less than one percent of the sample).

---

[17] Since we only observe managerial employees, we do not know if a new entrant to the sample in a given year is a new entrant to the firm or has instead been promoted from a clerical position. As suggested in Baker, Gibbs, and Holmstrom (1994a), however, such promoted workers would likely be treated the same way as new hires to the firm, since the promotion entails a large change in job tasks and thus the retirement of most of the task-specific human capital acquired in the pre-promotion job. See the end of this section for a related discussion.

Salary is measured as the real annual salary in 1988 dollars, deflated by the CPI. The salary data do not include bonuses, since bonus information is only available for 1981 to 1988. Baker, Gibbs, and Holmstrom (1994a) also ignore the bonus data and argue that bonuses change total compensation very little for most employees.[18] Using the full sample (including females and nonwhite males), they found that only twenty-five percent of employees received bonuses in the 1981 through 1988 time period and that these workers were heavily concentrated at the top of the job ladder. Since our sample restrictions eliminate workers in levels 4 and higher, ignoring bonus data should have little effect. Also, for the few workers at lower levels who receive bonuses, the bonuses account for a modest fraction of total compensation (the median bonus for workers who receive bonuses in levels 1, 2, and 3 is less than ten percent of salary, while it is less than fifteen percent for workers in level 4). Finally, a small number of observations concern employees operating in branches outside of the United States. Since the nominal salary data were recorded in local currencies, we follow Baker, Gibbs, and Holmstrom and drop these observations from our tests that require salary data.

Education is recorded in years in the data set, though we aggregate this variable into a set of dummy variables designed to capture different degrees. Specifically, we construct dummy variables for high school graduate (including some college), bachelors degree, MBA or other masters degree, and Ph.D. degree. As discussed in more detail in Section V, we believe the most plausible interpretation of our model is that a higher level of schooling serves as a signal that the worker belongs to a higher productivity group, and thus it is the receipt of a degree that is important rather than years of education. In other words, focusing on education as a signal, taking five years to complete a bachelors degree does not signal higher quality than taking four. We thus exploit only the variation in educational attainment that occurs at discrete cutoffs defined by years of typical degree completion.[19]

To be precise, a high school graduate is defined as a worker with twelve, thirteen, or fourteen years of education. A bachelors degree holder is defined as a worker with sixteen years of education. An

---

[18] A possible concern from omitting bonuses is that bonuses and promotions may be substitutes. Related to this issue, Daniel Parent in a private correspondence reported to us that in the PSID data investigated in Lemieux, MacLeod, and Parent (2009) analysis points to the opposite result, i.e., "bonuses and promotions are positively correlated, even controlling for unmeasured worker effects." Note that we also investigate this issue in our data set and find the same result found by Parent.

[19] We have also conducted our tests using years of education rather than degree dummy variables. Consistent with the above discussion, in general this approach is less consistent with our theoretical predictions than the approach of employing degree dummy variables.

MBA or other masters degree holder is defined as a worker with eighteen years of education. Finally, a Ph.D. is defined as a worker with twenty-one or more years of education (although there are no workers in the data set with exactly twenty-one years of education). We exclude from the sample workers with fifteen, seventeen, nineteen, and twenty years of education, since these workers do not fall clearly into one of our four degree categories. This exclusion sacrifices roughly four percent of our sample, where about ninety-seven percent of these excluded workers have seventeen years of education. A worker with seventeen years of education could be someone who took five years to complete a bachelors degree or someone who took three years and completed a two-year masters degree. Given the importance of distinguishing between workers with bachelors degrees and those with masters degrees, we exclude these workers from our sample.

In our theoretical model, workers in any of the N education groups could be employed in any of the jobs. In the data, however, some jobs are never held or almost never held by workers with specific education levels. There are seventeen job titles in the data set, labeled A through Q.[20] As seen in Table 1, no Ph.D.s are present in job titles A, B, H, I, and J, and no high school graduates are present in titles J and M. Further, several of the other job titles have a negligible fraction of workers from one or more education groups. Since our theoretical analysis is based on career paths that are regularly traversed by workers from all education groups, in our main specifications we restrict attention to workers in titles for which the fraction of occupancy is greater than one percent for each of the four groups. This selection criterion means that our main empirical analysis includes job titles C, D, G, K, and Q which are the job titles highlighted in italics and boldface in Table 1. We also consider alternative samples as sensitivity checks. First, we estimate the model using a less stringent rule for including job titles, allowing any title for which the occupancy rate is positive for each group, even if only due to a single worker. This rule includes job titles C, D, E, F, G, K, L, N, O, P, and Q. Second, we consider the full sample (see footnotes 23, 27, and 28).

In summary, we omit observations of workers in foreign plants, of workers who were already at the firm in 1969, for which the subjective performance rating is missing, for which job-title data are missing, for which years of schooling equals fifteen, seventeen, nineteen, or twenty, and in our main specifications for which the job title is other than C, D, G, K, and Q. Table 2 displays descriptive

[20] As in the original papers by Baker, Gibbs, and Holmstrom, the actual job titles are disguised to protect the anonymity of the firm.

statistics for the main variables in our analysis using the sample-selection rules just described. Note that our sample restrictions eliminate workers in levels 4, 5, 6, 7, and 8.

B) The Basic Tests

We begin our empirical analysis by considering the performance ratings and the concern that they may be biased measures of performance. Unfortunately, we do not have any description of the performance evaluation process in the firm. For example, we do not know whether evaluators were told to rate individuals in an absolute or relative manner.

Presumably, worker productivity rises with tenure at the firm, as in our theoretical model. One possibility, therefore, is that performance ratings are measured relative to others with the same level of tenure, and this would be problematic for employing the performance ratings as absolute measures of productivity as we do in our empirical work. To investigate this possibility, we estimate a model with performance as the dependent variable and tenure with the firm as the independent variable. We aggregate the performance ratings into two categories (equaling 1 if the rating equals 1 or 2 and 0 otherwise) so that we can estimate a fixed-effects logit model. As seen in the first column of Table 3, the coefficient on the tenure variable is positive and statistically significant at the five percent level. As seen in the second column of Table 3, the results are similar when we add various controls (years at level, years at title, job level, year dummies, and job-title dummies).

A second possibility is that performance ratings are measured relative to other individuals at the same job level. This would suggest that performance ratings would fall, on average, when a promotion occurs. In column 3 of Table 3 we investigate this possibility using a fixed-effects logit specification in which the dependent variable is the same one investigated in columns 1 and 2 but now the independent variable of interest is whether the worker was promoted the previous year. The coefficient on the promotion variable is positive and statistically significant at the five percent level. In column 4 we add the controls from column 2 and we again find a promotion coefficient that is positive and statistically significant. Although we cannot rule out the possibility that the performance ratings reflect some element of relative performance, these results are consistent with our assumption that performance ratings at the firm largely reflect absolute performance.

We now consider our first testable prediction which is captured in Corollary 1, i.e., given k sufficiently small, the minimum output level required for promotion, $y^P(S)$, decreases with the education

level. Because the Baker, Gibbs, and Holmstrom firm exhibits significant turnover at all job levels, we believe the evidence is consistent with the firm having a low level of firm-specific human capital. Hence, for the Baker, Gibbs, and Holmstrom firm, we interpret Corollary 1 as stating that the threshold level of output required to achieve promotion should be decreasing with the education level.[21]

In our theoretical model, we assume for simplicity that the relationship between skills and output is deterministic and that the employer perfectly observes a worker's output each period. These assumptions imply a simple cutoff rule for promotion based on current-period output. In a more realistic and generalized version of the model in which each period's output is stochastic, promotions would still be characterized by cutoffs rules, but the employer would be comparing expected future profit from promotion to expected future profit if the worker is not promoted. If output follows a Markov process, the best predictor of a worker's future productivity is the worker's current output, so the cutoff rule would be defined in terms of observed current-period output. When developing the empirical model below, we assume a cutoff rule based on current-period output. While this is consistent with a literal interpretation of the theoretical model, as just noted, it is also consistent with a more general framework in which output is stochastic and follows a Markov process, and this latter interpretation is more realistic.[22]

Since the threshold level of output required to achieve promotion is unobserved by the econometrician, we develop an empirical specification that allows us to test our first prediction using the observed data on promotions, schooling levels, and performance ratings. Letting the subscript i index workers and t index years, in what follows $PROMOTION_{it}$ is a dummy variable that equals one if worker i is promoted in year t and zero otherwise, P_{it-1} is the performance rating of worker i in year t-1, while HS_{it-1}, MA_{it-1}, and PHD_{it-1} are dummy variables each of which equals one if at date t-1 worker i's years of schooling indicates that schooling level being the highest level of educational attainment (as defined earlier) and zero otherwise.

---

[21] In a private correspondence, Michael Gibbs has indicated to us that, based on his knowledge of the actual identity of the firm, he also believes that the Baker, Gibbs, and Holmstrom firm is characterized by little firm-specific human capital.

[22] It shold be noted that if, instead of a Markov process, the firm simply observed output with noise in each period, there would be incremental information in lagged values of output, i.e., current-period output would not be a sufficient statistic for the sequence of prior outputs. To account for this possibility, we report results from specifications in which additional lags of performance are included.

Similar to the theoretical model, let y_{it-1} and y^P_{it-1} denote, respectively, worker i's output in t-1 and the minimum output in t-1 required for worker i to be promoted in t, where by definition this means $PROMOTION_{it}=1$ if $y_{it-1}-y^P_{it-1}\geq 0$ and $PROMOTION_{it}=0$ if $y_{it-1}-y^P_{it-1}<0$. Since both of these variables are unobserved by the econometrician, we specify them as latent index variables.

(3) $$y_{it-1}=h(P_{it-1})+e_{it-1}$$

(4) $$y^P_{it-1}=\psi_0+\psi_1 HS_{it-1}+\psi_2 MA_{it-1}+\psi_3 PHD_{it-1}+\mathbf{X_{it-1}}\boldsymbol{\tau}+v_{it-1}$$

In this specification h(·) is monotonically decreasing, while e and v are stochastic disturbances. Note that h(·) is decreasing because of the way the performance rating is defined, i.e., 1 is the highest rating and 5 the lowest. The vector of controls, $\mathbf{X_{it-1}}$, includes age, age squared, tenure at the firm, tenure at the job level, and dummies for job level, job titles, and years. Note that Corollary 1 implies $\psi_3<\psi_2<0<\psi_1$. That is, the minimum output level required to achieve promotion is decreasing with the schooling level, where the excluded group is the college educated.

Substituting (3) and (4) into the expressions for $PROMOTION_{it}$ given above yields (5a) and (5b).

(5a) $$PROMOTION_{it}=1 \text{ if } h(P_{it-1})-\psi_0-\psi_1 HS_{it-1}-\psi_2 MA_{it-1}-\psi_3 PHD_{it-1}-\mathbf{X_{it-1}}\boldsymbol{\tau}\geq v_{it-1}-e_{it-1}$$

(5b) $$PROMOTION_{it}=0 \text{ if } h(P_{it-1})-\psi_0-\psi_1 HS_{it-1}-\psi_2 MA_{it-1}-\psi_3 PHD_{it-1}-\mathbf{X_{it-1}}\boldsymbol{\tau}<v_{it-1}-e_{it-1}$$

Assuming h(·) is a linear function of P_{it-1}, (5a) and (5b) can be rewritten as (6a) and (6b), where $\mu_{it-1}=v_{it-1}-e_{it-1}$, $\beta_0=-\psi_0$, $\beta_1=-\psi_1$, $\beta_2=-\psi_2$, $\beta_3=-\psi_3$, $\boldsymbol{\delta}=-\boldsymbol{\tau}$, and $\xi<0$.

(6a) $$PROMOTION_{it}=1 \text{ if } \mu_{it-1}\leq\beta_0+\beta_1 HS_{it-1}+\beta_2 MA_{it-1}+\beta_3 PHD_{it-1}+\mathbf{X_{it-1}}\boldsymbol{\delta}+\xi P_{it-1}$$

(6b) $$PROMOTION_{it}=0 \text{ if } \mu_{it-1}>\beta_0+\beta_1 HS_{it-1}+\beta_2 MA_{it-1}+\beta_3 PHD_{it-1}+\mathbf{X_{it-1}}\boldsymbol{\delta}+\xi P_{it-1}$$

Assuming that μ_{it-1} has the standard normal distribution, the promotion rule is described by the following probit model.

(7) $$\text{Prob}(PROMOTION_{it}=1)=\boldsymbol{\Phi}(\beta_0+\beta_1 HS_{it-1}+\beta_2 MA_{it-1}+\beta_3 PHD_{it-1}+\mathbf{X_{it-1}}\boldsymbol{\delta}+\xi P_{it-1})$$

Since $\beta_j=-\psi_j$ for j=1,2, and 3, the prediction $\psi_3<\psi_2<0<\psi_1$ translates into $\beta_1<0<\beta_2<\beta_3$. In other words, controlling for worker performance and the additional control variables in **X**, the probability that a worker is promoted in any year t should be an increasing function of the worker's education level.

The first column of Table 4 reports results of our estimation of equation (7) for the basic sample, where to make interpretation easier we report marginal effects rather than probit coefficients. The results provide clear support for our first testable prediction. Other things equal, and in particular holding performance fixed, the probability of promotion is 5.9 percentage points lower for high school graduates than for bachelors degree holders, 5.2 percentage points higher for masters degree holders than for

bachelors degree holders, and 15.7 percentage points higher for Ph.D.s than for bachelors degree holders. As discussed earlier, we also estimate the model using a less stringent rule for including job titles, allowing any job title for which the occupancy rate is positive for each educational group. As shown in the second column of Table 4, the predicted results also hold given this less stringent rule for including job titles, with all three marginal effects strongly statistically significant.[23]

A potential concern is that the performance rating in year t-1 might not fully capture worker performance in the pre-promotion job (see footnote 22 for a related discussion). Specifically, relevant pre-promotion performance might span multiple years in which case earlier performance ratings would also matter. To investigate this issue, we estimated a specification that includes performance in years t-1 and t-2. These results are reported for the basic sample in the third column of Table 4. As can be seen, the marginal effects found in this alternative specification are qualitatively identical and similar in magnitude to those found in the first column.

We now turn to our second testable implication, which is that the wage increase due to promotion decreases with education. Letting $\ln w_{it}$ denote the natural logarithm of worker i's real annual salary as measured on the last day of year t, we consider the following regression specification.[24]

(8a) $$\ln w_{it} - \ln w_{it-1} = \gamma_0 + \gamma_1 HS_{it-1} + \gamma_2 MA_{it-1} + \gamma_3 PHD_{it-1} + \mathbf{Y_{it-1}}\boldsymbol{\lambda} + \omega P_{it-1} + \varepsilon_{it} \qquad \text{if } PROMOTION_{it}=1$$

(8b) $$= \alpha_0 + \alpha_1 HS_{it-1} + \alpha_2 MA_{it-1} + \alpha_3 PHD_{it-1} + \mathbf{Z_{it-1}}\boldsymbol{\rho} + \rho P_{it-1} + \upsilon_{it} \qquad \text{if } PROMOTION_{it}=0$$

The prediction that the wage increase due to promotion decreases with education refers to the wage increase relative to what the worker would have received in the absence of a promotion. In terms of our regressions, this wage premium is given by (8a) minus (8b). The theoretical prediction to be tested is therefore $\gamma_3-\alpha_3<\gamma_2-\alpha_2<0<\gamma_1-\alpha_1$. In regression (8a), $\mathbf{Y_{it-1}}$ is the same vector of controls included in our promotion probability test given in equation (7) except for the substitution of job-title transition dummies

[23] In both tests the difference between the marginal effects for masters degree holders and Ph.D.s is statistically significant at the five percent level. Also, our basic test yields a probability of promotion of 0.206 for bachelors degree holders and 0.260 for masters degree holders. This is the source for the "about twenty percent" statement in the Introduction for the decrease in promotion probability when education is decreased from masters degree to bachelors degree. Note that we have also conducted this test on the full sample and there was no change in the qualitative nature of the results.

[24] One discrepancy between the theory and the empirical testing is that in the theory the promotion process is deterministic. In other words, in contrast to our empirical testing, if a worker is promoted in equilibrium, there is no similar worker with whom to compare the promoted worker who is not promoted in equilibrium. Our empirical methodology relies on such comparisons. However, although we do not formally show it here, it is possible to introduce "slot constraints" into our theoretical framework with the result that little is changed except that equilibrium behavior would allow for such comparisons.

for job-title dummies. These job-title transition dummies are of the form d_{jk} indicating a transition from job title j in year t-1 to job title k in year t. This is important because the job title to which a worker is promoted is likely to affect the wage change from promotion. In regression (8b), $\mathbf{Z_{it-1}}$ differs from $\mathbf{Y_{it-1}}$ in two ways. First, it includes job-title dummies instead of job-title transition dummies. Second, it includes individual-specific fixed effects which we necessarily omit from equation (8a) because educational attainment is typically time invariant during a worker's tenure with the firm, and we are interested in the relationship between education and wage growth.[25]

Since the theory predicts that both the percentage and absolute wage increases due to promotion should decrease with education, we also consider the following specification.

(9a) $$w_{it} - w_{it-1} = \zeta_0 + \zeta_1 HS_{it-1} + \zeta_2 MA_{it-1} + \zeta_3 PHD_{it-1} + \mathbf{Y_{it-1}}\boldsymbol{\eta} + \kappa P_{it-1} + \iota_{it} \qquad \text{if } PROMOTION_{it}=1$$

(9b) $$= \varsigma_0 + \varsigma_1 HS_{it-1} + \varsigma_2 MA_{it-1} + \varsigma_3 PHD_{it-1} + \mathbf{Z_{it-1}}\boldsymbol{\chi} + \sigma P_{it-1} + o_{it} \qquad \text{if } PROMOTION_{it}=0$$

In this test everything is the same as before except the dependent variables are differences in wage levels as opposed to differences in the logarithm of wage levels, where the theoretical prediction is now $\zeta_3-\varsigma_3<\zeta_2-\varsigma_2<0<\zeta_1-\varsigma_1$.

The first two columns of Table 5 report results of our estimation of equations (8a) and (8b) for the basic sample. As seen in the lower panel of the table, the results support the theoretical prediction, i.e., we find $\gamma_3-\alpha_3<\gamma_2-\alpha_2<0<\gamma_1-\alpha_1$, although our estimate of $\gamma_1-\alpha_1$ is statistically insignificant. The point estimates suggest that, other things equal, the wage premium from promotion is 2.3 percentage points lower for masters degree holders than bachelors degree holders, 4.1 percentage points lower for Ph.D.s than for bachelors degree holders, and 0.4 percentage points higher for high school graduates than for bachelors degree holders.[26] The estimate of $\gamma_3-\alpha_3$ is significant at the ten percent level, and the estimate of $\gamma_2-\alpha_2$ is significant at the one percent level.

---

[25] Another way to understand this test is as follows. One can construct for each observation of a promoted worker a wage increase due to the promotion using the following three-step procedure. First, estimate regression (8b) for the subsample of observations where promotion did not occur. Second, for each observation in which the worker was promoted, use the parameter estimates from the first step to derive a predicted no-promotion wage increase. Third, subtract this predicted wage increase from the actual wage increase. After this construction, one can use this "wage increase from promotion" variable as the dependent variable in a regression that has the same independent variables as in (8a). The coefficients on the education dummy variables in this regression would be identical to the estimates of $\gamma_1-\alpha_1$, $\gamma_2-\alpha_2$, and $\gamma_3-\alpha_3$ found in Table 5. Note that including individual-specific fixed effects in (8b) allows us to more accurately estimate the "predicted wage increases."

[26] In this analysis the difference between masters degree holders and Ph.D.s, although it has the correct sign, is not statistically significant (Z=0.782). Also, this test yields an average percentage wage increase due to promotion of 0.053 for bachelors degree holders and 0.031 for masters degree holders. This is the source for the "over seventy

To test the robustness of these findings, we also estimated the wage-growth regression using the less stringent selection rule concerning job titles. The results, which are reported in the third and fourth columns of Table 5, are qualitatively similar to those for our main test. The estimated values for γ_1-α_1,, γ_2-α_2, and γ_3-α_3, respectively, are 0.493, -1.723, and -3.363. As was true for our main test, the estimate of γ_1-α_1 is statistically insignificant, whereas the estimate of γ_2-α_2 is significant at the one percent level, and the estimate of γ_3-α_3 is significant at the ten percent level. We also estimated the model on our basic subsample controlling for performance in years t-1 and t-2. As reported in the fifth and sixth columns of Table 5, these results are also qualitatively similar to those found in the first two columns. One difference is that the estimate of γ_1-α_1, which was positive and insignificant in our main analysis, becomes negative and insignificant when we include performance in t-1 and t-2.[27]

One possible explanation for the weak empirical support in Table 5 concerning high school graduates and our second theoretical prediction is related to the fact that our data set only contains the white collar part of the labor force (see footnote 17). Many of the high school graduates that we observe in our data set likely started their careers at the firm in a blue collar job and were promoted into a white collar job (this is likely much less frequent for the other education groups). Although we control as much as is feasible for the nature of the job by including controls for job level and job title transition, it is possible that workers being promoted into managerial positions out of the blue collar part of the firm are on systematically different career tracks for which promotions are associated with smaller wage increases.

In Table 6 we reproduce all the tests in Table 5 except we estimate equations (9a) and (9b) rather than (8a) and (8b), i.e., the dependent variables are defined in terms of wage levels rather than log wages. All but one of the differences in coefficients at the bottom of the table have the predicted sign. In contrast to the Table 5 results, however, only the estimated values for ζ_3-ς_3 are statistically significant, where for each test ζ_3-ς_3 is statistically significant at the one percent level.[28]

---

percent" statement in the Introduction for the increase in the average percentage wage increase due to promotion when the education level is decreased from masters degree to bachelors degree.

[27] We have also conducted our basic test using the full sample. There were two qualitative differences between results for this test and the results reported in columns 1 and 2. First, as in the fifth and sixth columns of Table 5, in this test γ_1-α_1 is negative and insignificant. Second, in this test γ_3-α_3, although it does have the correct sign, is greater than rather than less than γ_2-α_2, and further it is not statistically significant at standard significance levels.

[28] Conducting the basic test in Table 6 on the full sample yields results qualitatively the same as those found in columns 1 and 2 of Table 6.

We now turn to our third testable implication which is that the first two testable predictions should hold more strongly for workers receiving their first promotion than for workers receiving a subsequent promotion. In the first two columns of Table 7 we re-estimate the models in the first two columns of Table 4 for our "first promotion" subsample. This subsample consists of observations for which either the worker has yet to receive a promotion or has just been promoted. Comparing Tables 4 and 7 provides support for our third testable implication. Specifically, the marginal effects for high school graduates and masters degree holders are higher in absolute value for the first promotion subsample for each of the two tests. The evidence for Ph.D.s is more mixed. In the column 2 test the marginal effect for Ph.D. holders is higher for the "first promotion" subsample, while it is slightly lower in the column 1 test.

In the third and fourth columns of Table 7 we re-estimate the models from the first two columns of Table 4 for our "subsequent promotion" subsample. This subsample consists of observations for which the worker has previously received at least one promotion. Comparing columns 1 and 2 of Table 7 with columns 3 and 4 again provides support for our third testable implication. The marginal effects for high school graduates and masters degree holders are higher in absolute value for the first promotion subsample for each of the two tests. Also, similar to what was true in the comparison of Tables 4 and 7, in the columns 2 and 4 test the marginal effect for Ph.D. holders is higher for the first promotion subsample, while in the columns 1 and 3 test the marginal effect for Ph.D. holders is roughly the same across the two subsamples.[29]

We now consider the third testable implication as it relates to wage changes. Because few workers in our sample transition from being high school graduates to college graduates during the sample period, or from masters degree holders to Ph.D.s, when we estimate our wage change models on subsamples, one or both of these education variables is frequently dropped in the estimation due to collinearities. For this reason, when we estimate these models we use only workers who hold either a bachelors degree or a masters degree.

[29]One possibility concerning why the marginal effects for Ph.D.s do not vary systematically across the subsamples is that some Ph.D.s are hired into positions where the first few promotions are close to automatic. For workers hired into such positions signaling can be unimportant even beyond the first promotion which is consistent with the large and significantly positive marginal effect for Ph.D.s in Table 7.

Table 8 reports these results.[30] Columns 1 and 2 report results for the estimation of (8a) and (8b) for the "first promotion" subsample, while columns 3 and 4 report results for the estimation of (8a) and (8b) for the "subsequent promotion" subsample. Columns 5 and 6 report results for the estimation of (9a) and (9b) for the "first promotion" subsample, while columns 7 and 8 report results for the estimation of (9a) and (9b) for the "subsequent promotion" subsample. As indicated, all tests drop observations where the worker is either a high school graduate or holds a Ph.D. degree.[31] The results exactly match our third testable implication. In each case γ_2-α_2 or ζ_2-ς_2 has the predicted sign and is significant at least at the five percent level for the "first promotion" subsample, while γ_2-α_2 or ζ_2-ς_2 has the wrong sign and/or is not significant at the five percent level for the "subsequent promotion" subsample.

IV. ALTERNATIVE EXPLANATIONS

This section investigates a number of potential alternative explanations for the results found in the previous section. We consider five alternatives: i) education provides higher-level skills; ii) symmetric learning; iii) coarse information about performance; iv) biased performance ratings; and v) biased promotion contests. We find that the promotion-as-signal hypothesis matches the evidence better than any of the alternatives.[32]

Alternative 1: Education Provides Higher-Level Skills

One potential explanation for why education matters in promotion decisions even after controlling for performance is that education provides skills more useful at higher levels of the firm's job ladder. For example, consider two workers equally productive at job level 1, where one worker has a bachelors degree and the other an MBA. Suppose further that having an MBA provides a worker with skills that are only useful on jobs at level 2 or higher. Then, even though performance on level 1 is the

---

[30] All tests reported in Table 8 consider subsamples of our basic sample.

[31] Results are similar if, instead of restricting the sample to those workers with either a bachelors or a masters degree, we use all observations but with only one binary education variable on the right-hand side, defined as 1 if the worker has a masters degree or Ph.D. and 0 if the worker has a high school education or bachelors degree.

[32] An important perspective for thinking about promotions is that promotions are used to provide incentives for effort such as in Lazear and Rosen (1981), Rosen (1986), and MacLeod and Malcomson (1988). Biased promotion contests, i.e., alternative 5, is a variant of this argument that serves as an alternative explanation for some of our findings. In general, however, most models of this sort do not generate predictions consistent with our findings. For example, in most of this literature either workers are *ex ante* identical or the prize does not vary with worker attributes and such models cannot explain our finding that the wage increase due to promotion decreases with the worker's education level. See the Conclusion for a related discussion.

same, the firm will have a greater incentive to promote the MBA because his or her expected performance on level 2 is higher.

The first problem with this alternative explanation concerns our second main empirical finding. That is, the idea that education provides higher-level skills is inconsistent with our finding that, except for high school graduates, the wage increase due to promotion decreases with education. If the role of education in the promotion process is to provide higher-level skills, wage increases due to promotion should be positively, not negatively, related to education. The reason is that, because in that explanation more highly educated workers have skills that are more highly valued at high-level jobs, such workers should get particularly large wage increases upon promotion, since promotion for these workers is associated with particularly large increases in productivity.

Furthermore, one way to address this alternative explanation is to include predicted post-promotion performance as an explanatory variable in our promotion-probability tests. That is, suppose we estimate how workers with various education levels are expected to perform after promotion and include these predictions as an explanatory variable in our probit analysis. Then, if this alternative explanation is correct, education should become less important in the promotion decision since it matters only because it translates into higher expected post-promotion performance.

To pursue this idea, we first constructed a measure of predicted post-promotion performance as a function of the education level and controls. To do this we considered the subsample of promotions in each year t and estimated an ordered probit in which performance in year t+1 was the dependent variable and the independent variables were the right-hand-side variables in equation (7). Then for each observation (including both promotions and non-promotions) we used the resulting estimates to compute predicted probabilities for each of the five possible performance outcomes in year t+1. Denoting these predicted probabilities as $p_1, p_2, \ldots, p_5$, our estimate of predicted post-promotion performance conditional on the information in year t-1 is $\sum_{k=1}^{5} kp_k$. We then included this measure of predicted performance as an additional control in the probit equation for promotion in year t and the results were very similar to our main results in Table 4. That is, as reported in the last column of the table, the marginal effects for each education dummy variable is similar in magnitude and has a similar level of statistical significance after

including expected post-promotion performance. This leads us to again reject the alternative explanation for our results that higher education provides higher level skills.[33]

Alternative 2: Symmetric Learning

As is discussed in Section V, symmetric learning refers to a situation in which learning comes from publicly available information so at any date all firms have the same information and beliefs about each worker's ability. As an example, the benchmark analysis of Section II is a symmetric learning model since in each period each worker's output is publicly observed rather than privately observed by the worker's current employer.

Symmetric learning when output is deterministic (as in Section II's benchmark analysis) does not provide an explanation for our promotion probability findings, but symmetric learning can explain these findings when output is stochastic. With deterministic output the most recent output observation should contain all the relevant information about a worker's ability. So in such a world, if performance ratings are available, there is no reason for a firm to partially base promotion decisions on education levels.

But suppose there is symmetric learning and output is stochastic. Then a worker's most recent performance rating will not contain all the relevant information concerning the worker's true ability. Rather, previous performance ratings and the worker's education level (to the extent education and ability are positively correlated) can also contain relevant information. Hence, one possible explanation for why we find a positive relationship between education and probability of promotion is that, even after controlling for current performance, higher education implies higher expected ability.

Symmetric learning can also explain our wage growth findings. This follows from an argument related to one found in Gibbons and Waldman (1999a). That paper shows that symmetric learning can explain large wage increases upon promotion because, on average, promoted workers should be those for whom there were large improvements in beliefs concerning the worker's underlying abilities. Now consider symmetric learning when workers vary in their education levels and higher education translates on average into higher underlying ability. In such a setting it would be natural for promoted workers

[33] Note that we also considered whether education predicts post-promotion performance after including all of our standard controls including pre-promotion performance. We investigated this using both OLS and ordered probit. The answer is that, except for high school graduates, education is positively related to post-promotion performance. But as discussed above, controlling for this generally positive relationship in our promotion probit analysis does not reduce the positive relationship between education and promotion probability.

from lower education groups to experience larger improvements, on average, in beliefs concerning the workers' underlying abilities. In turn, these larger improvements in beliefs should translate into larger promotion wage increases for these lower education groups.[34]

One test of this argument concerns incorporating multiple performance ratings into our promotion-probability analysis. That is, if the reason education and promotion probability are positively related is that a worker's education level provides valuable incremental information about a worker's true ability, then the coefficients on the education variables should fall in absolute value as more performance ratings are added to the probit analysis (see Altonji and Pierret (2001) for a related analysis). Note, this is true as long as the performance ratings themselves are only moderately positively correlated, so that adding more performance ratings adds important incremental information. In Table 9 we provide a bivariate correlation matrix for performance in period t and three lagged values for performance. Not surprisingly there is positive serial correlation in performance ratings, but the correlations are sufficiently below one that adding performance ratings to our probit analysis contributes important incremental information.

Table 10 reports results of this test for our basic sample (note that the first two columns of this table also appear in Table 4). As reported in Table 10, the positive relationship between education and promotion probability is still strongly statistically significant even when four years of performance ratings are included. More importantly, the results in Table 10 are not consisistent with the absolute magnitudes of the education coefficients falling as more performance ratings are added. For example, although the absolute value of the coefficient on the high school variable is lower in column 4 than in column 1, the absolute value of the MA coefficient is basically unchanged and the Ph.D. coefficient is actually higher in column 4 than in column 1. These findings do not support symmetric learning as the correct explanation for our findings.[35]

---

[34] Gibbons and Waldman assume that learning occurs at discrete intervals rather than in a continuous fashion. This is realistic if, for example, learning occurs when a worker completes a project and project completion occurs at discrete intervals.

[35] We have also conducted this test restricting the sample for each regression to the 1796 data points for which performance in years t-1 through t-4 are all available. There was no change in the qualitative nature of the results. In addition, we conducted this test on the first promotion subsample. Focusing on the signs and magnitudes of the coefficients the results are quite similar to those found in Table 10. However, because sample sizes are quite small when performance ratings from multiple years are included, standard errors are large in these estimations and as a result education coefficients are typically not statistically significant.

A related test is to include a predicted post-promotion performance variable into a probit analysis that includes multiple performance ratings. The symmetric-learning argument suggests that including predicted post-promotion performance should significantly reduce the effect of education on promotion probability. In the last column of Table 10 we add the predicted post-promotion performance variable from above into the probit analysis that employs four performance measures.[36] The coefficients on the education variables all continue to have the predicted sign and two of the three education coefficients continue to be strongly statistically significant. As before, the results do not support symmetric learning being the correct explanation for our findings.

Alternative 3: Coarse Information About Performance

Another potential explanation for why education matters in promotion decisions is that performance ratings are coarse measures of true performance. Remember that the performance rating is an integer value between 1 and 5. Saying that the performance rating is a coarse measure of true performance simply means that five categories do not capture fine gradations of performance. To see why this idea can potentially explain our promotion probability results, consider two workers who both receive a one – the highest rating – but one of the workers actually had a higher true performance so this worker's probability of promotion is in fact higher. If, as seems quite plausible, considering all such worker pairs the workers with higher true performance on average have higher education levels, then in our probit analysis education may serve as a proxy for the unmeasured higher true performance. In turn, given higher true performance positively affects the probability of promotion, this is an explanation for the positive relationship we find between education and probability of promotion even though we control for the performance rating.

One problem with this argument concerns our wage-growth findings. If the only reason that education is positively related to promotion probability is that higher education is proxying for unmeasured higher true performance, then we would expect education to be positively related to the "gross" wage increase associated with promotion (by gross wage increase we mean the wage increase that does not net out what the worker would have received in the absence of promotion). But the first

[36] We attempted to construct a measure of predicted post-promotion performance that used the ordered probit technique described above, but which employed performance ratings in periods t-1, t-2, t-3, and t-4 rather than just t-1 as above. However, in this case the probit estimation failed to converge.

columns of Tables 5 and 6 indicate that the gross wage increase due to promotion is, except for the high school group, decreasing with education rather than increasing. Note that the prediction for the net wage increase due to promotion, which was our focus in Section III, is unclear. The reason is that, according to the coarse-information argument, higher education (which means on average higher true performance) should increase both the wage increase upon promotion and the wage increase in the absence of promotion.

A second problem with this argument is that it does not explain why education should positively affect promotion probabilities more for first promotions than subsequent promotions. If higher education is proxying for higher true performance, then higher education should be correlated with higher promotion probabilities for both types of promotions.

Another approach for investigating the coarse-information argument involves including performance ratings from multiple periods in the probit analysis. If a single period's performance rating is a coarse measure of true performance, then in aggregate performance ratings from multiple periods should more accurately capture true performance. In turn, this means that if the coarse-information argument is what is driving our finding of a positive relationship between education and promotion probability, then including ratings from multiple periods should significantly reduce the positive relationship between education and promotion probability. As already discussed, Table 10 reports promotion-probability results when multiple performance ratings are included. Given the table shows that the absolute value of the coefficient on the HS variable becomes smaller as more performance ratings are added, there is some evidence consistent with coarse information mattering for the high school group. But given this relationship does not hold for the other education groups, given the coefficients of interest are still statistically significant even when four performance ratings are included, and given our point in the above paragraph that this explanation does not match our other empirical findings, we do not believe that coarse information is the correct explanation for our empirical results.

Alternative 4: Biased Performance Ratings

Biased performance ratings serve as a potential explanation for our promotion probability findings.[37] Suppose that each worker's performance is measured relative to workers with the same level

[37] Gibbons and Waldman (1999a) argue that biased performance ratings may explain results found in Medoff and Abraham (1980,1981). Their argument relies on ratings being measured relative to firm tenure rather than being

of education and that, on average, higher levels of education translate into higher levels of performance. Now consider two workers with the same performance rating but different levels of education. Given each worker is evaluated relative to the average performance for workers with the same education level, the worker with higher education is likely to have a higher true performance level and thus, as we find, should also have a higher probability of promotion. Note that earlier we provided evidence suggesting that performance ratings were not measured relative to firm tenure or job level. Unfortunately, there is no similar test to perform that would show performance evaluations are not measured relative to the worker's education level.

There is, however, a problem with this argument which is the same as one of the problems discussed above concerning the coarse-information argument. That is, if the reason education is positively related to promotion probabilities even after controlling for the performance rating is that higher education captures unmeasured higher true performance, then higher education should be associated with higher promotion probabilities for both first promotions and subsequent promotions. As Table 7 shows, however, except for Ph.D. holders, the evidence for this relationship is weak for subsequent promotions.

Alternative 5: Biased Promotion Contests

Meyer (1991) investigates a T-period tournament model in which there are two workers and a promotion decision at the end of the T periods. Her basic point is that, if one focuses on performance in period T, then the firm should bias its decision rule so that the worker who was more productive in the first T-1 periods is promoted after period T even if this worker's period-T output is a little lower than the other worker's output. The logic is that when output is stochastic, each period's output is informative of worker ability so promotions are more efficient when outputs across many periods are considered rather than being determined solely by a single-period's output. Note that this argument is just a variant of the symmetric-learning explanation discussed above. Hence, consistent with that discussion, we believe the results in Table 10 indicate that this argument is unlikely to be the correct explanation for our findings.

---

absolute measures of performance, so the results reported in Table 3 throw some doubt on their argument. See Bernhardt (1995) for an alternative explanation of the Medoff and Abraham findings that relies on the same ideas of asymmetric information and promotions serving as signals investigated here.

Meyer (1992) puts forth an alternative argument for biased promotion tournaments. In that analysis there is a sequence of two contests and the firm biases the second contest in favor of the first contest's winner. The reason is that a small bias causes a second-order decrease in effort levels in the second contest but a first-order increase in effort levels in the first. This could potentially explain our promotion-probability findings. The idea here is that for promotions for workers starting at higher levels of the job ladder education may serve as a proxy for having been a "bigger" winner earlier on.

One problem with this explanation is that, given all the controls we include in our analysis, it is not clear whether this explanation is plausible. Nevertheless, ignoring this problem with the explanation, this argument does not hold for first promotions since there are no earlier promotion contests for which education could be serving as a proxy. But in Table 7 we found that education is more important for promotion probabilities when the sample is restricted to first promotions. We therefore again reject biased promotion contests as an explanation for our results.

V. DISCUSSION

In this section we discuss two issues. In the first subsection we discuss how our analysis contributes to the literature on learning in labor markets. In the second we consider how our analysis contributes to the extensive literature on education as a labor-market signal.

A) Symmetric Versus Asymmetric Learning

From a theoretical perspective, there are two basic approaches for modeling learning in labor markets. One approach, investigated in papers such as Harris and Holmstrom (1982) and Gibbons and Waldman (1999a), is that learning is symmetric, i.e., any information revealed about a worker's ability during the worker's career is public knowledge. The other approach, first investigated in Waldman (1984a) and Greenwald (1986), is that learning is asymmetric. That is, information about a worker's ability is only directly revealed to the worker's current employer, while other firms observe the current employer's actions such as promotion decisions and firing decisions in making inferences about the worker's ability.

A substantial empirical literature investigates learning in labor markets. Studies such as Gibbons and Katz (1992), Farber and Gibbons (1996), and Altonji and Pierret (2001) focus on symmetric learning, while Gibbons and Katz (1991) and our own study focus on asymmetric learning. The former papers find

evidence consistent with symmetric learning, while the latter find evidence consistent with asymmetric learning. Schonberg (2007) criticizes this literature for paying insufficient attention to developing testable implications that distinguish between the two types of learning. She derives such implications in the context of a specific model of labor-market turnover and then presents empirical evidence consistent with university graduates being characterized by asymmetric learning and high school graduates and dropouts being characterized by symmetric learning.[38]

Although we find Schonberg's analysis interesting, we disagree with her idea that for any group labor-market learning is primarily either symmetric or asymmetric. We believe the evidence clearly suggests that even within groups learning is somewhere between the pure symmetric and pure asymmetric cases. For example, as argued in Gibbons and Waldman (1999a), a pure asymmetric-learning model cannot easily explain the empirical findings of Baker, Gibbs, and Holmstrom (1994a,b). In particular, Baker, Gibbs, and Holmstrom find that, on average, workers receive wage increases even in periods in which they are not promoted and these wage increases vary across workers. This is difficult to reconcile with a pure asymmetric-learning model, but is easily captured in a model characterized by some symmetric learning. On the other hand, as we argued in previous sections, other aspects of the Baker, Gibbs, and Holmstrom data set suggest that asymmetric learning is also important.

To see what we have in mind, consider, for example, the manager of a large division of a Fortune 500 firm. To the extent everyone observes the overall success of the division's products, some learning about the manager's ability is of a symmetric or public nature. But the firm's CEO has access to other information relevant for judging the ability of the division head. The CEO can look at the details of the division's accounting numbers which are typically not available outside the firm. The CEO can also judge better than other potential employers the extent to which the division's success or failure is due to the division head.

Consistent with the discussion in this section, we believe an interesting direction for future research is to consider analyses characterized by a mix of symmetric and asymmetric learning and other

---

[38] To be more precise, Schonberg develops two testable implications for asymmetric learning. The first which she finds clear evidence for among university graduates is the result originally due to Greenwald that workers who separate should be drawn from the low end of the ability distribution. The second is that the return to ability should be positively related to firm tenure. For this prediction she finds mixed evidence among university graduates. However, although the second testable implication is predicted by the specific model of asymmetric learning that she considers, we do not believe it is a robust prediction of such models. So we believe that Schonberg's findings overall are consistent with asymmetric learning being important for university graduates.

intermediate cases between the two polar cases. One promising direction along this line is suggested by Granovetter (1973,1995) who focuses on how hires frequently occur through personal connections. This suggests that much of the learning in the labor market is neither symmetric nor asymmetric, but rather something between the two where some information about any specific worker leaks out from the worker's current employer but only in a limited way.[39]

B) Education as a Labor-Market Signal

There is an extensive literature that investigates education as a labor-market signal (see Riley (2001) for a survey). Here we discuss how our analysis contributes to this literature. We start with theory. The seminal theoretical paper in this literature is Spence (1973). In that paper education does not result in human-capital accumulation, but rather workers with higher ability have a lower cost of acquiring education. In turn, since education does not directly contribute to productivity, the socially-optimal level of education is zero. But Spence shows there are equilibria – in fact, many equilibria – characterized by positive levels of education. Note that Spence did not analyze his model as a formal game, and much of the theoretical literature that followed, such as Cho and Kreps (1987) and Riley (1979a), focuses on how signaling works in a one-period setting when a formal game-theoretic approach is employed.

From our perspective the more interesting theoretical extension is Riley's (1979b) analysis which considers careers that last more than a single period. Riley shows that, as a worker's career progresses and firms learn about true ability, the importance of the initial education signal as a factor determining compensation should decrease. Note that Altonji and Pierret's (2001) analysis of symmetric learning mentioned in the previous subsection is similar. They argue that, as careers progress and firms observe output realizations, education becomes less important and true ability more important as factors determining compensation.

[39] See Montgomery (1991) for a theoretical analysis along this line, and more recently Kahn (2009) and Pinkston (2009) conduct theoretical and empirical analyses that allow for both symmetric and asymmetric learning. Also, DeVaro (2008) estimates a structural model of employer recruitment choice in which hires occur via either personal connections or formal methods such as advertising. The focus in that analysis is on the role of recruitment strategies as information-generating devices in the labor market and, although the model does not distinguish between symmetric and asymmetric learning, that is one direction in which the framework might be extended.

Although we do not formally treat education as a signal since education levels are given exogenously in our analysis, it is possible to extend our analysis in this way.[40] The main result of such an extension would be that, in contrast to Riley (1979b) and Altonji and Pierret (2001), because learning is asymmetric after workers enter the labor market the importance of education signaling in determining compensation would not necessarily be a decreasing function of labor-market experience. That is, since education would increase promotion probabilities, education could be an important factor in compensation even late in careers because mostly only old workers with high ability and high education earn promotions to the top rungs of firms' job ladders.

A related point concerns the strength of workers' incentives to invest in education as a signal. In the standard multi-period education signaling model (which is the basis for recent estimates of the return to education signaling found in Kaymak (2007) and Lange (2007)), the return to signaling is the higher compensation workers receive early in careers prior to firms learning the worker's true ability. One of our points is that, if there is asymmetric learning after workers enter the labor market, then the return to education signaling is not just higher wages early in careers but also higher wages later in careers because, even after controlling for productivity, more highly educated workers should have higher probabilities of eventually reaching the top rungs of firms' job ladders. In other words, in a world with asymmetric learning and promotions, the return to education signaling is enhanced relative to a world where promotions are not a factor.

Now we turn to the empirical evidence. There is an extensive empirical literature on this topic, but much of it is subject to the criticism that the testable implications considered are also consistent with a world in which education simply serves to enhance human-capital accumulation. For example, Layard and Psacharopoulos (1974), Hungerford and Solon (1987), and Heywood (1994) all investigate whether earning an educational degree enhances compensation as the signaling story would suggest, but a world where education serves a pure human-capital-accumulation role makes the same prediction if individuals who drop out before earning a degree do so because they realize the education is not providing them with valuable human capital.

---

[40] The discussion that follows is related to recent analyses that appear in Ishida (2004a,b) concerning the promotion-as-signal hypothesis. See also Habermalz (2006) for a related analysis that does not incorporate the promotion-as-signal hypothesis.

Papers that are not subject to this criticism include Riley (1979b), Lang and Kropp (1986), and Altonji and Pierret (2001). Riley develops predictions based on the idea that some sectors rely on education as signals and some do not, and finds supporting evidence using the *Current Population Survey*. Lang and Kropp develop predictions concerning how an increase in educational levels at the bottom of the ability distribution affects educational attainment at the top through the operation of incentive compatibility constraints, and finds supporting evidence using changes in mandatory minimum education levels as the exogenous change affecting educational attainment at the bottom of the distribution. Finally, Altonji and Pierret find evidence consistent with the earlier theoretical discussion that the importance of education as a factor in compensation should decline as workers age.

Our empirical work adds to the existing evidence in favor of education having a signaling role. The theoretical extension we describe above would yield as a testable implication that, as we find in our empirical work, when there is little firm-specific human capital education improves promotion prospects even holding as fixed both the worker's current performance on the low-level job and his or her expected performance on the high-level job. Further, as we also find, the theory would predict that it is more the attainment of a degree rather than simply years of education that increases the probability of promotion (see footnote 19). Hence, as indicated, our empirical results are consistent with a world in which education has a signaling role.[41]

VI. CONCLUSION

An extensive theoretical literature argues that promotions serve as signals of worker ability. In this paper we first extended the theoretical literature on this topic by incorporating education into a standard model of the promotion-as-signal hypothesis and then derived three testable implications. The theory is based on the intuitively plausible idea that the signal associated with a promotion is more

[41] One could also construct a model yielding similar predictions that is characterized by uncertainty concerning the innate abilities of workers, but in which education serves a purely human-capital-enhancing function. Note, however, that for education not to serve as a signal given uncertainty concerning innate abilities, there would need to be no correlation between innate abilities and education levels. Although theoretically possible by, for example, having heterogeneous schooling levels arise solely from differences across individuals in their access to capital markets, such a model would not be a plausible description of the real world. In the real world innate ability, along with other factors such as family wealth and access to capital markets, is clearly an important determinant of educational attainment. In other words, our results are consistent with education having a signaling role, and we believe there is no plausible alternative in which education does not have a signaling role.

important for workers with low education levels. We then investigated these predictions using a data set covering the internal-labor-market history of a single medium-sized firm in the financial-services industry over a twenty-year period. Our empirical investigation supports the three predictions. First, holding performance fixed, the probability of promotion increases with a worker's education level. Second, except for high school graduates, wage increases due to promotion fall with educational attainment. Third, the first two predictions hold more strongly for first promotions than for subsequent promotions.

One interesting implication of our analysis concerns the correct way to model labor-market tournaments. The traditional approach, as found in the seminal analysis of Lazear and Rosen (1981), is that the firm commits to a promotion prize that is independent of the identity of the promoted worker. In contrast, we find that promotion prizes vary with the characteristics of the promoted worker. For example, early in careers promotion prizes are smaller for workers with more education. This suggests that the correct way to model labor-market tournaments may be by having the signal endogenously determine the size of the prize as explored in Zabojnik and Bernhardt (2001) (see also Gibbs (1995)). In that approach, firms do not commit to a wage structure in advance. Rather, workers provide effort or invest in human capital in order to increase the probability of being promoted and earn the higher wage associated with the resulting signal.

There are a number of directions in which this paper's analysis can be extended. First, we could incorporate effort choice or human-capital investment into our theoretical model and see whether this enrichment yields additional testable predictions. Second, we could investigate the validity of our theoretical predictions for other firms for which data are available. Third, we could investigate how our theoretical predictions hold up in a cross-section of firms and industries. For example, in industries such as academia where asymmetric information is less important because publication records serve as publicly observable measures of performance, our theory predicts that the testable predictions we have focused on should be less important.[42] Fourth, we could enrich the theoretical model by allowing turnover and then empirically investigate any new predictions that result.[43]

---

[42] The specific prediction is that, holding fixed the publication record, the quality of the Ph.D. granting institution should have a relatively small effect on the likelihood of being granted tenure and the wage premium associated with tenure. This should be easily testable using the type of data found in Coupe, Smeets, and Warzynski (2006).
[43] One prediction of such an extension is that individuals who separate from a firm will, on average, be individuals whose productivity is below the average productivity of individuals who look observationally equivalent but do not separate. We have conducted preliminary analysis of this prediction and found empirical support, i.e., after controlling for a number of observables the probability of turnover is negatively related to the performance rating.

APPENDIX

Proof of Proposition 1: As indicated in footnote 10, our focus is the unique equilibrium in which no workers are fired. We start with what happens when a worker is old. Consider wages. Because the initial employer can make counteroffers and because there is a small probability the initial employer will mistakenly not make a counteroffer when the initial employer has the smallest cost of committing that mistake, other firms are willing to offer a worker assigned to job j the worker's minimum possible output at one of these other firms which is based on who the initial employer assigns to job j in equilibrium. Further, since there is firm-specific human capital, a worker's first-period employer matches this offer whenever the firm does not mistakenly fail to make a counteroffer and when this is the case the worker remains with the first-period employer by assumption.

Now consider job assignments. Since output on job 2 rises faster with on-the-job human capital than output on job 1, for each schooling group S there must be a value $\eta^+(S)$ such that old worker i in schooling group S is assigned to job 1 (job 2) if $\eta_{it}<(\geq)\eta^+(S)$ (see footnote 9). In turn, given the above discussion concerning wages, the wage paid to a worker in schooling group S assigned to job 1 (job 2) is given by $d_1+c_1[\varphi_L+B(S)]f(1)+G(S)$ ($\max\{d_1+c_1\eta^+(S)+G(S),d_2+c_2\eta^+(S)+G(S)\}$).

Now consider $\eta^+(S)$ for a specific value S. Suppose $[\varphi_L+B(S)]f(1)<\eta^+(S)<[\varphi_H+B(S)]f(1)$. Then $\eta^+(S)$ is the value for η_{it} such that a firm is indifferent between assigning an old worker to jobs 1 and 2. In this case $\eta^+(S)$ satisfies (A1).

(A1) $$(1+k)[d_1+c_1\eta^+(S)]-[d_1+c_1(\varphi_L+B(S))f(1)]=(1+k)[d_2+c_2\eta^+(S)]-\max\{d_1+c_1\eta^+(S),d_2+c_2\eta^+(S)\}$$

Suppose $\eta^+(S)=\eta'$. Then (A1) reduces to $d_1+c_1(\varphi_L+B(S))f(1)=d_1+c_1\eta^+(S)$, which contradicts $\eta^+(S)=\eta'$. Suppose $\eta^+(S)<\eta'$. Then (A1) reduces to (A2).

(A2) $$(1+k)[d_2+c_2\eta^+(S)]-(1+k)[d_1+c_1\eta^+(S)]=[d_1+c_1\eta^+(S)]-[d_1+c_1(\varphi_L+B(S))f(1)]$$

But if $\eta^+(S)<\eta'$, then the left-hand side of this expression is strictly negative while the right-hand side is positive so we have a contradiction. Thus, if $[\varphi_L+B(S)]f(1)<\eta^+(S)\leq[\varphi_H+B(S)]f(1)$ for all S, then $\eta^+(S)>\eta'$ for all S and this, in turn, means $\max\{d_1+c_1\eta^+(S),d_2+c_2\eta^+(S)\}=d_2+c_2\eta^+(S)$ for all S.

Now suppose $\eta^+(S)=[\varphi_L+B(S)]f(1)$. Consider the return to promoting a worker whose value for η_{it} is given by $\eta_{it}=[\varphi_L+B(S)]f(1)+\gamma$, where γ is small. The extra productivity associated with such a promotion equals $[d_2+c_2[(\varphi_L+B(S))f(1)+\gamma]]-[d_1+c_1[(\varphi_L+B(S))f(1)+\gamma]]$ which is strictly negative for γ close

---

But note that this result could be explained by Jovanovic's (1979) job-search model where worker-firm matches are an experience good.

to zero. Starting from a situation in which $\eta^+(S)=[\varphi_L+B(S)]f(1)$, when the off-the-equilibrium path event of a worker not being promoted is observed by the market the inference is that the worker's on-the-job human capital is $[\varphi_L+B(S)]f(1)$ (this follows from our assumption concerning off-the-equilibrium path promotion decisions – see footnote 7). The extra cost of promoting such a worker is therefore zero. Thus, since the extra productivity of promoting such a worker is less than the extra cost, the firm will not want to promote the worker so we have a contradiction. Hence, $\eta^+(S)>\eta'$ for all S.

Now consider young workers. Given that from above we know that a firm earns positive expected profits from an old worker it employed when young, competition across firms means that the wage for young workers must exceed expected productivity. We also know that, given our assumption $\theta^E(N)f(0)<\eta'$, all young workers are assigned to job 1. Combining this with young workers being paid more than expected productivity yields $w_Y(S)>d_1+c_1(\theta^E(S)f(0))+G(S)$ for all S.

<u>Proof of Corollary 1</u>: From the proof of Proposition 1, we know given there is a positive number of promotions for workers of schooling level S_1 that (A3) must be satisfied.

(A3) $$(1+k)[d_1+c_1\eta^+(S_1)]-[d_1+c_1(\varphi_L+B(S_1))f(1)]=(1+k)[d_2+c_2\eta^+(S_1)]-[d_2+c_2\eta^+(S_1)]$$

Rearranging yields (A4).

(A4) $$(1+k)[(d_2+c_2\eta^+(S_1))-(d_1+c_1\eta^+(S_1))]-[(d_2+c_2\eta^+(S_1))-(d_1+c_1(\varphi_L+B(S_1))f(1))]=0$$

From Proposition 1 we know that $\eta^+(S_1)>\eta'$, so there must be a value for $\eta^+(S_1)$ higher than η' for which the left-hand side of (A4) equals zero. Further, if $\eta^+(S_1)=\eta'$, then the left-hand side of (A4) is strictly negative. The derivative of the left-hand side of (A4) with respect to η^+ equals $(1+k)(c_2-c_1)-c_2$, which must be strictly positive given the two preceding sentences.

Now consider S_2. Given there is a positive number of promotions for workers of schooling level S_2, (A5) must be satisfied.

(A5) $$(1+k)[(d_2+c_2\eta^+(S_2))-(d_1+c_1\eta^+(S_2))]-[(d_2+c_2\eta^+(S_2))-(d_1+c_1(\varphi_L+B(S_2))f(1))]=0$$

Suppose $\eta^+(S_2)=\eta^+(S_1)$. Given $B(S_2)>B(S_1)$, a comparison of (A4) and (A5) yields that the left-hand side of (A5) is positive if (A4) is satisfied. In turn, given that $(1+k)(c_2-c_1)-c_2>0$ and that (A5) is satisfied, it must be that $\eta^+(S_2)<\eta^+(S_1)$.

We now consider the relationship between $y^P(S_2)$ and $y^P(S_1)$. Subtracting (A5) from (A4) and rearranging yields (A6).

(A6) $$\eta^+(S_1)-\eta^+(S_2)=c_1[B(S_2)f(1)-B(S_1)f(1)]/[(1+k)(c_2-c_1)-c_2]$$

Let k* be the value for k such that $[\eta^+(S_1)-\eta^+(S_2)]f(1)/c_1f(0)=G(S_2)-G(S_1)$. (A6) tells us that for any k<k* such that $(1+k)(c_2-c_1)-c_2>0$, $[\eta^+(S_1)-\eta^+(S_2)]f(1)/c_1f(0)>G(S_2)-G(S_1)$. By definition $y^P(S)=d_1+[c_1\eta^+(S)f(0)/f(1)]+G(S)$. Given $[\eta^+(S_1)-\eta^+(S_2)]f(1)/c_1f(0)>G(S_2)-G(S_1)$ for all k<k* such that $(1+k)(c_2-c_1)-c_2>0$, we now have that $y^P(S_2)<y^P(S_1)$ if k is sufficiently small.

Proof of Corollary 2: From Proposition 1, the wage increase due to a promotion as a function of the worker's schooling level is given by (A7).

(A7) $$\Delta w^P(S)=[d_2+c_2\eta^+(S)-w_Y(S)]-[[d_1+c_1(\varphi_L+B(S)]f(1)-w_Y(S)]$$

This can be rewritten as (A8).

(A8) $$\Delta w^P(S)=(d_2-d_1)+c_2\eta^+(S)-c_1[\varphi_L+B(S)]f(1)$$

We now have that $\Delta w^P(S)$ is decreasing in S for schooling groups with a strictly positive probability of promotion because B(S) is increasing in S and, from Corollary 1, $\eta^+(S)$ is decreasing in S.

The percentage wage increase due to promotion is given by (A9), where $w_O^{NP}(S)$ is the wage paid to a worker with schooling level S who is not promoted.

(A9) $$\Delta_{\%}w^P(S)=[\Delta w^P(S)/w_O^{NP}(S)]100$$

Since $\Delta w^P(S)$ is decreasing in S, (A9) tells us that $\Delta w_{\%}^P(S)$ will also be decreasing in S if $w_O^{NP}(S)$ is increasing in S. From above we have $w_O^{NP}(S)=[d_1+c_1(\varphi_L+B(S)]f(1)$. Since B(S) is increasing in S, $w_O^{NP}(S)$ is increasing in S.

REFERENCES

Acemoglu, D. and J. Pischke, "Why Do Firms Train? Theory and Evidence," *Quarterly Journal of Economics*, 113, 1998, 79-119.

Altonji, J. and C. Pierret, "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 116, 2001, 313-350.

Baker, G., M. Gibbs, and B. Holmstrom, "The Wage Policy of a Firm," *Quarterly Journal of Economics*, 109, 1994a, 881-919.

Baker, G., M. Gibbs, and B. Holmstrom, "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics*, 109, 1994b, 921-955.

Belzil, C. and M. Bognanno, "The Promotion Dynamics of American Executives," Mimeo, Temple University, 2005.

Bernhardt, D., "Strategic Promotion and Compensation," *Review of Economic Studies*, 62, 1995, 315-339.

Chang, C. and Y. Wang, "Human Capital Investment Under Asymmetric Information: The Pigovian Conjecture Revisited," *Journal of Labor Economics*, 14, 1996, 505-519.

Cho, I. and D. Kreps, "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 1987, 179-221.

Coupe, T., V. Smeets, and F. Warzynski, "Incentives, Sorting and Productivity Along the Career: Evidence from a Sample of Top Economists," *Journal of Law, Economics, and Organization*, 22, 2006, 137-167.

DeVaro, J., "The Labor Market Effects of Employer Recruitment Choices," *European Economic* Review, 52, 2008, 283-314.

Doiron, D., "Layoffs as Signals: The Canadian Evidence," *Canadian Journal of Economics*, 28, 1995, 899-913.

Farber, H. and R. Gibbons, "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 111, 1996, 1007-1047.

Gibbons, R., "Incentives and Careers in Organizations," in D. Kreps and K. Wallis, eds., *Advances in Economic Theory and Econometrics*, Cambridge University Press: New York, 1997, 1-37.

Gibbons, R. and L. Katz, "Layoffs and Lemons," *Journal of Labor Economics*, 9, 1991, 351-380.

Gibbons, R. and L. Katz, "Does Unmeasured Ability Explain Inter-Industry Wage Differentials," *Review of Economic Studies*, 59, 1992, 515-535.

Gibbons, R. and M. Waldman, "A Theory of Wage and Promotion Dynamics Inside Firms," *Quarterly Journal of Economics*, 114, 1999a, 1321-1358.

Gibbons, R. and M. Waldman, "Careers in Organizations: Theory and Evidence," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Volume 3, North Holland, 1999b, 2373-2437.

Gibbons, R. and M. Waldman, "Enriching a Theory of Wage and Promotion Dynamics Inside Firms," *Journal of Labor Economics*, 24, 2006, 59-107.

Gibbs, M., "Incentive Compensation in a Corporate Hierarchy," *Journal of Accounting and Economics*, 19, 1995, 247-277.

Golan, L., "Counteroffers and Efficiency in Labor Markets with Asymmetric Information," *Journal of Labor Economics*, 23, 2005, 373-393.

Granovetter, M., *Getting a Job: A Study of Contacts and Careers*, Harvard University Press: Cambridge, MA, 1973.

Granovetter, M., *Getting a Job: A Study of Contacts and Careers, 2nd ed.*, University of Chicago Press: Chicago, IL, 1995.

Greenwald, B., "Adverse Selection in the Labour Market," *Review of Economic Studies*, 53, 1986, 325-347.

Grund, C., "Stigma Effects of Layoffs? Evidence from German Micro-Data," *Economics Letters*, 64, 1999, 241-247.

Habermalz, S., "More Detail on the Pattern of Returns to Educational Signals," *Southern Economic Journal*, 73, 2006, 125-135.

Harris, M. and B. Holmstrom, "A Theory of Wage Dynamics," *Review of Economic Studies*, 49, 1982, 315-333.

Heywood, J., "How Widespread are Sheepskin Returns to Education in the U.S.?," *Economics of Education Review*, 13, 1994, 227-234.

Hungerford, T. and G. Solon, "Sheepskin Effects in the Returns to Education," *Review of Economics and Statistics*, 69, 1987, 175-177.

Ishida, J., "Education as Advertisement," *Economics Bulletin*, 10, 2004a, 1-8.

Ishida, J., "Signaling and Strategically Delayed Promotion," *Labour Economics*, 11, 2004b, 687-700.

Kahn, L., "Asymmetric Information Between Employers," Mimeo, Yale University, 2009.

Kaymak, B., "Signaling, Self-Selection, and Employer Learning," Mimeo, University of Rochester, 2007.

Krashinsky, H., "Evidence on Adverse Selection and Establishment Size in the Labor Market," *Industrial and Labor Relations Review*, 56, 2002, 84-96.

Jovanovic, B., "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87, 1979, 972-990.

Lang, K. and D. Kropp, "Human Capital versus Sorting: The Effects of Compulsory Attendance Laws," *Quarterly Journal of Economics*, 101, 1986, 609-624.

Lange, F., "The Speed of Employer Learning," *Journal of Labor Economics*, 25, 2007, 1-35.

Layard, R. and G. Psacharopoulos, "The Screening Hypothesis and the Returns to Education," *Journal of Political Economy*, 82, 1974, 985-999.

Lazear, E., "Raids and Offer Matching," in R. Ehrenberg, ed., *Research in Labor Economics, Vol. 8*, JAI Press: Greenwich, CT, 1986, 141-165.

Lazear, E., "The Job as a Concept," in W. Bruns, ed., *Performance Measurement, Evaluations, and Incentives*, Harvard University Press: Boston, MA, 1992, 183-215.

Lazear, E. and S. Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89, 1981, 841-864.

Lemieux, T., W.B. MacLeod, and D. Parent, "Performance Pay and Wage Inequality," *Quarterly Journal of Economics*, 124, 2009, 1-49.

MacLeod, W.B. and J. Malcomson, "Reputation and Hierarchy in Dynamic Models of Employment," *Journal of Political Economy*, 96, 1988, 832-854.

Medoff, J. and K. Abraham, "Experience, Performance, and Earnings," *Quarterly Journal of Economics*, 95, 1980, 703-736.

Medoff, J. and K. Abraham, "Are Those Paid More Really More Productive?," *Journal of Human Resources*, 16, 1981, 186-216.

Meyer, M., "Learning from Coarse Information: Biased Contests and Career Profiles," *Review of Economic Studies*, 58, 1991, 15-41.

Meyer, M., "Biased Contests and Moral Hazard: Implications for Career Profiles," *Annales d'Economie et de Statistique*, 25/26, 1992, 165-187.

Milgrom, P. and S. Oster, "Job Discrimination, Market Forces, and the Invisibility Hypothesis," *Quarterly Journal of Economics*, 102, 1987, 453-476.

Montgomery, J., "Social Networks and Labor Market Outcomes: Toward an Economic Analysis," *American Economic Review*, 81, 1991, 1408-1418.

Myerson, R., "Refinements of the Nash Equilibrium Concept," *International Journal of Game Theory*, 7, 1978, 73-80.

Owan, H., "Promotion, Turnover, Earnings, and Firm-Sponsored Training," *Journal of Labor Economics*, 22, 2004, 955-978.

Pinkston, J., "A Model of Asymmetric Employer Learning with Testable Implications," *Review of Economic Studies*, 76, 2009, 367-394.

Ricart i Costa, J., "Managerial Task Assignment and Promotion," *Econometrica*, 56, 1988, 449-466.

Ricchetti, B.M., "*Essays in Labor and Organizational Economics*, Unpublished Ph.D. Dissertation, 2007, Cornell University.

Riley, J., "Informational Equilibrium," *Econometrica*, 47, 1979a, 331-359.

Riley, J., "Testing the Educational Screening Hypothesis," *Journal of Political Economy*, 87, 1979b, S227-S252.

Riley, J., "Silver Signals, Twenty-Five Years of Screening and Signaling," *Journal of Economic Literature*, 39, 2001, 432-478.

Rosen, S., "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics*, 13, 1982, 311-323.

Rosen, S., "Prizes and Incentives in Elimination Tournaments," *American Economic Review*, 76, 1986, 701-715.

Schonberg, U., "Testing for Asymmetric Employer Learning," *Journal of Labor Economics*, 25, 2007, 651-691.

Selten, R., "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 1975, 25-55.

Seltzer, A. and D. Merrett, "Personnel Policies at the Union Bank of Australia: Evidence from the 1888-1900 Entry Cohorts," *Journal of Labor Economics*, 18, 2000, 573-613.

Seltzer, A. and K. Simons, "Salaries and Career Opportunities in the Banking Industry: Evidence from Personnel Records of the Union Bank of Australia," *Explorations in Economic History*, 38, 2001, 195-224.

Song, Y., "Recall Bias in the Displaced Workers Survey: Are Layoffs Really Lemons?," *Labour Economics*, 14, 2007, 335-345.

Spence, M., "Job Market Signaling," *Quarterly Journal of Economics*, 87, 1973, 355-374.

van Herpen, M., K. Cools, and M. van Praag, "Wage Structure and the Incentive Effect of Promotions," *Kyklos*, 59, 2006, 441-459.

Waldman, M., "Job Assignments, Signalling, and Efficiency," *Rand Journal of Economics*, 15, 1984a, 255-267.

Waldman, M., "Worker Allocation, Hierarchies, and the Wage Distribution," *Review of Economic Studies*, 51, 1984b, 95-109.

Waldman, M., "Up-or-Out Contracts: A Signaling Perspective," *Journal of Labor Economics*, 8, 1990, 230-250.

Waldman, M., "Theory and Evidence in Internal Labor Markets," Mimeo, Cornell University, 2008 (Forthcoming *Handbook of Organizational Economics*).

Zabojnik, J. and D. Bernhardt, "Corporate Tournaments, Human Capital Acquisition, and the Firm Size-Wage Relation," *Review of Economic Studies*, 68, 2001, 693-716.

TABLE 1: Distribution of Job Titles for Each Educational Group

	High School	College	M.A.	Ph.D.
Job Title A	9.22	4.48	1.32	0.00
Job Title B	0.35	1.02	0.45	0.00
Job Title C	***16.08***	***14.98***	***11.08***	***6.54***
Job Title D	***6.84***	***10.66***	***10.07***	***8.60***
Job Title E	15.49	6.83	4.26	0.93
Job Title F	0.84	1.03	0.25	1.12
Job Title G	***26.78***	***23.26***	***24.06***	***17.48***
Job Title H	0.18	0.80	0.67	0.00
Job Title I	5.76	2.49	1.43	0.00
Job Title J	0.00	0.11	0.00	0.00
Job Title K	***1.57***	***1.66***	***1.70***	***2.52***
Job Title L	0.22	0.64	0.75	0.47
Job Title M	0.00	0.01	0.07	0.47
Job Title N	0.69	1.71	0.59	3.83
Job Title O	0.60	3.37	3.66	8.32
Job Title P	0.81	1.16	0.47	3.36
Job Title Q	***14.57***	***25.78***	***39.17***	***46.36***
TOTAL	**100%**	**100%**	**100%**	**100%**

TABLE 2: Descriptive Statistics

	Mean	Std. Deviation
Promotion	0.131	0.337
HS Graduate+	0.370	0.483
College Graduate	0.376	0.484
MA	0.223	0.416
PhD	0.031	0.173
Performance	1.901	0.770
Age	42.168	9.371
Years at Company	6.123	3.793
Years at Title	3.847	2.836
Years at Level	3.855	2.837
Job Level 1	0.137	0.343
Job Level 2	0.146	0.353
Job Level 3	0.360	0.480
Job Level 4	0.357	0.479

Notes: Computed on subsample that: i) includes only job titles C, D, G, K, and Q; ii) omits workers with years of education equaling 15, 17, 19, or 20; iii) omits observations with missing performance data; iv) omits observations for which the history of job titles is incomplete over the worker's career at the firm.

TABLE 3: Performance Regressions (Fixed-Effect Logit Models)

Dependent Variable: *HighPerf* = 1 if Performance = 1 or 2
= 0 otherwise

	(1)	(2)	(3)	(4)
	y = $HighPerf_{it}$	y = $HighPerf_{it}$	y = $HighPerf_{it+1}$	y = $HighPerf_{it+1}$
Years at Company	0.014 (2.48)**	2.302 (21.65)***	•	2.510 (25.27)***
Years at Title	•	-0.135 (2.67)***	•	-0.086 (1.59)
Years at Level	•	-0.068 (1.27)	•	-0.035 (0.60)
Promotion (t-1)	•	•	0.476 (6.66)***	0.249 (2.38)**
Job Level Dummies (t-1)	NO	YES	NO	YES
Job Title Dummies (t-1)	NO	YES	NO	YES
Year Dummies	NO	YES	NO	YES
Sample Size	N = 11,104	N = 9098	N = 10,155	N = 6256

Notes: Cell entries are coefficiences from fixed-effect logit models, with Z-statistics in parentheses. Statistical significance at the 10%, 5%, and 1% levels denoted by *, **, and ***.

TABLE 4: Probit Marginal Effects for Probability of Promotion in Year t

	(1)	(2)	(3)	(4)
	Basic Sample[1]	**Less Stringent Sample[2]**	**Basic Sample[1]**	**Basic Sample[1]**
HS Graduate+	-0.059 (5.09)***	-0.036 (4.78)***	-0.053 (3.89)***	-0.063 (4.77)***
MA	0.052 (3.96)***	0.040 (4.36)***	0.031 (1.98)**	0.049 (3.51)***
PhD	0.157 (4.05)***	0.093 (3.64)***	0.154 (3.17)***	0.160 (3.82)***
Performance (t-1)	-0.052 (7.30)***	-0.044 (9.44)***	-0.040 (4.11)***	-0.046 (3.07)***
Performance (t-2)	•	•	-0.035 (3.81)***	•
Age (t-1)	-0.038 (7.96)***	-0.017 (5.39)***	-0.032 (5.36)***	-0.038 (6.40)***
Age squared (t-1)	0.0003 (5.69)***	0.0001 (2.94)***	0.0003 (3.73)***	0.0003 (4.80)***
Years at Company (t-1)	-0.006 (1.99)**	-0.010 (4.92)***	-0.009 (2.47)**	-0.007 (2.21)**
Years at Title (t-1)	-0.013 (1.05)	-0.002 (0.36)	-0.002 (0.16)	-0.011 (0.86)
Years at Level (t-1)	0.044 (3.50)***	0.028 (4.27)***	0.031 (2.00)**	0.045 (3.54)***
Expected Performance (t+1)	•	•	•	-0.028 (0.50)
Job Level Dummies (t-1)	YES	YES	YES	YES
Job Title Dummies (t-1)	YES	YES	YES	YES
Year Dummies	YES	YES	YES	YES
Sample Size	N = 6514	N = 11,170	N = 4400	N = 6346
Pseudo R^2	0.14	0.19	0.14	0.14

[1] Includes only job titles C, D, G, K, and Q.

[2] Less stringent sample uses job titles C, D, E, F, G, K, L, N, O, P, and Q.

Notes: Z-statistics in parentheses below each estimate. Statistical significance at the 10%, 5%, and 1% levels denoted by *, **, and ***. All effects are evaluated at the means for all covariates. Marginal effects displayed for continuous covariates. For dummy covariates, cell entries are the differences in predicted probabilities when the dummy equals 1 and when it equals 0.

TABLE 5: OLS Estimates of Change in Annual Log-Wage

	(1)	(2)	(3)	(4)	(5)	(6)
	Basic Sample[1]		**Less Stringent Sample[2]**		**Basic Sample[1]**	
	Promotions	No Prom.	Promotions	No Prom.	Promotions	No Prom.
High School Graduate+	-0.404 (0.91)	-0.807 (1.06)	-0.275 (0.91)	-0.769 (1.20)	-0.339 (0.68)	0.848 (0.87)
MA	-1.210 (3.14)***	1.047 (1.63)	-0.948 (3.10)***	0.775 (1.39)	-0.515 (1.13)	1.147 (1.56)
PhD	-1.310 (1.47)	2.836 (1.31)	-0.891 (1.23)	2.472 (1.32)	-0.864 (0.81)	5.485 (2.52)**
Performance (t-1)	-0.397 (1.63)	-0.347 (3.23)***	-0.367 (2.02)**	-0.297 (3.32)***	-0.267 (0.083)	-0.408 (3.23)***
Performance (t-2)	•	•	•	•	0.594 (1.98)**	0.265 (2.36)**
Age (t-1)	-1.012 (6.10)***	•	-0.916 (7.27)***	•	-0.760 (3.67)***	•
Age squared (t-1)	0.010 (4.92)***	0.001 (0.75)	0.009 (5.87)***	0.001 (0.51)	0.007 (2.89)***	0.0003 (0.19)
Years at Company (t-1)	0.207 (1.71)*	0.008 (0.05)	0.185 (1.98)**	0.036 (0.28)	0.117 (0.90)	0.039 (0.20)
Years at Title (t-1)	-0.315 (0.84)	0.244 (0.67)	-0.131 (0.47)	0.100 (0.50)	-0.258 (0.51)	0.254 (0.60)
Years at Level (t-1)	0.656 (1.69)*	-0.664 (1.85)*	-0.672 (2.30)**	-0.473 (2.35)**	-0.655 (1.27)	-0.637 (1.53)
Constant	47.704 (11.15)***	7.216 (3.54)***	32.211 (9.95)***	9.355 (5.13)***	32.56 (6.46)***	5.661 (2.63)***
Job Level Controls (t-1)	YES	YES	YES	YES	YES	YES
Job Title Controls (t-1)	YES	YES	YES	YES	YES	YES
Year Dummies	YES	YES	YES	YES	YES	YES
Differences in Coefficients						
High School Graduate+	0.403 (0.46)		0.493 (0.70)		-1.187 (1.09)	
MA	-2.257 (3.01)***		-1.723 (2.70)***		-1.66 (1.92)*	
PhD	-4.146 (1.78)*		-3.363 (1.67)*		-6.345 (2.62)***	
Sample Size	N = 1302	N = 7442	N = 2030	N = 9153	N = 829	N = 5295

[1] Uses only job titles C, D, G, K, and Q.

[2] Less stringent sample uses job titles C, D, E, F, G, K, L, N, O, P, and Q.

Notes: All coefficients are multiplied by 100. Statistical significance at the 10%, 5%, and 1% levels denoted by *, ** and ***. Specification also includes interactions of the promotion dummy with all other covariates in addition to the education interactions. Age variable is dropped in the "no promotions" models due to collinearities in the presence of individual fixed effects.

TABLE 6: OLS Estimates of Change in Wage from Year t-1 to t

	(1)	(2)	(3)	(4)	(5)	(6)
	Basic Sample[1]		**Less Stringent Sample[2]**		**Basic Sample[1]**	
	Promotions	No Prom.	Promotions	No Prom.	Promotions	No Prom.
High School Graduate+	-21.969 (0.09)	-63.812 (0.12)	-38.024 (0.23)	-328.854 (0.75)	-28.121 (0.10)	777.222 (1.16)
MA	5.384 (0.03)	377.702 (0.87)	67.476 (0.41)	189.803 (0.50)	126.476 (0.47)	597.698 (1.18)
PhD	-325.929 (0.69)	5855.571 (4.00)***	-206.394 (0.53)	4494.95 (3.52)***	-527.581 (0.85)	7808.706 (5.17)***
Performance (t-1)	-262.359 (2.04)**	-576.698 (7.83)***	-258.899 (2.65)***	-497.883 (8.06)***	-364.490 (1.95)*	-555.056 (6.31)***
Performance (t-2)	•	•	•	•	504.096 (2.88)***	-45.626 (0.58)
Wage(t-1)	-0.021 (1.64)	-0.111 (13.70)***	-0.029 (3.36)***	-0.093 (13.57)***	-0.007 (0.39)	-0.068 (7.58)***
Age (t-1)	-314.811 (3.54)***	•	-262.709 (3.81)***	•	-308.602 (2.52)**	•
Age squared (t-1)	2.960 (2.69)***	-8.647 (8.05)***	2.492 (2.92)***	-7.707 (8.75)***	2.830 (1.91)*	-7.240 (5.88)***
Years at Company (t-1)	39.880 (0.62)	794.310 (6.64)***	1.384 (0.03)	680.298 (7.29)***	8.504 (0.11)	571.371 (4.12)***
Years at Title (t-1)	-93.349 (0.47)	316.202 (1.29)	51.228 (0.34)	75.191 (0.55)	-137.926 (0.47)	273.018 (0.94)
Years at Level (t-1)	-329.055 (1.61)	-290.231 (1.19)	-371.913 (2.37)**	-112.486 (0.82)	-271.442 (0.90)	-287.026 (1.00)
Constant	25554.76 (11.06)***	20057.63 (12.44)***	12748.81 (7.32)***	27631.25 (12.49)***	13176.13 (4.47)***	18459.14 (9.56)***
Job Level Controls (t-1)	YES	YES	YES	YES	YES	YES
Job Title Controls (t-1)	YES	YES	YES	YES	YES	YES
Year Dummies	YES	YES	YES	YES	YES	YES
Differences in Coefficients						
High School Graduate+	41.843 (0.07)		290.829 (0.63)		-805.343 (1.10)	
MA	-372.319 (0.77)		-122.327 (0.29)		-471.223 (0.82)	
PhD	-6181.5 (4.02)***		-4701.34 (3.52)***		-8336.29 (5.11)***	
Sample Size	N = 1302	N = 7442	N = 2030	N = 9153	N = 829	N = 5295

[1] Uses only job titles C, D, G, K, and Q.

[2] Less stringent sample uses job titles C, D, E, F, G, K, L, N, O, P, and Q.

Notes: Statistical significance at the 10%, 5%, and 1% levels denoted by *, ** and ***. Specification also includes interactions of the promotion dummy with all other covariates in addition to the education interactions. Age variable is dropped in the "no promotions" models due to collinearities in the presence of individual fixed effects.

TABLE 7: Probit Marginal Effects for Probability of Promotion in Year t

	(First-promotion subsample)		(Subsequent Promotion Sample)	
	(1)	**(2)**	**(3)**	**(4)**
	Basic Sample[1]	**Less Stringent Sample[2]**	**Basic Sample[1]**	**Less Stringent Sample[2]**
HS Graduate+	-0.095 (5.36)***	-0.057 (3.53)***	-0.025 (1.63)	-0.022 (3.11)***
MA	0.104 (4.60)***	0.097 (4.57)***	0.018 (1.11)	0.013 (1.60)
PhD	0.154 (2.52)**	0.141 (2.58)***	0.150 (3.06)***	0.061 (2.55)**
Performance (t-1)	-0.043 (4.02)***	-0.046 (4.71)***	-0.062 (6.38)***	-0.042 (8.85)***
Performance (t-2)	•	•	•	•
Age (t-1)	-0.043 (6.55)***	-0.021 (3.24)***	-0.034 (4.78)***	-0.017 (5.20)***
Age squared (t-1)	0.0004 (5.03)***	0.0001 (1.47)	0.0003 (3.28)***	0.0001 (3.58)***
Years at Company (t-1)	-0.007 (0.24)	-0.018 (0.54)	-0.015 (3.74)***	-0.009 (4.68)***
Years at Title (t-1)	-0.020 (1.11)	-0.009 (0.62)	-0.008 (0.44)	-0.002 (0.34)
Years at Level (t-1)	0.042 (1.16)	0.039 (1.08)	0.057 (3.17)***	0.029 (4.44)***
Expected Performance (t+1)	•	•	•	•
Job Level Dummies (t-1)	YES	YES	YES	YES
Job Title Dummies (t-1)	YES	YES	YES	YES
Year Dummies	YES	YES	YES	YES
Sample Size	N = 2960	N = 4113	N = 3546	N = 7048
Pseudo R^2	0.16	0.15	0.15	0.21

[1] Includes only job titles C, D, G, K, and Q.
[2] Less stringent sample uses job titles C, D, E, F, G, K, L, N, O, P, and Q.
Notes: Z-statistics in parentheses below each estimate. Statistical significance at the 10%, 5%, and 1% levels denoted by *, **, and ***. All effects are evaluated at the means for all covariates. Marginal effects displayed for continuous covariates. For dummy covariates, cell entries are the differences in predicted probabilities when the dummy equals 1 and when it equals 0.

TABLE 8: OLS Estimates of Change in Wage from Year t-1 to t

	Changes in Log-Wages				Changes in Wage Levels			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	First Promotion Sample		**Subsequent Promotion Sample**		**First Promotion Sample**		**Subsequent Promotion Sample**	
	Promotions	No Prom.	Promotions	No Prom.	Promotions	No Prom.	Promotions	No Prom.
MA	-0.021 (3.50)***	0.138 (2.49)**	-0.004 (0.70)	0.004 (0.49)	-382.535 (1.45)	6744.885 (1.93)*	214.326 (0.64)	7.011 (0.01)
Performance (t-1)	-0.008 (1.95)*	0.006 (1.64)	0.005 (1.16)	-0.005 (2.79)***	-280.261 (1.56)	106.522 (0.48)	189.949 (0.76)	-572.426 (4.01)***
Age (t-1)	-0.009 (2.83)***	0.025 (0.94)	-0.010 (2.95)***	-0.003 (0.89)	-228.332 (1.72)*	1714.062 (1.01)	-373.521 (1.85)*	7.856 (0.03)
Age squared (t-1)	0.0001 (2.07)**	0.0001 (0.16)	0.0001 (2.22)**	0.00002 (0.65)	2.032 (1.17)	-3.113 (0.84)	3.443 (1.35)	-3.631 (1.61)
Years at Company (t-1)	0.011 (0.92)	•	0.003 (1.16)	•	299.709 (0.55)	•	2.262 (0.02)	•
Years at Title (t-1)	-0.004 (0.47)	-0.019 (0.55)	0.008 (1.08)	0.011 (0.28)	-167.628 (0.43)	-1055.895 (0.49)	445.494 (0.95)	1277.049 (0.45)
Years at Level (t-1)	-0.014 (0.87)	-0.010 (0.22)	-0.023 (2.87)**	-0.014 (0.36)	-475.349 (0.69)	-440.066 (0.16)	18459.14 (9.56)***	-1254.108 (0.44)
Constant	0.366 (5.42)***	-0.888 (0.96)	0.294 (4.21)***	0.196 (1.50)	13547.50 (4.54)***	-55372.79 (0.94)	13172.85 (3.08)***	12353.71 (1.25)
Job Level Controls (t-1)	YES	YES	YES	YES	YES	YES	YES	YES
Job Title Controls (t-1)	YES	YES	YES	YES	YES	YES	YES	YES
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES
Differences in Coefficients								
MA	-0.159 (2.86)***		-0.008 (0.79)		-7127.42 (2.03)**		207.314 (0.28)	
Sample Size	N = 477	N = 1353	N = 489	N = 3050	N = 477	N = 1353	N = 489	N = 3050

Notes: All tests reported in Table 8 consider subsamples of our basic sample, using only job titles C, D, G, K, and Q.

Statistical significance at the 10%, 5%, and 1% levels denoted by *, ** and ***. Z-statistics in parentheses below each estimate.

TABLE 9: Correlation Matrix for Performance Ratings over Time

	$Performance_t$	$Performance_{t-1}$	$Performance_{t-2}$	$Performance_{t-3}$
$Performance_t$	1.000			
$Performance_{t-1}$	0.581*	1.000		
$Performance_{t-2}$	0.394*	0.590*	1.000	
$Performance_{t-3}$	0.249*	0.398*	0.610*	1.000

Note: * indicates correlation is statistically significantly different from zero at the 1% level. Correlations computed using "stringent sample" (i.e. job titles C,D,G,K,Q).

TABLE 10: Probit Marginal Effects for Probability of Promotion in Year t Controlling for Various Lags of Performance

	(1)	(2)	(3)	(4)	(5)
	Performance in Year t-1	**Performance in Years t-1 and t-2**	**Performance in Years t-1, t-2, and t-3**	**Performance in Years t-1,t-2,t-3, and t-4**	**Performance in Years t-1,t-2,t-3, and t-4**
HS Graduate+	-0.059 (5.09)***	-0.053 (3.89)***	-0.046 (2.82)***	-0.041 (2.04)**	-0.027 (1.19)
MA	0.052 (3.96)***	0.031 (1.98)**	0.055 (2.85)***	0.051 (2.11)**	0.061 (2.37)**
PhD	0.157 (4.05)***	0.154 (3.17)***	0.129 (2.08)**	0.191 (2.11)**	0.194 (2.00)**
Performance (t-1)	-0.052 (7.30)***	-0.040 (4.11)***	-0.034 (2.78)***	-0.021 (1.29)	-0.051 (1.99)**
Performance (t-2)	•	-0.035 (3.81)***	-0.040 (3.29)***	-0.029 (1.82)*	-0.033 (2.01)**
Performance (t-3)	•	•	-0.018 (1.75)*	-0.025 (1.71)*	-0.028 (1.91)*
Performance (t-4)	•	•	•	-0.007 (0.57)	-0.006 (0.48)
Expected Performance (t+1)	•	•	•	•	0.126 (1.55)
Sample Size	N = 6514	N = 4400	N = 2932	N = 1819	N = 1796
Pseudo R^2	0.14	0.14	0.15	0.13	0.13

Notes: Each specification is estimated on the subsample of job titles C, D, G, K, and Q. Apart from the the number of lagged performance measures, the specification is identical to our main specification in Column 1 of Table 3. Z-statistics in parentheses below each estimate. Statistical significance at the 10%, 5%, and 1% levels denoted by *, **, and ***. All effects are evaluated at the means for all covariates. Marginal effects displayed for continuous covariates. For dummy covariates, cell entries are the differences in predicted probabilities when the dummy equals 1 and when it equals 0.