Postschooling Training Investment and Employer Learning

Barış Kaymak
Université de Montréal

Wage growth among young workers relates positively to ability and negatively to education, conditional on ability. This was interpreted as evidence for statistical discrimination with employer learning. I show that this pattern is also consistent with a version of the Ben-Porath model of skill formation in which (i) workers differ in their learning ability and (ii) job training is a substitute for formal schooling. Data on job training from the NLSY confirm both modeling extensions for young workers. Nonetheless, for more experienced workers, job training and formal schooling appear to be positively related, even after controlling for ability.

Workers with higher cognitive ability, as measured by aptitude tests, enjoy faster wage growth throughout their careers. However, among workers with similar abilities, those with higher educational attainment have slower wage growth. This pattern is particularly strong for young workers and lasts for about 10–15 years.

These empirical regularities are considered as evidence for statistical discrimination based on educational attainment. When a prospective worker's productivity is not directly observed, employers rely on indicators of quality, such as education, to set a wage rate. Eventually, they learn to distinguish between workers with similar educational backgrounds on the basis of their job performance, which brings the wage rate closer to productivity.[1] The statistical discrimination model thus predicts that correlates of productivity that are unobservable to employers do not explain wages of young workers but become important determinants with experience. Furthermore, if these correlates are related to education, then educational attainment is expected to be less and less important for experienced workers (Farber and Gibbons 1996; Altonji and Pierret 2001).

In this paper, I develop and test a human capital model that is also consistent with the empirical regularities above. It is natural for workers with better learning skills to have a comparative advantage in acquiring new

[1] This implicitly assumes a competitive labor market in which potential employers have the same information as the current employer. Several papers in the learning literature have explored deviations from this assumption (see, e.g., Gibbons and Katz 1991; Schönberg 2007).

[*Journal of Human Capital,* 2014, vol. 8, no. 3]

skills, not only through formal education but also through job training and formation. Consequently, these workers have higher educational attainment, on average, and they enjoy faster earnings growth throughout their career. It is conceivable, however, that, conditional on ability, workers with less schooling invest more in their skills to compensate for their lack of formal education and thereby enjoy faster earnings growth. In this case, earnings of workers with similar ability but different schooling levels converge to each other, and education becomes a poor determinant of earnings, especially for experienced workers.

I formalize this argument using a variation of the Ben-Porath (1967) model of life cycle skill formation and derive conditions under which the model delivers the observed empirical patterns of wage growth by ability and education. These conditions are critical for delivering salient features of the wage data, such as the concavity of the wage experience profile. They are, therefore, met in reasonable formulations of the Ben-Porath model.

The key features that allow the model to match the observed wage patterns are the substitutability of job training for education and the complementarity between training time and ability in acquisition of new skills. Several testable implications emerge as a result. The model implies that workers with higher ability invest more time in job training and that training activity is decreasing in education for workers of comparable ability. Furthermore, the return to training investment is similar to the return to education, especially for younger workers. Later on, potential complementarity between training time and human capital may lead to a larger marginal return to training.

I test these implications using data on training programs from the 1979 cohort of the National Longitudinal Survey of Youth (NLSY). Results from the Armed Forces Qualification Test (AFQT) are used as a proxy for ability following the literature. The data confirm a strong positive link between ability and training investment. Conditional on ability, workers with higher educational attainment spend less time on training activities in total. Moreover, the (annualized) return to training is estimated to be 7.8 percent for young workers, which is similar to the return to a year of education at 8.3 percent. These findings lend support to the human capital model.

Further investigation of training activity by experience nonetheless reveals some conflicting evidence. It appears that training and education are substitutes around the time of labor market entry, but a positive correlation between education and training emerges with experience. Although workers with higher educational attainment invest less in training early on, total training investments are similar after 15 years of experience for workers with similar abilities. By contrast, the gap in training investment by ability widens with experience as predicted by the human capital model. This suggests that the negative association between education and wage growth for workers with similar abilities remains a qualitative diagnostic for the statistical discrimination model. Moreover, although the link between training and ability is statistically strong, observed investment in training is not quantitatively sufficient to explain the patterns in wage growth given the estimated return to job training. These findings suggest that both models are at play in the determination of wage growth around labor market entry, whereas the employer learning model is likely the more dominant model after a few years of experience.

The implications of the statistical discrimination model for wage growth were also used to test whether employers discriminate against workers on the basis of race (Altonji and Pierret 2001). Yet, it is possible that employers are fully informed, and race is simply correlated with investment in human capital and the ability to acquire new skills. The data on training activity in the NLSY, however, do not show any systematic difference by race. This undermines the human capital story for understanding the racial differences in earnings growth and reinforces the statistical discrimination theory as a more viable interpretation of racial differences in wage growth for young workers.

This paper is closest, in spirit, to the paper by Kahn and Lange (2010), who test the employer learning model using direct productivity measures from a single, large firm (the Baker-Gibbs-Holmstrom data). They find that the variation in individual productivity is the predominant factor in pay dispersion, suggesting a reinterpretation of the employer learning literature. The findings here confirm this conclusion using a more specific model of productivity changes, one that depends on job training, but in a more general sample.

Section I presents the model and outlines the key conditions that allow the model to generate the observed differences in wage growth by ability and education. Section II describes the data, and Section III reviews the wage regressions that provide the base for the recent employer learning literature. Section IV evaluates the implications of the human capital model for training and wages. Section V presents conclusions.

I. A Model of Skill Formation with Heterogeneous Agents

The objective of this section is to develop a human capital model in which wage growth depends positively on ability and negatively on educational attainment as in a statistical discrimination model.[2] This requires ability to be complementary to investment in human capital and formal education to be a substitute. To achieve the former, ability is interpreted as the efficiency with which a worker acquires new skills, also known as the ability to learn after Heckman, Lochner, and Taber (1998a). This interpretation

[2] The human capital model developed here is treated as a distinct alternative to the employer learning model, thereby abstracting from potential interactions between the two models. Gibbons and Waldman (1999, 2006) study in this vein the task assignment problem under uncertainty and link changes in marginal product to the information acquired by the employer via promotions.

allows workers with higher abilities to learn more skills for a given amount of time and facilitates acquisition of skills both during school and later at work. A positive relation between ability, education, and wage growth emerges as a result.

It is relatively less straightforward to generate a negative relationship between education and wage growth with a human capital model. In a typical learning-by-doing model, for instance, education is positively associated with future wage growth because higher education raises a worker's market value and encourages labor supply, which is the source of human capital accumulation in that model. In an on-the-job training model à la Ben-Porath, education and training intrinsically are substitutes. However, the particular heterogeneity structure one introduces in the model may generate a seeming complementarity. For instance, Willis and Rosen (1979) study a setting in which workers differ in their cost of education as captured by the discount rate. In a Ben-Porath model, this type of variation predicts a positive relationship between education and wage growth since workers with higher discount rates not only spend less time at school but also invest less in training. I show below that an alternative formulation of the Ben-Porath model, one in which differences in educational attainment, conditional on ability, arise from variation in preferences for formal schooling, can generate the observed patterns of wage growth in the data. As long as there are nonincreasing returns to human capital, those who prefer to quit school earlier graduate with a smaller stock of human capital and invest more heavily in training activities later on.

Formally, suppose that the economy is populated by a set of workers with finite careers. The function $K(t)$ denotes the human capital stock of a worker at age t. In the terminology of Heckman et al. (1998a), $K(t)$ is referred to as the *ability to earn*. The market for human capital is competitive. A unit of human capital is valued at rate R. Each worker is endowed with a unit of work effort at any time. Workers can accumulate human capital by investing a fraction l of their effort to training. A worker's current earnings are

$$y(t) = RK(t)[1 - l(t)]. \tag{1}$$

The human capital stock of a worker evolves according to the following production function:

$$\dot{K}(t) = z_i g(l(t), K(t)), \tag{2}$$

where z_i is the individual-specific skill parameter that captures differences in ability to accumulate additional skills.[3] The production function is in-

[3] The human capital formation depends on individual investment choices only and abstracts from potential spillover effects as analyzed in Ehrlich and Kim (2007).

creasing in both arguments and displays diminishing returns in each argument. It is assumed that the depreciation rate of human capital is zero.[4]

To capture the differences in preferences for schooling, let the flow variable ψ_i denote the monetary equivalent of the net utility gain associated with schooling, which may originate from differences in social background, parents' attitudes toward education, or, simply, personal tastes for school as opposed to work. The terms ψ_i and z_i are independent.[5] If the utility function is linear in earnings, ψ_i can also be interpreted as an idiosyncratic cost of schooling. The technology for human capital production is the same at school and at work. If the worker chooses to attend school, however, he has to devote all his time to human capital accumulation: $l(t) = 1$. A worker's problem is to choose the optimal investment level to maximize his lifetime earnings. The individual control problem can be formalized as follows:

$$\max_{l(t)\in[0,1]} \int_0^{T+s} e^{-rt}\{RK(t)[1 - l(t)] + \psi_i I_{l(t)\geq 1}\}\,dt$$

subject to

$$\dot{K}(t) = z_i g(l(t),\ K(t)).$$

The term $I_{l(t)\geq 1}$ is an indicator function that equals one if the worker is engaged in full-time human capital production; T denotes the length of a worker's career, which is assumed to be independent of his education level.[6] The Hamiltonian for the control problem above is

$$\mathcal{H} = e^{-rt}\{RK(t)[1 - l(t)] + \psi l(t) I_{l(t)\geq 1}\} + \lambda(t) z_i g(l(t),\ K(t)),$$

where $\lambda(t)$ is the shadow value of investment in human capital, and $K(0)$ is given and is common to all workers. The interior solution is fully characterized by the following equations:

$$-\frac{\partial \mathcal{H}}{\partial K} = -e^{-rt}R[1 - l(t)] - \lambda(t) z_i g_K(l(t),\ K(t)) = \dot{\lambda}(t), \tag{3}$$

$$\frac{\partial \mathcal{H}}{\partial l} = -e^{-rt}K(t)R + \lambda(t) z_i g_l(l(t),\ K(t)) = 0, \tag{4}$$

$$\lambda(T + s)K(T + s) = 0. \tag{5}$$

[4] This assumption does not alter our theoretical results and is consistent with the estimates reported in Browning, Hansen, and Heckman (1999).

[5] This assumption is innocuous. The results presented here would hold as long as the correlation between the two terms is less than perfect.

[6] The results are robust to a more general formulation of career length, $T(s)$, if $T(s) \leq T + s$. This includes the case in which all workers retire at the same age.

Equation (3) implies that the shadow value of investment is decreasing in time as the worker gets closer to the end of his career. This generates a decreasing investment profile and an increasing wage profile over a worker's career.

Equation (4) is the interior optimality condition for the time invested in training. As $\lambda(t)$ decreases, the marginal return to training investment decreases. Meanwhile, capital stock accumulates, raising the opportunity cost of training time. If the complementarity between capital stock and time investment, g_{lK}, is not too large, then the model generates a monotonically decreasing investment profile and an increasing earnings profile over the worker's life.

Workers with higher ability have higher marginal return to investment conditional on the stock of human capital, and they spend more time on training activities. This generates lower earnings at the beginning of their career and yields faster wage growth. If these workers also attain higher education levels, as discussed next, then education and wage growth are positively related.

A. Schooling Period

Workers who attend school spend their entire time endowment on human capital accumulation at school. To analyze the schooling choice, consider the first-order condition when $l(t)$ approaches one from the right, $\partial\mathcal{H}^{+}/\partial l$. An agent chooses to specialize in schooling if the marginal return to investment is strictly greater than the marginal cost at $l(t) = 1$:

$$\frac{\partial\mathcal{H}^{+}}{\partial l} = e^{-rt}[-K(t)R + \psi_i] + \lambda(t)z_i g_l(1,\ K(t)) > 0. \tag{6}$$

This occurs if the initial level of human capital is low and the marginal return to training is high. As the worker invests in his capital, $K(t)$ increases, raising the time cost of investment. At the same time the marginal return to additional human capital declines as the worker's horizon shortens, given, once again, that g_{lK} is not too high. Therefore, the specialization period happens only at the beginning of a worker's life. The worker leaves school when the inequality above is reversed.

Equation (6) yields a testable implication of the model. At the time a worker leaves school, equation (6) holds with equality; that is, the marginal return to a year of investment in training equals the marginal return to the last year of schooling. This also holds, on average, conditional on ability, since z_i and ψ_i are independent. This is a direct consequence of the substitutability of training for schooling. Furthermore, if existing capital and time are complements in the production of new human capital, $g_{lK} > 0$, then the marginal return to training increases as workers build human capital. A second implication, therefore, is that the return to training increases with experience. I test both these implications below.

Equation (6) also shows that workers with higher-utility attachments to school, ψ_i, or with higher ability, z_i, are more likely to stay longer at school. These are the driving forces of the heterogeneity in educational attainment in the model. When ability is not directly observable to the econometrician, then higher wage growth caused by higher ability can be mistakenly attributed to education. This is intuitively similar to the classical concept of ability bias in the estimation of the return to education (Griliches 1977). In fact, conditional on ability, a higher education level results in a higher stock of human capital at graduation, which reduces the training investment over the life cycle as a result of the decreasing return to capital stock in the production of human capital. The next section formalizes this result.

B. *Ability, Education, and Wage Growth*

Since ability raises the speed of human capital accumulation in this model, those with higher ability experience faster wage growth, on average. To see this, consider the average wage growth over a worker's career:

$$\begin{aligned}\frac{w(T+s)-w(s)}{T} &= \frac{RK(T+s)-RK(s)[1-l(s)]}{T}\\ &= \frac{zR\int_0^{T+s} g(l(t),\ K(t))\,dt + Rl(s)K(s)}{T},\end{aligned}$$

where the terminal condition $l(T+s)=0$ is imposed. Workers with higher ability not only produce more human capital for a given level of investment, $l(t)$, but also invest more intensively in training according to equation (4). This implies that the term above is unambiguously increasing in z.

Perhaps less evident is the negative effect of educational attainment on the rate of wage growth conditional on ability. When the observed earnings for a worker with s years of education and x years of experience is denoted by $w(x, s)$, the growth rate of earnings at experience level x is

$$\begin{aligned}\frac{w_x(x,\ s)}{w(x,\ s)} &= \frac{K_x(x,\ s)[1-l(x,\ s)] - l_x(x,\ s)K(x,\ s)}{K(x,\ s)[1-l(x,\ s)]}\\ &= \frac{zg(l(x,\ s),\ K(x,\ s))}{K(x,\ s)} - \frac{l_x(x,\ s)}{1-l(x,\ s)}.\end{aligned} \tag{7}$$

The dependence on z is suppressed for simplicity. Equation (7) shows that wage growth depends on two components. The first component is how much additional earning capacity is generated by training, and the second component is how much more of that capacity is released from learning activities and reallocated to earning activities. The effect of education on these components generally depends on the properties of the human capital production function $g(l, K)$. In what follows I formally analyze two classes of functions broadly used in the literature.

1. Case I: $g(l,\ K) = f(l)K^{\alpha}$

This case assumes that the effect of time and human capital stock is multiplicatively separable and includes the commonly used Cobb-Douglas form (e.g., Heckman, Lochner, and Taber 1998a, 1998b). Consider first the extreme case in which $\alpha = 1$; that is, production of new capital is a linear function of the level of current human capital. In this case, the capital stock cancels in the optimality conditions described in (3)–(5), which implies that paths for $\lambda(t)$ and $l(t)$ do not depend on the capital stock at the time of graduation.

The wage growth defined in equation (7) can be rewritten as

$$\frac{w_x(x,\ s)}{w(x,\ s)} = zf(l(x)) + \frac{l_x(x)}{1 - l(x)}.$$

Since the time investment in training is independent of education, the only variation in wage growth comes from z. Conditional on z, log wage and log capital stock profiles over experience are parallel for different educational choices. This critical case displays a *relative neutrality*, in the sense that the marginal return to and the marginal cost of training time relative to potential earnings are independent of the human capital stock.

When the marginal return to training time increases less than proportionally to potential earnings, the workers with higher potential earnings upon graduation find it optimal to spend less time on training activities. The following proposition establishes this result (the formal proof is relegated to App. A).

Proposition 1. If $g(l(t),\ K(t)) = f(l)K^{\alpha}$, then the rate of wage growth over experience is decreasing in education, conditional on z, if $\alpha \leq 1$, with equality if and only if $\alpha = 1$.

There are three distinct channels with which $\alpha < 1$ obtains the result in the proposition. First, with $\alpha < 1$, the marginal productivity of human capital is strictly decreasing. As a result, the rate of growth of potential earnings is decreasing in the level of capital stock for a given rate of investment. Second, the complementarity between time and capital stock, g_{lK}, weakens. As a result, workers with higher capital stock invest less time in human capital accumulation, which results in reduced growth of potential earnings for workers with the same ability level but higher educational attainment. Third, since actual earnings equal potential earnings net of investment expenditures, the rate of decline in investment expenditures relative to earnings also matters for the differences in the growth rates of earnings. Workers with higher capital stock not only spend less time training but also choose to decrease their training time more slowly with experience. As less effort is released from the training activities and allocated to work, earnings grow more slowly. Thus, the differences in the growth rates of earnings are more pronounced than those of potential earnings.

It was assumed that the length of a worker's career is fixed regardless of the time spent at school. If a worker's career is limited by age, for example, because of mandatory retirement, the workers with more schooling bear an additional cost to investment in education. After the schooling period is over, they face a shorter work horizon and invest even less in training, which amplifies the result above.

2. Case II: $g(l,\ K) = h(lK)$

Another widely used functional restriction on the technology of human capital production assumes that current human capital raises the return to market time and training time equally. Formally, this is achieved when the training time and the current capital stock enter the production function multiplicatively:

$$g(l(t),\ K(t)) = h(l(t)K(t)) = h(Q(t)),$$

where $h' > 0$, $h'' < 0$, and $Q(t)$ denotes the total outlays on training. Note that this is neither a special nor a more general case of the functional form in case I.

In this case, an additional unit of human capital acquired today simply raises the human capital stock available for rent throughout a worker's career but does not otherwise affect future investment decisions.[7] Hence, the investment decisions of better-educated workers with larger capital stocks are the same as those with inferior education levels. As a result the life cycle profiles of the level of potential earnings for workers with different educational attainments are parallel. We refer to this case as *absolute neutrality*.

The rate of growth of earnings can be written as follows:

$$\frac{w_x(x,\ s)}{w(x,\ s)} = \frac{zh(Q(x)) - Q_x(x)}{K(x,\ s) - Q(x)}. \tag{8}$$

The numerator of this expression is identical for all workers with the same ability, at any experience level, since they have the same training expenditure. The only difference is the stock of human capital, which is increasing in education. Since workers with higher education start their careers with higher levels of human capital, their earnings are always equally higher. This immediately implies that the changes in earnings over experience are the same for all workers, but the growth rate is lower for workers with more education. This is merely a scale effect.

Figure 1 depicts the training intensity for two workers with the same learning ability but different schooling levels. The optimal path for the worker with higher education lies below that for the other worker because he has less time to reap the benefits of his investment and he has a higher

[7] To see this, note that the shadow value of a unit of human capital at the margin is simply the present discounted value of rental rates for human capital until retirement: $\lambda(t) = Re^{-rt}[1 - e^{-r(T-t)}]/r$. The optimality condition for training then becomes $z_i h'(Q(t))\{1 - e^{-r[T(s)-t]}\} = r$, which depends only on time (or rather on the work horizon).

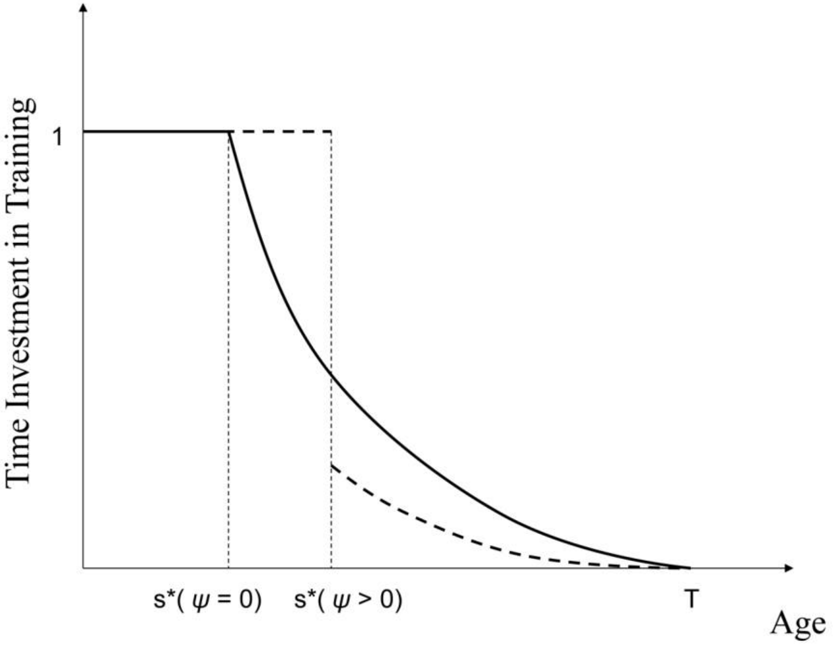

Figure 1.—Training intensity with absolute neutrality. The figure shows the fraction of time invested in training as a function of age for two workers with the same learning ability, z, but different utility gains from formal schooling, ψ. The dashed line corresponds to the worker with a higher ψ, who remains longer at school.

stock of human capital. The total investment expenditure is the same for both workers as it is determined only by age. This generates parallel potential earnings profiles as a function of age (fig. 2*A*). However, the worker with higher education has a shorter career; therefore, he faces a lower marginal rate of return to investment conditional on experience. Figure 2*B* shows potential and observed earnings as a function of experience. The shifted profiles display a lower level of investment expenditures, $Q_s(s, x) < 0$, and a lower growth in potential earnings at any experience level, $K_{xs} < 0$.

In order to link the above argument for potential earnings to observed earnings, we also need to consider the rate of decline in the investment expenditures. If the training expenditure is convex in age, indicating that the investment decreases faster for younger workers, then the observed earnings growth is also lower for the educated, conditional on ability.[8] The convexity of the training expenditures in turn depends on a third-order condition on the production function. The following proposition formalizes the result.

[8] Note that the denominator in eq. (7) is still increasing in s. If, at any time, the worker with less education were to catch up, then the two workers would have the same wage profile from then on.

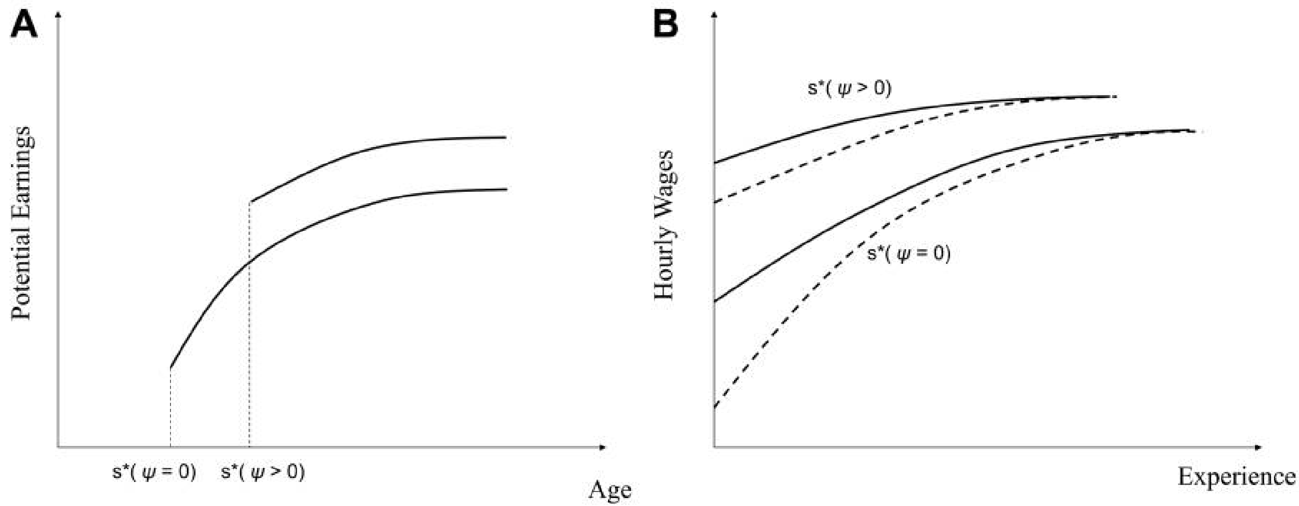

Figure 2.—Human capital, earnings, and experience under absolute neutrality. The figure shows potential (solid) and actual (dashed) earnings as a function of age and experience for two workers with the same learning ability, z, but different utility gains from formal schooling, ψ.

Proposition 2. Suppose that $g(l(t), K(t)) = h(l(t)K(t))$. Then the rate of wage growth over experience is decreasing in education, conditional on z, if

$$\frac{h''' h'}{(h'')^2} = k \geq 1,$$

where k is the hyperbolic absolute risk aversion in the context of utility functions.

The exponential function, $h(x) = -e^{\theta x}/\theta$, satisfies this condition with equality. The Cobb-Douglas function, which is a widely used parameterization of the human capital production function in the literature, also satisfies this condition.[9] Nevertheless, this condition is not necessary. In fact, when this condition is satisfied with equality, the proposition still holds because of the scale effect discussed above. The condition may seem arbitrary in the context of production functions, but it is critical for generating earnings profiles that are increasing and concave in experience. The following proposition states the equivalence of the two empirical regularities.

Proposition 3. The rate of wage growth over experience is decreasing in education, conditional on z, if wage profiles are log concave in experience.

This is a direct result of the absolute neutrality assumption. Recall from the numerator of equation (8) that the change in wage level at time (or age) t is summarized as a function of the investment level, $Q(t)$. Since $Q(t) = Q(x + s)$, the rates of decline in optimal investment at the margin with respect to age, experience, and education are equivalent: $\dot{Q}(t) = Q_x(s + x) = Q_s(s + x)$; hence the parallel wage levels. The only other variable that determines wage growth besides investment is the capital stock, which appears in the denominator in (8). But the capital stock grows faster at school than at work since, by definition, workers spend less time on learning during training than they do at school ($l(t) < 1$ if $t > s$). Therefore, the change in wage growth with respect to experience is higher than the change with respect to education. If the former is negative because of log concavity of wages, the latter has to be negative (see App. A for the formal derivation).

While the log concavity of the wage-earnings profile is a generally accepted fact, workers may experience linear or even convex profiles in some cases for the first few years on a job (Heckman, Lochner, and Todd 2008). The definition of log concavity used in proposition 3 is a weak one and therefore allows for linear profiles. When wages rise linearly with experience, the rate of wage growth is still decreasing in education since more educated workers have higher wage levels. The proposition excludes strictly convex log wage profiles.

[9] The exponential function displays a constant absolute risk aversion of one in utility theory. Constant relative risk aversion has $k > 1$, and quadratic has $k = 0$.

C. Wage Dispersion over Experience

Besides the variation in wage growth by education and ability, there are two other predictions often attributed to the statistical discrimination model with employer learning. First, as employers learn workers' true productivity, the variance of earnings, conditional on education, grows over time. Workers with the same education level start out with similar wages, but as employers learn to distinguish among these workers, wages become more disperse.

On the other hand, the variance of earnings, conditional on ability, diminishes over time. In fact, if education is a pure indicator, without any value added to productivity, the variance converges to zero. Both of these predictions are also consistent with the human capital model presented above.

Take two workers with the same level of education but with different ability levels. We have established above that the worker with higher ability has a higher wage growth over experience. There are two reasons. First, the worker with higher ability is better at accumulating human capital, leading to faster wage growth for similar investments. Second, since the worker with lower ability is relatively overinvested in human capital through education, he chooses to invest less in training, further slowing down his wage growth. This widens the wage gap for people with similar education, leading to higher variance over time conditional on education.[10]

If we compare workers with the same ability level but different educational attainments, the wage gap will be largest at the beginning. As the worker with lower education invests more in training, he compensates for his lack of education and the wage gap narrows (see fig. 2*B*). The variance conditional on ability, therefore, diminishes with experience.

The typical employer learning model also predicts that wages follow a random walk because the flow of new information about a worker's quality is random and orthogonal to his earnings (Farber and Gibbons 1996). By contrast, wages in an on-the-job training model typically do not follow a martingale because innovations to wages result from investment in human capital, which is nonrandom by construction. The current model is not an exception. Whether wages empirically follow a random walk is a long-running debate (see Meghir and Pistaferri [2011] for a survey), one to which this paper offers little to contribute.

In what follows, the main ingredients of the model that help replicate the predictions of the statistical discrimination model are empirically tested. The two key components of the model are that training activity is associated positively with ability and negatively with educational attainment conditional on ability. This dependence is particularly strong during the early career and weakens with experience. Moreover, because educa-

[10] If the high-ability worker invests sufficiently more in training, he may start out at a lower wage. In this case the variance may decrease at the beginning for a brief period. Heckman et al. (2008) provide some evidence for this prediction.

tion and training are perfect substitutes, the return to education and the return to training investment should be similar at the margin for workers with similar learning ability, especially at the beginning of a worker's career.

II. Data

The data are taken from the 1979 cohort of the NLSY. The NLSY79 is a panel study of men and women who were aged 14–21 in 1978. The subjects were surveyed annually since 1979 (biannually after 1994) on several economic variables including their education, family background, and labor market outcomes. The survey contains detailed information on workers in transition from schooling to the labor market when most training activity and employer learning occur. The NLSY79 also reports the training activities of workers on and off the job throughout the survey, and it contains the scores of the AFQT, a variable that is widely used as a proxy for cognitive ability in the literature. Details of the data are reported in Appendix B.

The main drawback of the data is that the observed incidence of training is rather low since the respondents report only formal training spells. The fraction of respondents who report participation in training programs between two consecutive interviews is around 15 percent throughout the sample. Several activities on the job could qualify as training, though not necessarily be reported as such.[11] Therefore, the training observed in the data is likely a lower bound for the actual investment in human capital. Barron, Berger, and Black (1997) estimate the share of formal training in total training activity to be around one-seventh. The lack of measures is the main drawback of the NLSY79. This is likely to affect the results presented here if the component of informal training that is orthogonal to formal training relates to education and ability in a different way. This caveat should be kept in mind when interpreting the findings below.

The training data in the NLSY79 are incomplete prior to 1988. First, the 1987 survey of the NLSY79 omitted the training section. However, this does not limit the analysis since the missing information was filled in retroactively during the 1988 survey, which led to a jump in the incidence of training reported in 1988. A similar change is observed after 1994, when the survey moved to biannual frequency. Second, whereas the respondents were allowed to report up to three training spells before 1988, after 1988, this limit was raised to four. Nevertheless, the respondents were asked if they had a fourth (fifth beginning in 1988) training program to report, although additional information was not collected if the response was af-

[11] For instance, if a worker spends a few hours reading the manual of a new machine that he needs to operate or spends an hour listening to a coworker's advice on how to perform his task better, he would be considered investing in his human capital.

firmative. On the basis of this question, it is possible to calculate the number of workers constrained by these limits. The limit was binding for a total of 80 observations (about 0.2 percent of the sample) in all years and, hence, is not likely to be consequential for the results presented here.

Another limitation of the training data prior to 1988 is that the type or the intensity of the training activity was not recorded for training spells that lasted less than a month. This prohibits an accurate measure of total training hours for short spells, which, unfortunately, constitute about a third of all training spells. The literature has often ignored the data prior to 1988 for this reason. However, this approach misses out on a significant portion of the training spells right around the time of labor market entry, which are crucial for the purpose at hand. The approach here is to make use of the earlier data as much as possible. One option is to use the detailed data in later surveys to impute the missing information in earlier years. Alternatively, one can drop all the short spells with missing information from the analysis. The estimation results are robust independent of which approach is adopted. The main text reports results from the latter option, where the short spells prior to 1988 were not counted in total training investment. Since these spells are short, the implications for the estimates are not remarkable. The findings using the imputed data are reported in Appendix C.

To test the implications of the model for training, first, an annual flow variable for training investment is calculated going back exactly 1 year from the survey date. Figure 3 shows the annual hours of total training and on-the-job training by experience. Most training occurs at the onset of a worker's career, as predicted by the model. Concentrating on the calendar year prior to the survey results in the loss of some training spells that occur between interviews that are more than a year apart. Consequently, the incidence of training is even lower when a flow measure is used. Only 6.9 percent of the workers report some investment in training during the year prior to the survey. This figure is 2.6 percent if we restrict attention to company training only. This is the main reason behind the low hours of training in figure 3. Among those who report some training investment, the average number of hours of investment is 321, equivalent to 8 weeks of full-time engagement. This figure is 425 hours for those within the first year of experience and declines to 280 hours at 10 years of experience and to 181 at 20 years of experience.

The low incidence of training observed in the data results in a heavy abundance of zeros, making it harder to draw robust inference. To get around this problem, a cumulative variable is created in total hours of training investment. This reduces the incidence of zeros by cumulating training spells over a worker's career and by drawing in additional training spells between interviews that are further apart. For the sample used in the analysis the incidence of zeros for total investment is 52 percent. This percentage declines from 76 percent among workers with less than 1 year of experience to 35 percent among those with 15 years of experience.

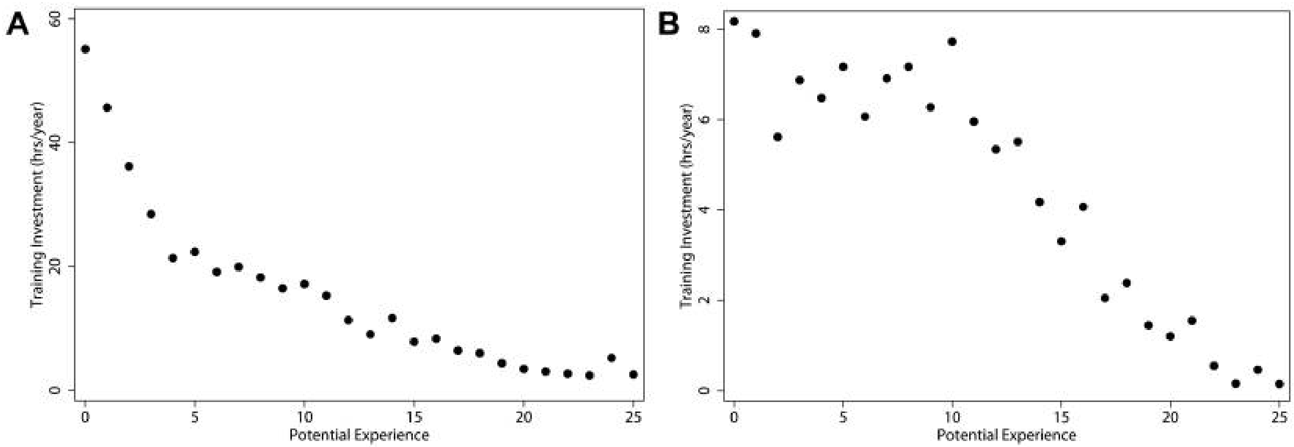

Figure 3.—Job training by experience, NLSY 1979 men: *A*, annual training hours; *B*, annual on-the-job training

III. Ability, Education, and Wage Growth

Table 1 shows the wage regressions that constitute the motivation for the employer learning literature. Column 1 shows the results of a regression of log wages on education, AFQT score, and an interaction of education with experience. The control variables are indicators for years of experience, survey years, and racial background. As expected, wages vary positively with both ability and education. Return to a year of education is 8.0 percent, and the market value of a one standard deviation higher AFQT score is 9.3 percent.

The interaction term between education and experience is statistically zero. Column 2 adds the interaction of AFQT with experience. Results show that wages grow 6.8 percent faster for every 10 years for a one standard deviation higher AFQT score above the mean. More interestingly, the coefficient on education interacted with experience is now significantly lower. A year of education translates into 1.6 percent lower wage growth per decade. These findings are consistent with those of Altonji and Pierret (2001) and Lange (2007).

Altonji and Pierret (2001) use differences in wage growth by race to test whether employers engage in statistical discrimination based on race. Column 3 includes an interaction term between experience and an indicator for race. When ability is excluded from the regression, black workers seem to have similar starting salaries, conditional on education, but they have 10.8 percent lower wage growth per decade. Once ability is controlled for, however, the starting salaries are about 7.1 percent lower for blacks, and they have 5.1 percent lower wage growth every 10 years.

Interpreted within the confines of a statistical discrimination model, the findings above suggest that employers initially use educational attainment

TABLE 1
Education, Ability, and Wage Growth

	Dependent Variable: Log Hourly Wages			
	(1)	(2)	(3)	(4)
Education	.080	.094	.080	.093
	(.006)	(.006)	(.006)	(.006)
AFQT	.093	.041	.093	.046
	(.009)	(.011)	(.009)	(.011)
Black	−.112	−.112	−.025	−.071
	(.009)	(.022)	(.029)	(.030)
Education × exp/10	.002	−.016	.001	−.015
	(.006)	(.006)	(.006)	(.007)
AFQT × exp/10		.068		.062
		(.011)		(.013)
Black × exp/10			−.108	−.051
			(.027)	(.029)
Observations	26,792	26,792	26,792	26,792

Note.—All specifications control for indicators of potential experience, survey year, and race. Standard errors are clustered by respondent and are reported in parentheses. The sample includes NLSY 1979 men with 15 years of experience or less.

to distinguish between workers of different ability. As workers' true abilities are revealed by their performance on the job, education becomes less important relative to the AFQT score. The findings on the growth patterns by race indicate that employers also use race, or another variable that is correlated with race, to distinguish workers' abilities. Since the average racial wage gap is about 11 percent, this constitutes about two-thirds of the racial wage gap. The remaining component of the wage gap emerges as wages further reflect ability over experience.

Lange (2007) shows that the effect of employer learning on differences in wage growth by education and ability is particularly strong early on and dissipates as employers learn workers' true abilities. The human capital model studied above also predicts that the impact of training on wage growth is strongest at the beginning: as workers approach retirement, the value of new human capital decreases and workers invest less in training. This can be seen in table 2, which estimates the differences in wage growth

TABLE 2
Education, Ability, and Wage Growth

	Dependent Variable: Log Hourly Wages			
	(1)	(2)	(3)	(4)
Education	.080	.102	.081	.100
	(.007)	(.007)	(.007)	(.007)
AFQT	.093	.009	.093	.018
	(.009)	(.013)	(.009)	(.014)
Black	−.114	−.111	.033	−.043
	(.022)	(.022)	(.036)	(.039)
Education × exp/5:				
First 5 years	.000	−.022	−.001	−.020
	(.006)	(.008)	(.006)	(.008)
6–10 years	.001	−.014	.000	−.013
	(.004)	(.005)	(.004)	(.005)
11–15 years	.001	−.010	.000	−.010
	(.003)	(.004)	(.003)	(.004)
AFQT × exp/5:				
First 5 years		.087		.079
		(.016)		(.018)
6–10 years		.058		.053
		(.009)		(.010)
11–15 years		.043		.038
		(.006)		(.007)
Black × exp/5:				
First 5 years			−.147	−.068
			(.044)	(.049)
6–10 years			−.097	−.046
			(.022)	(.025)
11–15 years			−.072	−.034
			(.015)	(.017)
Observations	26,792	26,792	26,792	26,792

Note.—All specifications control for indicators of potential experience and survey year. Standard errors are clustered by respondent and are reported in parentheses. The sample includes NLSY 1979 men with 15 years of experience or less.

separately for 5-year periods. The decline in the coefficient of education is 2.2 percent for the first 5 years of experience, 1.4 percent for the next 5, and 1.0 percent between 10 and 15 years of experience. Similarly, the rise in the coefficient of AFQT is 8.7 percent for the first 5 years, 5.8 percent for the next 5, and 4.3 percent between 10 and 15 years of experience. A similar pattern is observed when the interaction of an indicator for black and potential experience is included in the regressions in columns 3 and 4.

IV. Testing the Human Capital Model with Job Training

The capacity of the human capital model to explain the patterns of wage growth discussed above depends on the value of job training in the market and on the differences in training investment by ability and education. Thus I begin the analysis by estimating the return to job training.

A. *Return to Job Training for Young Workers*

Table 3 reports the results from the regression of log wages on education, ability, and training. The training variables included in the regression are cumulative training investment and hours of training activity during the year before the survey. The former is akin to years of schooling in that it captures the total time input to human capital production. The second variable reflects the opportunity cost of training activity in terms of forgone earnings during training. All training variables are annualized using 52

TABLE 3
Return to Job Training

	Dependent Variable: Log Hourly Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
Education	.083	.083				
	(.005)	(.005)				
AFQT	.088	.087				
	(.009)	(.009)				
Cumulative training/(52 × 40)	.079		.093		.078	
	(.013)		(.014)		(.030)	
Current training/(52 × 40)	−.137		−.169		−.156	
	(.044)		(.038)		(.044)	
Cumulative on-the-job training/(52 × 40)		.082		.111		.126
		(.013)		(.015)		(.028)
Current on-the-job training/(52 × 40)		.171		−.041		.018
		(.055)		(.048)		(.062)
Fixed worker effects	No	No	Yes	Yes	Yes	Yes
Observations	26,792	26,792	28,042	28,042	9,432	9,432

Note.—All specifications control for indicators of potential experience, survey year, and race. Robust standard errors are clustered by respondent and are reported in parentheses. The sample includes NLSY 1979 men with 15 years of experience or less. Columns 5 and 6 use observations with less than 5 years of experience.

40-hour weeks. The control variables are indicators for potential experience, race, and survey year.

The estimated return to a year of general job training is 7.9 percent. The return to a year of formal company training is 8.2 percent. Both of these are similar to the return to a year of education, estimated at 8.3 percent. Workers who are currently engaged in general training activities are subject to a 13.7 percent reduction in wages. This effect is consistent with the model in which time is the main input, albeit quantitatively small for a year of full-time training. In fact, workers who have recently received company training appear to have 17.1 percent higher wages. One possibility is that the receipt of company training is correlated with components of worker productivity other than those included in the regression. To get around this, columns 3 and 4 include fixed worker effects in the regression. Since education and ability are also fixed worker traits, they are not included in these estimations. The effect of current training on wages decreases slightly to −16.9 percent for general training and drastically to −4.1 percent for company training, although the latter is not statistically significant. While these estimates suggest that workers do not bear a large part of the training costs, as modeled above, it is also possible that the cost of training is internalized through an employment margin in which the worker does not receive any pay for a period of time, or through a wage margin, but over a longer term. Another possibility is that training investment is more prone to measurement error relative to years of schooling and, therefore, subject to a more severe attenuation bias.

When fixed effects are included, the return to a year of training is 9.3 percent for general training and 11.1 percent for company training, slightly higher than the return to education. The slight increase in the return to training suggests that, like education, the unobserved (fixed) worker characteristics are correlated negatively with training investment.

When there are complementarities between capital stock and time investment, the model predicts that the marginal return to training is similar to that of education for young workers and increases with experience. To test this, columns 5 and 6 repeat the fixed-effects specification for workers with 5 years of experience or less. The return to general training is 7.8 percent, which is lower than the 9.3 percent estimated in the whole sample, although the difference is not statistically significant. While this seems to somewhat support the model's prediction, the return to company training goes in the opposite direction. The estimated return to company training is 12.6 percent for young workers, higher than the 9.6 percent estimated for the sample.

One may be concerned that the use of potential experience instead of actual experience introduces a bias in our results. When actual experience is replaced by potential experience, the error term includes the difference between the actual and potential experience, which is generally negative. If training primarily takes place during actual employment, this difference may be closer to zero for workers with higher accumulated training, generating an upward bias in the estimate of the return to training. A similar bias may also exist in the estimation of the return to education and AFQT if these characteristics are correlated with employment. To address this concern, table 4 reports the estimates using actual experience. The return to a full year of training is estimated at 7.6 percent. The return to education is around 5.4 percent, close to the return to 9 months of job training. When fixed effects are used, the return to training is 11.3 percent, slightly higher than 9.6 percent when potential experience is used. For workers with less than 5 years of experience, the return is lower at 8.6 percent. Overall, table 4 shows that the results are statistically robust to the use of actual experience instead of potential experience.[12]

The estimates of the return to training suggest that training investment is as important as schooling, which supports the notion that training and education are substitutes. The results on the secondary predictions of the model are less clear. Results based on general training activities lend somewhat stronger support to the model than on-site company training.

B. *Ability, Education, and Training Intensity*

The results in the previous section attribute a crucial role for training investment in wages. Then, could the differences in wage growth by education and ability be explained, at least in part, by differences in training? The results presented in Section III are consistent with the human capital model if workers with ability are more likely to receive training and workers with higher education and blacks are less likely to receive training, conditional on ability. In this section I directly test for these results.

Table 5 presents the marginal effects from a Tobit regression of training investment on educational attainment, AFQT, and race. The control variables are indicators for survey years, potential experience, and race. The results show that training investment is strongly related to ability. A one standard deviation higher AFQT score is associated with 132 hours of additional training activity in total and 121 hours of additional company training. Conditional on ability, however, training activity is negatively related to educational attainment. Workers with an additional year of education have 30 hours less training activity in total and receive 27 hours less company training.[13] This is direct evidence for the human capital model presented in the previous section.

By contrast, there is no apparent relationship between race and training, once education and ability are controlled for. The results indicate that blacks have slightly less training activity, but the difference is not statisti-

[12] Since the main conclusion here is that the return to education is comparable to that of training, omitted variables are a concern only if they are correlated with either training or education or with both, but in the opposite direction.

[13] Tobit seems appropriate given the number of zeros in the data. When a least-squares regression is estimated, the corresponding coefficients of ability are 155 hours for total training and 145 for company training, and those of education are −53 for total training and −52 for company training.

TABLE 4
Return to Job Training: Robustness to Actual Experience

	Dependent Variable: Log Hourly Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
Education	.054	.054				
	(.004)	(.004)				
AFQT	.069	.068				
	(.009)	(.009)				
Cumulative training/(52 × 40)	.076		.113		.086	
	(.012)		(.015)		(.030)	
Current training/(52 × 40)	−.155		−.172		−.140	
	(.043)		(.038)		(.046)	
Cumulative on-the-job training/(52 × 40)		.079		.130		.127
		(.013)		(.015)		(.028)
Current on-the-job training/(52 × 40)		.132		−.040		.030
		(.056)		(.049)		(.065)
Fixed worker effects	No	No	Yes	Yes	Yes	Yes
Observations	26,789	26,789	28,036	28,036	9,426	9,426

Note.—All specifications control for indicators of actual experience, survey year, and race. Robust standard errors are clustered by respondent and are reported in parentheses. For comparison with table 3, the sample includes NLSY 1979 men with 15 years of potential experience or less. Columns 5 and 6 use observations with less than 5 years of potential experience.

TABLE 5
The Determinants of Training

	Training Type			
	All (1)	On-the-Job (2)	All (3)	On-the-Job (4)
Education	−29.8	−26.8	−67.6	−58.8
	(8.7)	(8.0)	(12.6)	(11.9)
AFQT	131.6	120.6	72.5	61.1
	(19.5)	(18.1)	(24.6)	(23.4)
Black	−63.9	−59.4	−67.0	−61.4
	(40.6)	(37.6)	(64.4)	(61.8)
Education × exp/10			48.8	40.9
			(11.5)	(10.7)
AFQT × exp/10			69.6	69.5
			(19.6)	(18.8)
Black × exp/10			5.2	4.1
			(51.0)	(49.6)
Observations	26,792	26,792	26,792	26,792

Note.—Results from Tobit regressions. All specifications control for indicators of potential experience, survey year, and race. Robust standard errors are clustered by respondent and are reported in parentheses. The sample includes NLSY 1979 men with 15 years of experience or less.

cally significant. These findings lend support to the validity of the statistical discrimination hypothesis by ruling out human capital formation by training as a viable explanation for racial differences in earnings growth.

Comparable estimates are hard to come by as the earlier literature on the determinants of training has focused on the incidence of training and data on ability were often unavailable. Blundell, Dearden, and Meghir (1996) and Lynch and Black (1998), among others, find a positive relation between the receipt of training and education. This is not conditional on ability. Altonji and Spletzer (1991) include Scholastic Aptitude Test scores of high school seniors in their regressions and find that the receipt of training depends positively on both ability and education for workers with more than 10 years of experience in the National Longitudinal Survey of the High School Class of 1972. Two other studies obtained mixed results. Lynch (1992) reports a higher incidence of training for high school graduates in the NLSY in 1983. For workers with postsecondary education, off-the-job training activity is lower while on-the-job training is similar. Veum (1995) studies the young workers in the NLSY during 1986–90 and finds that the incidence of training is increasing in education, conditional on the AFQT score, with the exception of apprenticeships.

These findings generally suggest that job training and education are positively related, in contrast to the results presented here. The discrepancy appears to stem from the difference between the incidence and the intensity of training. To demonstrate, the incidence of training was regressed on years of education and AFQT, controlling for indicators for potential experience, survey year, and racial background. The findings show that both education and ability are associated with a higher probability of training activity: the coefficient on education is 1.3 percent (standard error [SE] 0.23 percent), and the coefficient on AFQT is 3.5 percent (SE 0.45 percent). Among workers with some training activity, however, the intensity of training varies negatively with education and positively with ability. When total hours of training during the year prior to the survey is regressed on years of education and AFQT (with the same control variables as above), the coefficient on education is −15.4 (SE 3.21) and the coefficient on ability is 7.7 (SE 6.4).

Recall from table 2 that the differences in wage growth by education and ability are particularly strong early in the career. To generate this, the differences in total training investment by education and ability should widen with experience. To test this implication, columns 3 and 4 of table 5 display the marginal effects of ability and education on total training investment by experience. The results show that workers with higher ability invest more in training both initially and later on. A worker with a one standard deviation higher AFQT score engages in 73 hours of additional training initially and continues to invest 70 hours more per decade. A similar result is obtained for company training, where ability is associated with 59 hours of more company training initially and 70 hours of additional investment per decade.

The results show an interesting pattern for the relationship between education and training investment. Conditional on ability, a worker with 1 more year of education invests 68 hours less in training at the beginning of his career but invests 49 hours more per decade during the first 15 years of his career. This stands in contrast to the human capital model discussed above. While education and training are negatively related at the outset of a worker's career, they are rather positively related in general. Those with higher education are likely to experience a higher wage growth conditional on ability as they accumulate longer training hours.

Although the results are qualitatively consistent with the human capital model, the estimated differences in training activity reported in table 5 are not likely to generate large differences in wage growth given the estimated returns to training investment in table 3. A worker with one standard deviation higher ability starts out with 72 hours of additional training. According to the estimates in column 3 of table 3, he initially accepts a 0.59 percent ($= 72 \times -0.169$) cut in wages. Over the course of 15 years he invests an additional 105 hours in training, cumulating 177 hours of training. Evaluated at 9.3 percent return for a full year of training, this amounts to 0.79 percent higher wages relative to a worker with average ability. The implied total wage growth is 1.38 percent over the first 15 years of his career. The coefficient on the interaction of AFQT and potential experience in column 2 of table 1 implies that the observed impact of AFQT on wage growth is 10.2 percent over 15 years, approximately seven times as large as the figure predicted by the training measures. A similar calculation can be made for education. An additional year of education is associated with 67.6 hours less training, which implies 0.55 percent higher wages initially. After 15 years, the difference between cumulated training hours decreases to nearly zero, implying that differences in training by education can explain about 23 percent of the observed 2.4 percent decline in wage growth by education over the 15 years of a worker's career.[14]

An alternative way to test how far the training model goes in explaining the observed patterns in wage growth by education and ability is to include the training measures in the wage regressions. The results are reported in table 6. Relative to the estimates in table 2, the coefficients on the interactions of education and AFQT with potential experience are closer to zero, although the differences are quantitatively small (about 10 percent of the coefficient in table 2) and statistically insignificant.

The results suggest that when training activity is not accounted for, wage regressions are likely to exaggerate the role of employer learning for differences in wage growth by education and ability. Although the training patterns in table 5 are qualitatively consistent with the human capital

[14] The exact calculations are

$$\frac{0.169 \times 72/(52 \times 40) + .093 \times (72 + 70 \times 15/10)/(52 \times 40)}{0.068 \times 15/10}$$
$$= \frac{0.59\% + 0.79\%}{10.2\%} = 14\% = 1/7$$

and

$$\frac{0.169 \times 67.6/(52 \times 40)}{.016 \times 15/10} = 23\%.$$

TABLE 6
Education, Ability, and Wage Growth

	Dependent Variable: Log Hourly Wages			
	(1)	(2)	(3)	(4)
Education	.081	.101	.082	.100
	(.007)	(.007)	(.007)	(.007)
AFQT	.088	.012	.088	.020
	(.009)	(.013)	(.009)	(.014)
Black	−.112	−.111	.025	−.044
	(.022)	(.022)	(.036)	(.039)
Education × exp/5:				
First 5 years	.001	−.019	.000	−.018
	(.006)	(.008)	(.006)	(.008)
6–10 years	.001	−.012	.001	−.011
	(.004)	(.005)	(.004)	(.005)
11–15 years	.000	−.010	.000	−.009
	(.003)	(.004)	(.003)	(.004)
AFQT × exp/5:				
First 5 years		.080		.072
		(.016)		(.018)
6–10 years		.053		.048
		(.009)		(.010)
11–15 years		.039		.034
		(.006)		(.007)
Black × exp/5:				
First 5 years			−.138	−.066
			(.044)	(.049)
6–10 years			−.092	−.045
			(.022)	(.025)
11–15 years			−.067	−.033
			(.015)	(.017)
Observations	26,792	26,792	26,792	26,792

Note.—All specifications control for indicators of potential experience, survey year, cumulative hours of total and company training, and current training investment. Standard errors are clustered by respondent and are reported in parentheses. The sample includes NLSY 1979 men with 15 years of experience or less.

model, the observed levels of training activity suggest that this bias is quantitatively small. If the informal training activity that goes unmeasured in the data displays similar qualitative features, perhaps the differences in wage growth by education and ability could very well be explained by a training model. If, for instance, informal training is seven times larger than the measured formal training as reported in Barron et al. (1997) and if we were willing to speculate that informal training relates to education and ability in the same way as formal training, then total training activity could potentially explain most, if not all, of the observed differences in wage growth.

V. Conclusion

A model of statistical discrimination, where employers use education and racial background as indicators of otherwise unobserved productivity, is a coherent interpretation of observed differences in wage growth among young workers. It provides a rational interpretation for the racial wage gap and explains why education has a diminishing role for wages of more experienced workers with similar abilities. This paper developed and tested a model of human capital as an alternative to the employer learning model.

The findings based on the data on job training from the NLSY lend partial support to the human capital model. Increased intensity of training activity could potentially explain the faster wage growth among young workers with higher cognitive ability. On the other hand, the substitution of training for education appears to be short-lived in the data. Therefore, the negative relationship between wage growth and educational attainment among young workers with comparable abilities remains a challenge for the human capital model. Second, the effect of ability on training investment, while statistically significant, is not large enough to explain the observed rise in wage growth by ability given the estimated return to job training. Part of this is likely due to the fact that most job training is informal and goes unmeasured in the data. Finally, the absence of a meaningful relationship between racial background and job training casts doubt on the human capital theory as a suitable interpretation of the widening racial wage gap by experience.

Appendix A

Technical Appendix

Claim 1. The rate of wage growth by experience is increasing in ability z conditional on educational attainment.

Proof. The proof is done in two steps. First, it is shown that the rate of wage growth is strictly increasing in the role of existing capital stock on human capital accumulation $g_k(l, k)$. Second, it is shown that when $g_k = 0$, wage growth is increasing in ability conditional on education. The derivative of equation (7) is

$$\frac{d(w_x/w)}{dz} = \frac{[zg_K(dK/dz) + zg_l(dl/dz) + g]K - zg(dK/dz)}{K^2} - \frac{dl_x/(1-l)}{dz},$$

which is increasing in g_k since $dK/dz > 0$ given s.

Next, suppose that $g_k = 0$, that is, $\dot{K} = zf(l)$. In this case, total capital stock is linear in z at any experience level. To see this, note that $K(t) = K(0) + \int_0^t zf(l(t))dt$. Equation (7) becomes

$$\frac{w_x}{w} = \frac{zf(l(t))}{K(0) + z\int_0^t f(l(t))dt} - \frac{l_x(t)}{1-l(t)}.$$

Note that, conditional on the path of $l(t)$, wage growth is increasing in z. Next, I show that when $g_k = 0$, investment intensity, $l(t)$, is increasing in ability. This is evident from the optimality condition (4), which becomes

$$e^{-rt}R\left[K(0) + zf(1)s + \int_0^x f(l(t))\,dt\right] = \lambda(t)z_i f'(l)$$

under the assumption for g. The term on the left is the marginal cost, and the term on the right is the marginal benefit of time investment in training. When $K(0) = 0$, z increases the marginal cost and benefit of training proportionally, leading to the same training paths for two workers with different abilities. When $K(0) > 0$, marginal cost responds less than proportionally to ability, encouraging high-ability workers to invest more in training. QED

CLAIM 2. If $g(l(t), K(t)) = f(l)K^{\alpha}$, then the rate of wage growth over experience is decreasing in education conditional on z if $\alpha \le 1$, with equality if and only if $\alpha = 1$.

Proof. With this formulation, we have a finite time control problem in which workers differ only in their capital stock upon graduation from school. The Hamiltonian conditions become[15]

$$-\frac{\partial \mathcal{H}}{\partial K} = -e^{-rt}R(1 - l) - \lambda z_i f(l)\alpha K^{\alpha-1} = \dot{\lambda}(t), \tag{A1}$$

$$\frac{\partial \mathcal{H}}{\partial l} = -e^{-rt}KR + \lambda z_i f'(l)K^{\alpha} = 0, \tag{A2}$$

$$\lambda(T + s)K(T + s) = 0. \tag{A3}$$

Case 1, $\alpha = 1$: Notice that if $\alpha = 1$, this system identifies a time path for $l(t)$ and $\lambda(t)$ that is independent of the existing capital stock. The wage growth over experience is given by the following formula:

$$\frac{\dot{w}}{w} = \frac{\dot{K}(1 - l) - \dot{l}K}{K(1 - l)} = z_i f(l) - \dot{l}/(1 - l),$$

where the second equality follows from the hypothesis of the proposition. Since the time investment, l, is independent of the initial capital stock, wage growth conditional on z_i is identical for all education groups.

Case 2, $\alpha < 1$: The wage growth over experience in this case is

$$\frac{\dot{w}}{w} = z_i f(l)K^{\alpha-1} - \dot{l}/(1 - l). \tag{A4}$$

The first term is decreasing in current capital stock. The second term is positive and is also decreasing in capital stock. To see this, rewrite the first-order condition,

$$f'(l) = \frac{e^{-rt}RK^{1-\alpha}}{\lambda z_i}.$$

Notice that the concavity of $f(l)$ implies that the fraction of time allocated to training is decreasing in the current capital stock for a given λ. Taking logs and differentiating with respect to time, we get

[15] Time arguments are suppressed for convenience.

$$\frac{f''(l)\dot{l}}{f'(l)} = -r + (1-\alpha)z_i f(l)K^{\alpha-1} - \dot{\lambda}/\lambda. \tag{A5}$$

The second term on the right-hand side of this equation is decreasing in K. To investigate the third component, combine the first two Hamiltonians:

$$\dot{\lambda} = -e^{-rt}R[(1-l) + \alpha f(l)/f'(l)].$$

The time path for λ depends only on l. The response in $\dot{\lambda}$ for a given change in l is

$$\frac{\partial\dot{\lambda}}{\partial l} = -e^{-rt}R\{-1 + \alpha[1 - f''f/(f')^2]\}.$$

A rise in time investment slows down the decline in λ if α is low enough. If this is the case, than the right-hand side of equation (A5) is definitely decreasing in K (by the chain rule $\partial\dot{\lambda}/\partial K = \partial\dot{\lambda}/\partial l \times \partial l/\partial K < 0$). Since $f'' < 0$, equation (A5) implies $d\dot{l}/dK > 0$, and hence the wage growth in equation (A4) is decreasing in capital stock.

If α is closer to one, then $\partial\dot{\lambda}/\partial l < 0$, implying that the sign of $d\dot{l}/dK$ is indeterminate. But note that this derivative is monotonic in α since $\partial^2\dot{\lambda}/\partial\alpha\partial l < 0$ and that at best, when $\alpha = 1$, $d\dot{l}/dK = 0$. Therefore, for any $\alpha < 1$, $\dot{l}$ is increasing in current capital and l is decreasing. This ensures that a worker with higher capital upon graduation (due to more schooling) will invest less today and his investment profile will decrease less than a worker with less capital $(-\partial\dot{l}/\partial K < 0)$. Since the terminal point for both workers is the same, $l(T+s)\lambda(T+s) = 0$, the investment profiles will not cross. Hence, wage growth is decreasing in capital stock at all levels of experience. QED

Claim 3.

$$\frac{\partial^2 \ln w}{\partial sx} < \frac{\partial^2 \ln w}{\partial x^2} = \ln \ddot{w}.$$

Proof. Recall that the wage growth is

$$\frac{w_x}{w} = \frac{zh(Q(x+s)) - Q_x(x+s)}{K(x,\, s) - Q(x+s)}.$$

Since $t = x + s$, we have $Q_x(x+s) = Q_s(x+s) = \dot{Q}(t)$, and hence $w_x = \dot{w}$ and $w_{xs} = w_{xx} = \ddot{w}$. Consequently, the numerator changes at the same rate with respect to education, experience, and age. We are done if the denominator rises faster with education than it does with time/experience. Since education and experience enter symmetrically in the investment, this is true if $K_s(x, s) > K_x(x, s)$. The equation for the capital stock at experience level x is

$$\begin{aligned} K(x,\, s) &= K(s) + \int_s^{s+x} zh(Q(t))\,dt \\ &= K(0) + \int_0^s zh(K(t))\,dt + \int_s^{s+x} zh(Q(t))\,dt. \end{aligned}$$

Taking the derivative with respect to x and s separately and taking the difference, we get

$$\begin{aligned} K_s - K_x &= zh(K(s)) + zh(Q(x+s)) - zh(Q(s)) - zh(Q(x+s)) \\ &= z(h(K(s)) - h(Q(s))) = z(h(K(s)) - h(l(s)K(s))) \\ &\geq 0, \end{aligned}$$

where the last inequality follows from the fact that $l(s) \leq 1$ at the time of graduation (which in turn comes from $\psi \geq 0$). QED

CLAIM 4. The rate of wage growth is decreasing in education if $rzh'''(\cdot) \geq [zh''(\cdot)]^2[1 - e^{-2r(T-t)}]$.

Proof. From the previous claim we have, $w_{xs} = w_{xx} = \ddot{w}$. Since the wage itself (the denominator) is increasing in education, it would suffice to find a condition that guarantees $\ddot{w} \leq 0$:

$$\ddot{w} = zh'(Q)\dot{Q} - \ddot{Q}.$$

Recall the optimality condition for training from footnote 7:

$$z_i h'(Q(t))\{1 - e^{-r[T(s)-t]}\} = r. \tag{A6}$$

Applying the implicit function theorem to equation (A6), we get

$$\dot{Q} = -\frac{r^2 e^{-r(T-t)}}{[1 - e^{-r(T-t)}]^2 zh''(Q)}.$$

Taking the time derivative one more time and simplifying terms, we get

$$\ddot{Q} = \frac{r^3 e^{-r(T-t)}[1 - e^{-2r(T-t)}]zh''(Q) - r^4 e^{-2r(T-t)} h'''(Q)/h''(Q)}{[1 - e^{-r(T-t)}]^4 [zh''(Q)]^2}.$$

Plugging these in the equation for $\ddot{w}$ yields

$$\ddot{w} = r^2 e^{-r(T-t)} \left\{ z^2 h' h'' [1 - e^{-r(T-t)}]^2 - r[1 - e^{-2r(T-t)}]zh'' + r^2 e^{-r(T-t)} \frac{h'''}{h''} \right\}.$$

When we replace $zg'(Q)$ with $r/[1 - e^{-r(T-t)}]$ from equation (A6), this term is nonpositive if and only if

$$rzh'''(\cdot) \geq [zh''(\cdot)]^2[1 - e^{-2r(T-t)}].$$

QED

Appendix B

Data

The NLSY79 includes three sampling groups: a random sample representative of the population, a supplemental sample of blacks and poor whites, and a military sample. The sample used in the analysis is restricted to the 3,003 male respondents from the nationally representative sample: 192 respondents with missing AFQT scores were dropped from the sample, and 32 additional respondents who had less than 8 years of completed education were excluded. Wages are defined as hourly earnings, and all nominal values are converted to 2004 dollars using the consumer price index. Additional observations were excluded if the respondent reported less than \$1 or more than \$100 in hourly earnings or more than

2,000 hours of training in a calendar year, or if the respondent was enrolled at school at the time of the survey. This leaves a total of 2,733 respondents with 26,792 observations for the estimations. The AFQT score was standardized by each age group as the respondents were of different ages when they took the test.

To keep a record of training activities, the respondents were asked whether they had participated in any vocational or technical training programs. Then, detailed information on the start and end dates of each training spell, the intensity of training—measured in hours per week—and the type of the training activity were collected. In particular, the respondents indicated one of the following categories: (1) business college, (2) nursing program, (3) apprenticeship, (4) vocational/technical institute, (5) barber/beauty school, (6) flight school, (7) correspondence school, (8) company training, and (9) other. Beginning in 1988, category 8 was split into three as (8) formal company training, (9) seminars or training programs at work not run by employer, and (10) seminars or training programs outside of work. I group the reported categories in two different ways and analyze them separately. First, I test the implications of the model using all training activities. Second, I present the results separately for on-site company training (category 8 before 1988 and categories 8 and 9 after 1988). These categories are often considered as on-the-job training spells in the literature (e.g., Parent 1999).

Appendix C

Robustness Analysis

This appendix reports the results using the imputed hours of training for short spells (less than a month). To construct the imputed data, total hours of training for spells that lasted less than a month were projected on educational attainment, AFQT score, and indicators of potential experience and racial background using surveys after 1988. The predictions from the regression were used to fill in the

TABLE C1
Return to Job Training: Imputed Training Hours for Short Spells Prior to 1988

	Dependent Variable: Log Hourly Wages		
	(1)	(3)	(5)
Education	.083		
	(.005)		
AFQT	.088		
	(.009)		
Cumulative training/(52 × 40)	.080	.094	.079
	(.031)	(.014)	(.030)
Current training/(52 × 40)	−.139	−.169	−.156
	(.044)	(.038)	(.044)
Fixed worker effects	No	Yes	Yes
Observations	26,792	28,042	9,432

Note.—All specifications control for indicators of potential experience, survey year, and race. Robust standard errors are clustered by respondent and are reported in parentheses. The sample includes NLSY 1979 men with 15 years of experience or less. Column 3 uses observations with less than 5 years of experience. The column numbers refer to the corresponding specifications from table 3.

TABLE C2
THE DETERMINANTS OF TRAINING: IMPUTED TRAINING HOURS FOR SHORT SPELLS PRIOR TO 1988

	Training Type: All	
	(1)	(2)
Education	−23.4	−50.9
	(8.6)	(11.8)
AFQT	139.6	76.1
	(19.5)	(23.4)
Black	−67.9	−56.2
	(40.5)	(61.6)
Education × exp/10		36.3
		(11.3)
AFQT × exp/10		76.3
		(19.5)
Black × exp/10		−12.5
		(47.9)
Observations	26,792	26,792

Note.—Results from are Tobit regressions. All specifications control for indicators of potential experience, survey year, and race. Robust standard errors are clustered by respondent and are reported in parentheses. The sample includes NLSY 1979 men with 15 years of experience or less.

corresponding values for short spells prior to 1988. Since the type of the training activity is missing for these spells, tables C1 and C2 report the results for total training only.

References

Altonji, Joseph G., and C. Pierret. 2001. "Employer Learning and Statistical Discrimination." *Q.J.E.* 116 (1): 313–50.

Altonji, Joseph G., and James R. Spletzer. 1991. "Worker Characteristics, Job Characteristics, and the Receipt of On-the-Job Training." *Indus. and Labor Relations Rev.* 45 (1): 58–79.

Barron, J. M., M. C. Berger, and D. A. Black. 1997. "How Well Do We Measure Training?" *J. Labor Econ.* 15 (3): 507–29.

Ben-Porath, Yoram. 1967. "The Production of Human Capital and the Life Cycle of Earnings." *J.P.E.* 75, no. 4, pt. 1 (August): 352–65.

Blundell, Richard, L. Dearden, and Costas Meghir. 1996. "The Determinants and Effects of Work-Related Training in Britain." Manuscript, Inst. Fiscal Studies, London.

Browning, Martin, Lars Peter Hansen, and James J. Heckman. 1999. "Micro Data and General Equilibrium Models." In *Handbook of Macroeconomics*, vol. 1A, edited by John B. Taylor and Michael Woodford. Amsterdam: North-Holland.

Ehrlich, Isaac, and Jinyoung Kim. 2007. "The Evolution of Income and Fertility Inequalities over the Course of Economic Development: A Human Capital Perspective." *J. Human Capital* 1 (1): 137–74.

Farber, Henry S., and Robert Gibbons. 1996. "Learning and Wage Dynamics." *Q.J.E.* 111 (4): 1007–47.

Gibbons, Robert, and L. F. Katz. 1991. "Layoffs and Lemons." *J. Labor Econ.* 9: 351–80.

Gibbons, Robert, and M. Waldman. 1999. "A Theory of Wage and Promotion Dynamics inside Firms." *Q.J.E.* 114:1321–58.

———. 2006. "Enriching a Theory of Wage and Promotion Dynamics inside Firms." *J. Labor Econ.* 24 (1): 59–107.

Griliches, Zvi. 1977. "Estimating the Returns to Schooling: Some Econometric Problems." *Econometrica* 45 (1): 1–22.

Heckman, James J., Lance J. Lochner, and Chris Taber. 1998a. "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents." *Rev. Econ. Dynamics* 1:1–58.

———. 1998b. "Tax Policy and Human-Capital Formation." *A.E.R. Papers and Proc.* 88 (2): 293–97.

Heckman, James J., Lance J. Lochner, and Petra E. Todd. 2008. "Earnings Functions and Rates of Return." *J. Human Capital* 2 (1): 1–31.

Kahn, Lisa, and Fabian Lange. 2010. "Employer Learning, Productivity and the Earnings Distribution: Evidence from Performance Measures." IZA Working Paper no. 5054, Inst. Study Labor, Bonn.

Lange, Fabian. 2007. "The Speed of Employer Learning." *J. Labor Econ.* 25 (1): 1–35.

Lynch, Lisa M. 1992. "Private-Sector Training and the Earnings of Young Workers." *A.E.R.* 82 (1): 299–312.

Lynch, Lisa M., and Sandra E. Black. 1998. "Beyond the Incidence of Employer-Provided Training." *Indus. and Labor Relations Rev.* 52 (1): 64–81.

Meghir, Costas, and Luigi Pistaferri. 2011. "Earnings, Consumption and Life Cycle Choices." In *Handbook of Labor Economics*, vol. 4B, edited by David Card and Orley Ashenfelter, 773–854. Amsterdam: Elsevier Sci.

Parent, Daniel. 1999. "Wages and Mobility: The Impact of Employer-Provided Training." *J. Labor Econ.* 17 (2): 298–317.

Schönberg, Uta. 2007. "Testing for Asymmetric Employer Learning." *J. Labor Econ.* 25:651–91.

Veum, Jonathan R. 1995. "Sources of Training and Their Impact on Wages." *Indus. and Labor Relations Rev.* 48 (4): 812–26.

Willis, Robert J., and Sherwin Rosen. 1979. "Education and Self-Selection." *J.P.E.* 87, no. 5, pt. 2 (October): S7–S36.