The Earnings Impact of Training Duration in a Developing Country

An Ordered Probit Selection Model of Colombia's *Servicio Nacional de Aprendizaje* (SENA)

Emmanuel Jimenez and Bernardo Kugler

ABSTRACT

This study estimates the earnings impact of one of the most extensive in-service training programs in the developing world, Colombia's SENA, through a comparison of nongraduates' and graduates' earnings profiles. The possible influence of selection is corrected through the use of an ordered probit model to determine participation in short and long courses. The correction causes a significant decline in SENA's earnings effect. The results are robust with respect to alternative specifications of the selection model.

I. Introduction

It is widely recognized that investment in human capital is a prerequisite for economic development. The formation of skills through vocational or in-service training is an important and growing component of such investment. Among developing countries, the commitment to vocational and technical training is most established in Latin America. By 1972, all Latin American countries had established government-run vocational training courses (Kugler and Reyes 1978). One of the oldest and largest of

The authors are economists in the education sector of The World Bank. Jimenez is also a professor of economics at the University of Western Ontario. *The authors would like to acknowledge an anonymous referee's comments, as well as the computing assistance of Messrs. Haeduk Lee and Kyuee-Ha Pahk, in the preparation of this paper. They also express their appreciation to SENA for making the data available for this study. The views expressed herein are those of the authors and should not be attributed to SENA or the World Bank.*

these is Colombia's SENA (Servicio Nacional de Aprendizaje). SENA was instituted in 1957, spends an equivalent of about one-tenth of the public budget for education, is spread over the whole country, and has trained about 15 percent of the urban labor force.

Despite the large investment in developing countries, the economic impact of training institutions has traditionally received little research attention, especially in comparison with the studies on formal education (Metcalf 1985). This paper attempts to estimate the impact of SENA training on workers' incomes, through a comparison of SENA graduates' earnings with those of a control group. This data set is based on a sample survey completed in 1981 and is unique for a developing country because of the detailed income information it contains.

A typical problem that arises with analyses of micro-level cross-section data is the possibility that the choice of a control group influences measured differences in earnings between participants and nonparticipants. Suppose that ability and parents' socioeconomic characteristics affect present earnings of individuals. In order to be able to attribute these measured differences to the impact of participating in SENA, the researcher must be reasonably certain that they are not due partly to the fact that those who participate have different characteristics from those who do not. One way to accomplish this is to control for characteristics in the earnings functions. However, this procedure would not be sufficient if the sampled individuals, as economically rational agents, select the options that optimize their income-earnings potential. In this case, the researcher would be unable to infer an unbiased estimate of the true earnings potential of participating in SENA, since the estimate would be based solely on participants. The sample would simply not yield any observations of participants with nonparticipant characteristics—nonoptimal choices are not observed. By a similar argument, the researcher could not obtain an unbiased estimate of the earnings stream of those choosing not to participate.

Over the past decade, important methodological advances have been made in order to resolve statistically this so-called selection problem. Because it arises in many economic topics, a considerable literature has emerged.[1] In the human capital literature, the bias is treated by estimating an additional equation that determines the occurrence of a binary event (college attendance versus not in Willis and Rosen 1979 and Kenny et al. 1979) and then using the estimated equation to correct for the truncation. More recently the model has been extended to allow for the possibility that

1. Much of it has been based on Heckman's seminal work (1976, 1979) which treats the bias as a specification error, and his application to the labor force participation decision. Some other examples are Lee and Trost (1978) for housing tenure choice, Lee (1978) for union membership. See Maddala (1983) for a thorough review.

individuals have more than two (polychotomous) choices. A general model developed by Lee (1983) has been applied to correct for the bias in the impact of two-year technical school training by Trost and Lee (1984), who first estimate the determinants of the choice to attend high school only, technical school after high school, and college after high school. Another model (Garen 1984) poses a slightly different model which extends the selection bias approach to the case where the choice variable is continuous and the choice set is ordered, such as the choice of years of schooling. An ordered choice set is one in which one value of the choice variable is clearly greater or less than another value.

This paper will attempt to adopt this methodology to test for the importance of selectivity in estimating the earnings effect of SENA. In the process, some insights into the determinants of selection into SENA will be discussed. Program participants are confronted with discrete choices of how much SENA training to take. In particular, the data base allows us to distinguish between program participants who choose to take SENA's short courses (three to six months in duration) from those who take SENA's long courses (up to two and a half years in duration). An ordered probit selection model, which sorts the sample according to the length of training, is presented to account for this institutional characteristic. This model is an extension of Garen's ordered continuous choice sets to ordered discrete choice sets. In many cases, such as the one investigated here, continuous choices cannot be observed.

II. The Model

The first step in addressing the selection problem is to model the decision to attend SENA. SENA training can be taken as long courses, which last on average a little less than two years and are meant primarily for individuals who have finished primary and junior secondary school but who have little or no practical work experience. In addition, there are SENA short courses, which last an average of two and one-half months. Opportunity costs are minimized on this option. Finally, there are individuals who chose no SENA training at all. On average, these individuals accumulate the same number of years of formal schooling as long-course graduates. The sampled workers are assumed to compare the net present values of their expected future incomes (V) under each mode of training activity and choose that activity where V is highest.

Let J represent the outcome of a trichotomous choice by an individual. If an individual decides to take no SENA courses $J = 0$; if a short course $J = 1$; and if a long course $J = 2$. If a representative individual chooses not to attend SENA, the present value of earnings (V_0) is:

(1) $$V_0 = \int_0^N y_0(t) \exp(-rt)\, dt \qquad (\text{if } J = 0);$$

where $y_0(t)$ = earnings at time t of someone who has never undergone vocational training at SENA; N is the duration of that individual's working life; and r is a rate of time discount. If a representative individual chooses to attend SENA, the present value of earnings would depend upon whether short courses (V_1) or long courses (V_2) were chosen:

(2) $$V_1 = \int_{T_1}^n y_1(t) \exp(-rt)\, dt \qquad (\text{if } J = 1);$$

(3) $$V_2 = \int_{T_2}^n y_2(t) \exp(-rt)\, dt \qquad (\text{if } J = 2);$$

where T_1 and T_2 are, respectively, the duration (in years) of short and long course training. Each individual chooses the training mode J associated with the highest V.

Current earnings at time t for each training mode are determined by:

(4) $$y_0(t) = \bar{y}_0 \exp[g_0(t-s)] \qquad \text{for } J = 0;$$

(5) $$y_1(t) = \bar{y}_1 \exp[g_1(t-T_1-s)] \qquad \text{for } J = 1; \text{ and}$$

(6) $$y_2(t) = \bar{y}_2 \exp[g_2(t-T_2-s)] \qquad \text{for } J = 2;$$

where $\bar{y}_0$, and $\bar{y}_1$, and $\bar{y}_2$ denote initial (posttraining) earnings; s is years of schooling; and g_0, g_1, and g_2 denote the rate of growth of earnings.[2] The specifications of initial earnings are assumed to depend upon:

(7) $$\ln \bar{y}_J = f_J(X_J) + e_J$$

for J = 0, 1, 2; X_J is a vector of observable explanatory variables; f_J, are functions whose parameters are to be estimated; and e_J denote permanent person-specific unobserved components reflecting unmeasured factors influencing earnings potential. One of these unmeasured effects is natural ability. In addition, the rate of discount is determined by:

(8) $$r = k(S) + e$$

where k is a function and S is a vector of observable explanatory variables that affect the SENA decision through their effect on the discount rate. These variables include the socioeconomic characteristics of trainees' families, which determine the ability of any individuals to forego opportunity costs. The term e is a permanent unobserved component influencing finan-

2. For simplicity, the rate of growth of earnings is assumed constant in this paper. It can also be assumed to be stochastically determined by X_J without affecting the model, because of linearity assumptions.

cial barriers to undergoing vocational training. The vector of error terms in (7), and (8) are assumed to be randomly distributed.

All trainees make economically rational decisions regarding the choice of training length. The optimal hours length J^* is chosen such that $\ln (V_J^*/V_i) > 0$ for all $J^* \neq i$. Substituting (4) to (6) and (8) into (1) to (3), the evaluation of the integrals, and linearization will yield that the optimal length of course, $\ln (V_j^*/V_i) = J^*$, is determined by:

$$(9) \quad J^* = Z\phi + \epsilon$$

where Z includes a vector of constants, X and S, and ϵ is a random error term, which is assumed to be normally distributed with zero mean and a variance of σ_ϵ^2. Since the categories of participation are measured by the length of the course, $J = 2$ if $c < J^*$, $J = 1$ if $0 < J^* < c$, $J = 0$ if $J^* \leq 0$ where c is an unobserved cutoff point which delineates the length of course. The probabilities of choosing each mode are:

$$(10) \quad \Pr(\text{LONG}) = \Pr(J = 2) = \Pr(c < J^*) = \Pr(c < Z\phi + \epsilon) = F[(Z\phi - c)/\sigma_\epsilon]$$

$$(11) \quad \Pr(\text{SHORT}) = \Pr(J = 1) = \Pr(0 < J^* < c) = \Pr(0 < Z\phi + \epsilon < c) = F[(c - Z\phi)/\sigma_\epsilon] - F(-Z\phi/\sigma_\epsilon)$$

$$(12) \quad \Pr(\text{no SENA}) = \Pr(J = 0) = \Pr(J^* \leq 0) = 1 - F(Z\phi/\sigma_\epsilon)$$

If $F(\,.\,)$ is a standard normal *c.d.f.*, this is an ordered probit model that can be estimated by maximum likelihood.

Application to the selection problem:

Let the earnings functions be of the standard semilogarithmic form. If α's, β's, and γ's are parameters to be estimated, then:

$$(13) \quad \ln(y_0) = \alpha_0 + \alpha_1 s + \alpha_2 x + \alpha_3 x^2 + \alpha_4 sx + \mu_0 \qquad \text{if } J = 0$$

$$(14) \quad \ln(y_1) = \beta_0 + \beta_1 s + \beta_2 x + \beta_3 x^2 + \beta_4 sx + \mu_1 \qquad \text{if } J = 1$$

$$(15) \quad \ln(y_2) = \gamma_0 + \gamma_1 s + \gamma_2 x + \gamma_3 x^2 + \gamma_4 sx + \mu_2 \qquad \text{if } J = 2,$$

where s denotes years of schooling and x, years of experience. It is assumed that μ_i are normal and $E(\mu_i) = 0$; $E(\mu_i \mu_i) = \sigma_{ij}$ for $i = j$ and 0 otherwise; and the covariance between μ_i and the explanatory variables is zero. These assumptions are valid if the samples are drawn randomly from a population. Thus, for a given level of s and x, estimates of the parameters by OLS can be used to calculate the effect of training (e.g., $\ln(y_1) - \ln(y_0)$, for the short course effect) since:

(16) $E(\ln y_0 \mid s,x) = \alpha_0 + \alpha_1 s + \alpha_2 x + \alpha_3 x^2 + \alpha_4 sx + E(\mu_0 \mid J = 0)$

(17) $E(\ln y_1 \mid s,x) = \beta_0 + \beta_1 s + \beta_2 x + \beta_3 x^2 + \beta_4 sx + E(\mu_1 \mid J = 1)$

(18) $E(\ln y_2 \mid s,x) = \gamma_0 + \gamma_1 s + \gamma_2 x + \gamma_3 x^2 + \gamma_4 sx + E(\mu_2 \mid J = 2)$

The moments of the conditional normal show that the last terms of (16) to (18) can be interpreted in terms of the conditional probabilities (10) to (12). The last term of (17) can be written as:

$$E(\mu_1 \mid J = 1) = E(\mu_1 \mid 0 < J \leq c)$$

(19)
$$= \frac{\sigma_{1\epsilon}}{\sigma_\epsilon}\left\{-\frac{f[(c - Z\phi)/\sigma_\epsilon] + f[(-Z\phi)/\sigma_\epsilon]}{F[(c - Z\phi)/\sigma_\epsilon] - F[(-Z\phi)/\sigma_\epsilon]}\right\} = \beta_5 \lambda_1,$$

where $\sigma_{1\epsilon}$ = correlation coefficient between μ and ϵ, $\sigma_{1\epsilon}/\sigma_\epsilon = \beta_5$, the term within $\{.\} = \lambda_1$. The latter can be computed from the ordered probit equation and added as an explanatory variable in (14). If $\lambda_1 < 0$, and if $\beta_5 < 0$, then, short-course enrollees are a positive, self-selected group from the earnings distribution. Similarly, it can be shown that:

(20)
$$E(\mu_0 \mid J = 0) = \frac{\sigma_{0\epsilon}}{\sigma_\epsilon}\left\{-\frac{f[(-Z\phi)/\sigma_\epsilon]}{F[(-Z\phi)/\sigma_\epsilon]}\right\} = \alpha_5 \lambda_0,$$

(21)
$$E(\mu_2 \mid J = 2) = \frac{\sigma_{2\epsilon}}{\sigma_\epsilon}\left\{\frac{f[(c - Z\phi)/\sigma_\epsilon)]}{1 - F[(c - \phi)/\sigma_\epsilon)]}\right\} = \gamma_5 \lambda_2,$$

where $\lambda_0 < 0$ and $\lambda_2 > 0$.

The λ's can be obtained from the ordered probit equations describing the choice of how much SENA training to take. The inclusion of λ_0, λ_1, and λ_2 into the earnings functions would allow the estimation of α_5, β_5, and γ_5. Together, these values can be used to estimate the selection bias. By the principle of comparative advantage described in Willis and Rosen (1979), the prior hypotheses are that $\alpha_5 < 0$ and $\gamma_5 > 0$. The sign of β_5 is ambiguous, since short-course enrollees will have values of J^* in the middle range.

III. Data

The data used in this analysis are drawn from the Bogota subsample of an extensive nationwide survey of graduates conducted by SENA itself between 1979–81. The sampling procedures are described in SENA-Holanda (1982). The graduates were asked detailed questions about the courses that they took, the socioeconomic status of their family, and their employment and earnings history. In addition, SENA surveyed a control group of workers, who never undertook any SENA training but who worked in the same types of firms as the graduates.

The analysis focuses only on male workers, who constituted more than one-half of the entire sample. After females and those observations with incomplete information are deleted, the sample consists of 1,004 control group individuals and 1,348 SENA graduates. SENA courses can be divided into long courses, which last approximately two years and some of which have a substantial in-plant training component, and short courses, which last about three months. Of the SENA graduates in the sample, 310 were short-course and 1,038 were long-course attendees.

The means of the key variables are summarized in Table 1 for each of the subgroups used in the analysis. The mean earnings of the SENA subsample is approximately 9 percent higher than that of the non-SENA subsample. Schooling attainment is slightly less, due mostly to the relatively low mean years of schooling for those taking short course, SENA graduates also tend to have more on-the-job experience, measured as age minus years of schooling minus the duration of SENA training. Again, the short-course individuals account for much of this difference, since they average about 18.7 years of experience, as compared to long-course graduates, with only about 12 years. Non-SENA individuals have about 12.5 years.

The second group of variables includes the determinants of the choice of participating in SENA short or long courses or not participating at all—the Z vector in the notation of the preceding section. The first six variables of the Z vector contain information about the socioeconomic status of the individual's father—his schooling and an index of his occupational prestige, as measured by a two-digit industrial classification. The non-SENA sample tends to come from households with fathers having a greater educational attainment and who were engaged in higher prestige occupations. The next three variables also suggest that there is very little difference in the migration histories of the subgroups. Finally, the SENA group tends to have had its primary schooling interrupted more often than the non-SENA group.

Unfortunately, the data base did not contain any information about innate ability, such as Intelligence Quotient (IQ). Since the survey was conducted only as a once-and-for-all cross-section, it is not possible to net out possible ability effects by analyzing only differences in earnings before and after training. Thus, if innate ability is an important source of self-selection, its effect can only be inferred from the error terms. Part of the error in the earnings equations can be interpreted as the unobserved component of comparative advantage.

Table 1 indicates that there may be systematic differences, not only between SENA and non-SENA participants, but among SENA graduates themselves. Short-course graduates, who have less formal schooling and more experience than the rest of the sample, also tend to come from more

Table 1
Means (and Standard Deviations) of Key Variables

			SENA		
Variable	Description	Non-SENA	Overall	Short-course graduates	Long-course graduates
Earnings Function Variables					
LY	Log (earnings)	12.03	12.12	11.57	12.28
		(0.58)	(0.90)	(1.04)	(0.79)
s	Years of schooling	9.90	9.26	7.02	9.92
		(3.90)	(3.46)	(3.47)	(3.16)
x	Years of experience	12.48	13.52	18.66	11.99
		(9.09)	(9.26)	(11.98)	(7.64)
Z Variables					
FPRIM	= 1 if father had some primary education	0.15	0.22	0.31	0.20
		(0.36)	(0.42)	(0.46)	(0.40)
FPRIMED	= 1 if father finished primary education	0.33	0.37	0.29	0.39
		(0.47)	(0.48)	(0.45)	(0.50)
FSECED	= 1 if father had any secondary education	0.31	0.26	0.16	0.28
		(0.46)	(0.44)	(0.37)	(0.45)
FHIGHED	= 1 if father had any higher education	0.11	0.05	0.02	0.05
		(0.31)	(0.21)	(0.15)	(0.22)
FNRED	= 1 if father's education not reported	0.04	0.02	0.02	0.02
		(0.20)	(0.14)	(0.13)	(0.15)
FOCC	Occupational Status index (1 = high; 10 = low)	4.96	5.26	5.38	5.23
		(1.46)	(1.26)	(1.16)	(1.29)
MOBILB	= 1 if moved since birth	0.55	0.54	0.75	0.48
		(0.50)	(0.50)	(0.44)	(0.50)
MOBILBP	= 1 if moved between birth and primary	0.22	0.22	0.21	0.23
		(0.42)	(0.42)	(0.41)	(0.42)
MOBILP	= 1 if moved since primary	0.46	0.43	0.66	0.36
		(0.50)	(0.50)	(0.47)	(0.48)
RURALED	= 1 if primary school in rural area	0.15	0.20	0.44	0.13
		(0.36)	(0.40)	(0.50)	(0.34)
IN-TERRUP	= 1 if primary schooling interrupted	0.15	0.30	0.23	0.32
		(0.36)	(0.46)	(0.42)	(0.47)
Correcting Terms from Ordered Probit					
λ_0	Equation (20)	−.878	n.a.	n.a.	n.a.
		(.197)			
λ_1	Equation (21)	n.a.	n.a.	−.014	n.a.
				(.273)	
λ_2	Equation (22)	n.a.	n.a.	n.a.	.853
					(.203)
Number of observations		1,004	1,348	310	1,038

Note: n.a. means not applicable.

mobile families where the father stopped formal schooling at the primary level. If these factors affect the choice to enter short or long course SENA training, it is important to apply the ordered model described earlier to control for possible selectivity biases.

IV. Results

The results of the first step of the analysis—the estimation of the ordered probit model of how much SENA training to pursue—are presented in Table 2. The signs of the other coefficients indicate that, in relation to individuals whose fathers had no schooling, those whose fathers had some or completed primary were more likely to pursue more SENA training. However, those whose fathers attended secondary or higher education were likely to pursue less SENA training. Thus, SENA is not likely to attract the lowest (for affordability reasons) or the highest socioeconomic groups who can afford the higher opportunity costs of formal academic schooling. It is more likely to attract those whose primary schooling has been interrupted.

The results of Table 2 are used to generate estimates of λ_0, λ_1, and λ_2. These are then used to test and correct for selectivity bias in the earnings

Table 2
Reduced-Form Ordered Probit

Variable	Coefficient	Standard Error
INTERCEP	−0.035	0.1730
FPRIM	0.215	0.1228
FPRIMED	0.1639	0.1166
FSECED	−0.0021	0.1209
FHIGHED	−0.3881	0.1539
FNRED	−0.2694	0.1807
FOCC	0.0330	0.0212
MOBILB	−0.0875	0.0835
MOBILBP	0.0133	0.0713
MOBILP	−0.2137	0.0786
RURALED	0.0658	0.0777
INTERRUP	0.5206	0.0598
LOG-LIKL	−2,254.9	

functions. The mean levels (and standard deviations) of λ are shown in the last three rows of Table 1. Only $\hat{\lambda}_1$ is statistically insignificant.

Table 3 summarizes the estimated earnings functions. The first three columns of coefficients and standard errors present standard OLS earnings functions parameters without any corrections for selectivity. The last three columns include the estimated selection variables, $\hat{\lambda}_0$, $\hat{\lambda}_1$, and $\hat{\lambda}_2$, computed from the ordered probit equation, in the earnings functions regressions. The coefficients of schooling and experience are of the right sign and are of reasonable magnitude. With no experience, and with Mincer's assumptions, the rate of return to schooling, under both specifications (corrected and uncorrected for selectivity) is about 14 percent for the SENA subsample and 13 percent for the non-SENA subsample.

The interpretation of the selection bias, as measured by LAMBDA times its coefficient, is of primary interest in this paper. LAMBDA is an index of the likelihood of each observation being in-SENA due to individual characteristics. According to Table 1 the value of LAMBDA is negative for

Table 3
Earnings Functions[a]

	Uncorrected OLS			Corrected for Selection		
Variable	Non-SENA	SENA Short-Course Graduates	SENA Long-Course Graduates	Non-SENA	SENA Short-Course Graduates	SENA Long-Course Graduates
Intercept	10.2180	9.9688	10.2835	10.3770	9.9521	10.0950
	(.1000)	(.3219)	(.1828)	(.1256)	(.3185)	(.2089)
s	0.1281	0.1427	0.1366	.1260	.1428	.1379
	(.0077)	(.0283)	(.0142)	(.0077)	(.0280)	(.0141)
x	0.0854	0.0375	0.0973	.0848	.0399	.0988
	(.0081)	(.0220)	(0.170)	(.0081)	(.0218)	(.0169)
x squared	−0.0010	−0.0007	−0.0016	−.0011	−.0007	−.0016
	(.0001)	(.0003)	(.0003)	(.0001)	(.0003)	(.0003)
sx	−0.0025	0.0022	−0.0019	−.0025	.0019	−.0020
	(.0005)	(.0015)	(.0015)	(.0005)	(.0015)	(.0011)
LAMBDA (λ)	n.a.	n.a.	n.a.	.1559	−.2712	.2009
				(.075)	(.1958)	(.1093)
R squared	.37	.28	.19	.37	.28	.19

Note: n.a. means not applicable.
a. Standard errors in parentheses. These errors are correctly estimated, with the use of the LIMDEP program. All figures are rounded off to the nearest ten thousandths.

non-SENA sample and for SENA short-course graduates and positive for SENA long-course graduates. The coefficient for the long-course group is positive (as hypothesized, given $\lambda_1 < 0$) and is statistically significant at 90 percent in the earnings functions of Table 3. For the short-course group, the coefficient of LAMBDA is negative (as hypothesized) and thus the bias is also positive although at a lower significance level. We conclude that for the SENA subsample, an ordinary least squares (OLS) regression on earnings would overstate the true intercept, since it would include a positive selection term. Had individuals with socioeconomic characteristics similar to the non-SENA subsample taken SENA, average SENA earnings would be lower than observed. For the non-SENA group, the coefficient of LAMBDA is positive and significant. An OLS regession for this subsample would understate the intercept term, which is contrary to the hypothesis of comparative advantage. There would tend to be negative selection into the non-SENA group from the overall earnings distribution.

The positive selection into the training program indicates that affordability plays an important role. A more stable and higher status family background increases the chances of additional SENA training. These variables also affect schooling and thus, earnings.

Thus, if one were to use earnings functions estimated by OLS, the SENA effect would be overestimated for both short and long courses. This is because effect of including a selection term on the intercepts of the SENA (positive) and non-SENA (negative) subsamples are not offset by large changes in the coefficients (compare corrected and uncorrected panels in Table 3). The extent of the bias depends upon assumed levels of schooling and experience. As an illustration, Table 4 compares the earnings of SENA graduates with those who have not had SENA training, holding constant for different levels of schooling and experience. In these examples, formal academic schooling is fixed at two alternative levels: the junior secondary level (incomplete secondary), ten years, which is also the overall sample's mean level of schooling; and the complete secondary level, or 12 years. Experience is made to range from 19 to 29 years for the short-course sample; and twelve to 20 for long course individuals. Short-courses are designed for older participants.

According to Table 4, the correction for selectivity would significantly lower the magnitude of SENA's effect on earnings, as estimated by predicted earnings differentials from OLS estimates. The effect of SENA short courses rises with schooling and experience. For individuals with incomplete secondary schooling but with 29 years of experience, the SENA earnings effect based on OLS estimates would be 18 percent; however, if selectivity is taken into account, that effect would be nil. For an individual with complete secondary and 19 years of experience, the SENA effect drops from 24 percent to 4 percent due to selectivity; with more experience at 24 years, the

Table 4
Uncorrected and Selectivity-Corrected Effects of SENA on Earnings

			Earnings Effect of SENA (%)[a]	
Course Type	(1) Years of Schooling	(2) Years of Experience	(3) Uncorrected OLS	(4) Selectivity Corrected
Short	10	19	.9	−14.6
Short	10	29	18.4	−0.1
Short	12	19	24.3	4.0
Short	12	24	39.9	16.4
Long	10	12	33.7	−1.8
Long	10	20	35.2	−.9
Long	12	12	37.9	1.7
Long	12	16	40.5	3.5

a. Percentage differential from the non-SENA control group, at given levels of schooling and experience.

impact of a SENA short course would drop from 40 percent to 16 percent upon the selectivity correction.

Similar results hold for the effect of SENA long-courses, as shown in the last four rows of Table 4. The extent to which the SENA effect is lowered by the selectivity correction is also significant. For example, a secondary school graduate with 12 years of experience would gain 38 percent in earnings due to SENA, based on OLS estimates—but only 2 percent based on corrected estimates.

Most of the difference between the uncorrected and corrected earnings effects [Column (3) minus Column (4), Table 4] can be attributed to the impact on the intercept terms of including LAMBDA in the earnings function—rather than to the impact on the coefficients of the schooling and experience variables. For short courses the corrected minus uncorrected differences in earnings effects range between 14 and 24 percentage points, for the values of s and x shown in Table 4. In comparison, differences in earnings effects due solely to difference in the earnings functions coefficients span only about 1 to 5 percentage points. The rest is due to the intercept term. For long courses, the corrected minus uncorrected difference in earnings effects is 36 to 37 percentage points; the earnings effects difference due to the coefficients is only about 4 to 6 percentage points.

The set of explanatory variables in the earnings functions may include some variables which also explain participation. Alternative specifications of the earnings functions, in which earnings are determined not only by schooling and experience, but also by some of the explanatory variables in the participation equation, were also run to determine robustness. The results are presented in the appendix for three specifications involving various subsets of the explanatory variables in the ordered probit equation. The qualitative results reported in Table 4 are not affected. In particular, for the same values of schooling and experience in Table 4, the SENA effect calculated from uncorrected OLS regressions is considerably overestimated. The precise quantitative magnitude of the selectivity corrected earnings effect is somewhat sensitive to the specification. In another experiment, neither the quantitative nor the qualitative results change when a schooling variable is added to the ordered probit equation.

V. Final Remarks

The principal result of this study is that socioeconomic characteristics of trainees affect the decision on how much SENA type of training to undertake. Consequently, the estimation of an earnings effect from cross-section comparisons of SENA and non-SENA samples would be biased. An ordered probit decision model is used to correct for this bias in the earnings function. Initial results indicate that this bias may be positive and significant. When taken into account, at mean levels of schooling and experience, the impact of SENA long courses falls and that of short courses becomes negligible. The finding that the SENA effect is influenced by selection is robust for different specifications of the ordered probit and earnings functions.

A caveat in interpreting this result is that the data provide no direct measure of ability. If SENA systematically selects the less able for training, then there would be negative selection that could at least partially affect the positive selection due to socioeconomic background. In this case, the earnings differentials described above may be underestimates.

Appendix

Alternative Earnings and Probit Function Specifications

It is possible that some of the socioeconomic factors directly affect earnings, as well as the choice to attend SENA short and long courses. Appendix Table A1 presents reestimated earnings functions that include various subsets of the socioeconomic variables in the regression. The inclusion of these variables have relatively little effect on the schooling and experience variables. For example, the rates of return to education, calculated using Mincer's method, are about the same as those implied by Table 3 in the text. As might be expected, the magnitude and significance of LAMBDA's coefficients do vary, since LAMBDA is a nonlinear function of the socioeconomic variables. This is particularly true for the selection term for the non-SENA group in the second specification, as it virtually disappears. However, the signs of the coefficients do not change—the direction of the selection term is not affected in any of the specifications of Appendix Table A1.

The coefficients in Table A1 are used to compute earnings differentials corrected for selectivity. Most notably, the finding that the impact of SENA declines considerably with the addition of the correction term is preserved; corrected earnings effects are still significantly smaller than uncorrected effects, as shown in Table A2. For almost all the specifications, the SENA effect is negligible or negative. However, the magnitude of the difference between uncorrected and corrected earnings effects is somewhat sensitive to the specification. In Specifications 1 and 3, in which variables about household mobility are included in the earnings functions, the extent to which OLS overestimates the corrected earnings effect is magnified. The extent of the overestimation lessens, particularly for short-course graduates, for Specification 2, in which only variables that measure father's characteristics are included.

Another possible specification is that pretraining schooling—measured as years of schooling at the time the choice of which SENA course to attend[3]—can also affect the choice to attend SENA. The results are shown in Table A3. The left-hand panel shows the impact of pretraining initial schooling (labeled as $\bar{s}$) on the ordered probit equation. More years of initial schooling are negatively associated with the choice to attend more SENA training, given socioeconomic status. The right-hand panel shows the effect of deriving LAMBDA from an ordered probit equation that includes initial schooling. A comparison of these three columns with the results of the earnings functions in Table 3 of the text reveals that neither the intercept nor the coefficients change greatly. Consequently, the corrected earnings differentials in Column (4) of Table A4 are not much different from those shown in Column (4) of Table 4 in the text. The inclusion of pretraining schooling in the ordered probit does not alter the quantitative nor qualitative aspects of the results.

3. Non-SENA individuals are assumed to make their decisions about entry into SENA at the same age as the SENA counterparts.

Table A1
Alternative Specification of the Earnings Functions

	Specification 1			Specification 2			Specification 3		
Variables	Non-SENA	SENA Short	SENA Long	Non-SENA	SENA Short	SENA Long	Non-SENA	SENA Short	SENA Long
Intercept	10.6080	10.2170	9.9372	10.4000	9.3312	10.1380	11.3050	9.9725	9.5882
	(0.1597)	(0.3562)	(0.2416)	(0.1548)	(0.4632)	(0.2903)	(0.6681)	(0.3861)	(0.7706)
Schooling (s)	0.1231	0.1284	0.1359	0.1187	0.1299	0.1347	0.1194	0.1203	0.1351
	(0.0079)	(0.0296)	(0.0143)	(0.0079)	(0.0292)	(0.0143)	(0.0109)	(0.0299)	(0.0143)
Experience (x)	0.0855	0.0459	0.0986	0.0865	0.0477	0.1000	0.0864	0.0505	0.0999
	(0.0082)	(0.0217)	(0.0168)	(0.0081)	(0.0220)	(0.0170)	(0.0113)	(0.0219)	(0.0168)
x^2	−0.0011	−0.0007	−0.0016	−0.0011	−0.0007	−0.0016	−0.0011	−0.0007	−0.0016
	(0.0001)	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0003)
sx	−0.0025	0.0010	−0.0018	−0.0025	0.0008	−0.0021	−0.0025	0.0003	−0.0019
	(0.0005)	(0.0015)	(0.0011)	(0.0005)	(0.0015)	(0.0011)	(0.0007)	(0.0015)	(0.0011)
LAMBDA	0.3778	−0.3717	0.3914	0.0061	−0.2231	0.1206	1.1923	−1.7230	0.6762
	(0.1194)	(0.3616)	(0.1941)	(0.0950)	(0.2147)	(0.1310)	(0.7085)	(1.4650)	(0.8142)
FPRIM				0.0262	0.2260	0.0401	0.1973	−0.1210	0.1161
				(0.0707)	(0.1577)	(0.1157)	(0.1417)	(0.3445)	(0.1607)

FPRIMED				0.0132	0.4656	0.0615	0.1403	0.1498	0.1236
				(0.0661)	(0.1692)	(0.1112)	(0.1134)	(0.2658)	(0.1353)
FSECED				0.0816	0.7043	0.1157	0.0792	0.5746	0.1064
				(0.0682)	(0.1949)	(0.1161)	(0.0945)	(0.2012)	(0.1161)
FHIGHED				0.1359	0.5822	0.1807	−0.1424	1.0377	0.0183
				(0.0872)	(0.4032)	(0.1551)	(0.2402)	(0.8048)	(0.2988)
FNRED				−0.0386	0.2864	0.0112	−0.2407	0.5281	−0.1149
				(0.0951)	(0.4136)	(0.1852)	(0.1867)	(0.6116)	(0.2461)
FOCC				−0.0276	0.0636	−0.0036			
				(0.0126)	(0.0487)	(0.0195)			
MOBILB	0.0401	−0.2115	−0.1412				−0.0062	−0.0627	−0.1578
	(0.0497)	(0.1826)	(0.0789)				(0.0812)	(0.2293)	(0.0949)
MOBILBP	−0.0112	0.1118	0.1316				−0.0022	0.0671	0.1397
	(0.0412)	(0.1395)	(0.0679)				(0.0579)	(0.1414)	(0.0685)
MOBILP	−0.1103	0.2020	0.0050				−0.2267	0.4990	−0.0364
	(0.0484)	(0.1909)	(0.0802)				(0.1195)	(0.3528)	(0.1393)
RURALED	0.0061	−0.4019	0.1169				0.0437	−0.4093	0.1345
	(0.0462)	(0.1317)	(0.0728)				(0.0710)	(0.1572)	(0.0800)
INTERRUP	0.1243	−0.0695	0.0706				0.4166	−0.7620	0.1662
	(0.0603)	(0.2164)	(0.0799)				(0.2611)	(0.7644)	(0.2709)
R squared	0.3735	0.3087	0.1977	0.3782	0.3130	0.1946	0.3798	0.3289	0.1994

Table A2
Uncorrected and Selectivity-Corrected Effects of SENA on Earnings: Alternative Specifications[a]

Course Type	(1) Years of Schooling	(2) Years of Experience	Earnings Effect of SENA (%) (3) Uncorrected OLS	(4) Selectivity Corrected
Specification 1				
Short	10	19	−4.3	−20.2
Short	10	29	8.8	−8.1
Short	12	19	8.5	−7.9
Short	12	24	18.9	1.3
Long	10	12	26.8	−36.3
Long	10	20	27.0	−35.2
Long	12	12	31.6	−33.6
Long	12	16	33.8	−32.0
Specification 2				
Short	10	19	−19.5	−23.4
Short	10	29	−7.3	−12.3
Short	12	19	−5.6	−11.5
Short	12	24	3.7	−2.8
Long	10	12	30.1	18.5
Long	10	20	30.9	20.0
Long	12	12	35.2	23.6
Long	12	16	37.1	25.7
Specification 3				
Short	10	19	−10.6	−62.4
Short	10	29	2.9	−57.9
Short	12	19	0.0	−58.1
Short	12	24	9.1	−54.9
Long	10	12	27.6	−76.3
Long	10	20	31.4	−75.6
Long	12	12	33.4	−75.2
Long	12	16	37.2	−74.5

a. Values of socioeconomic characteristics: father's education set at incomplete primary or less; occupational status set at the middle of the scale; and sample individuals belong to households who moved since individual's primary school.

Table A3
Corrected Earnings Functions Where LAMBDA is Estimated from an Ordered Probit Equation that Includes a Schooling Variable

Ordered Probit		Corrected Earnings Functions			
Variable	Coefficient	Variable	Non-SENA	SENA Short-Course Graduates	SENA Long-Course Graduate
Intercept	.2010	Intercept	10.3480	9.8733	10.0980
	(.1894)		(.1360)	(.3214)	(.2018)
$\bar{s}$	−.0375	s	0.1252	0.1498	0.1363
	(.0094)		(.0079)	(.0282)	(.0141)
FPRIM	.2753	x	0.0844	0.0419	0.0989
	(.1228)		(.0081)	(.0218)	(.0169)
FPRIMED	.2694	x squared	−0.0010	−0.0007	−0.0016
	(.1176)		(.0001)	(.0003)	(.0003)
FSECED	.1358	sx	−0.0025	0.0017	−0.0020
	(.1232)		(.0005)	(.0015)	(.0011)
FHIGHED	−.2103	LAMBDA	0.1070	−0.3509	0.2230
	(.1578)		(.0762)	(.1850)	(.1048)
FNRED	−.1807				
	(.1804)				
FOCC	.0281				
	(.0214)				
MOBILB	−.0977				
	(.0839)				
MOBILBP	.0139				
	(.0720)				
MOBILP	−.2109				
	(.0791)				
RURALED	.0055				
	(.0785)				
INTERRUP	.5483				
	(.0610)				
LOG-LIKL	−2,246.7	R^2	.37	.28	.19

Table A4
Earnings Effects from Corrected Earnings Functions in Table A3

Course Type	(1) Years of Schooling	(2) Years of Experience	Earnings Effect of SENA (%)[a] (3) Uncorrected OLS	(4) Selectivity Corrected
Short	10	19	.9	−12.5
Short	10	29	18.4	0.7
Short	12	19	24.3	7.7
Short	12	24	39.9	19.5
Long	10	12	33.7	0.5
Long	10	20	35.2	0.5
Long	12	12	37.9	4.0
Long	12	16	40.5	5.4

a. Percentage differential from the non-SENA control group, at given levels of schooling and experience.

References

Garen, J. 1984. "The Returns to Schooling: A Selectivity Bias Approach With a Continuous Choice Variable." *Econometrica* 52(5):1199–2218.

Gomez, H., and Libreros E. 1984. "Formacion Profesional y Mercados de Trabajo." In *Educacion Formacion Profesional y Empleo*, 251–68. Bogota: SENA.

Heckman, J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5:475–92.

———. 1979. "Sample Selection Bias as A Specification Error." *Econometrica* 47:153–61.

Heckman, J., and R. Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions: An Overview." *Journal of Econometrics*, 30(1/2):239–67.

Jimenez, E., and B. Kugler. 1985. "Does SENA Matter—Some Preliminary Results on the Impact of Colombia's National Training System on Earnings." The World Bank, Education and Training Department, Discussion Paper.

Kenny, L., L. F. Lee, G. S. Maddala, and R. P. Trost. 1979. "Returns to College Education: An Investigation of Self-Selection Bias and the Project Talent Data." *International Economic Review*. 20(3):775–89.

Kugler, B., and A. Reyes. 1978. "Financing of Technical and Vocational Training in Latin America." In *The Financing of Education in Latin America*, 296–320. Washington: Inter American Development Bank.

Lee, C. B. 1985. "Financing Technical Education in LDCs: Economic Implication From a Survey of Training Modes in the Republic of Korea." EDT Discussion Paper, The World Bank.

Lee, L. F. 1978. "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables." *International Economic Review* 19(2):415–33.

———. 1983. "Generalized Econometric Models With Selectivity." *Econometrica* 51(2):507–12.

Lee, L. F., and R. Trost. 1978. "Estimation of Some Limited Dependent Variable Models With Application to Housing Demand." *Journal of Econometrics* 8:357–82.

Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

Metcalf, D. 1985. *The Economics of Vocational Training: Past Trends and Future Considerations*. World Bank Staff Working Paper No. 713.

Puryear, J. 1977. *Estudio Comparativo de la Formacion Professional en Colombia*. Montevideo: El Servicio Nacional de Aprendizaje.

SENA-Holanda. 1982. *Sistema de Planeacion de Recursos Humano*. Bogota: SENA.

Trost, R., and L. F. Lee. 1984. "Technical Training and Earnings: A Polychotomous Choice Model With Selectivity," *Review of Economics and Statistics* 66(1):151–56.

Willis, R., and S. Rosen. 1979. "Education and Self-Selection." *Journal of Political Economy* 87(5, Part 2):S7–S36.