MODELING EDUCATION SYSTEM PERFORMANCE WITH DEMOGRAPHIC DATA

AN INTRODUCTION TO THE PROFLUXO MODEL

by

Philip R. Fletcher
ILO/UNDP-IPEA/IPLAN

Sergio Costa Ribeiro
LNCC/CNPq

Brasília, DF
March 8, 1989

MODELING EDUCATION SYSTEM PERFORMANCE WITH DEMOGRAPHIC DATA

AN INTRODUCTION TO THE PROFLUXO MODEL

Introduction

Official statistics produced by national ministries of education play important roles in education planning and evaluation, where they are widely used in predicting future enrollments and monitoring education system performance.

Today, it is recognized that serious errors are present in the school statistics published by ministries of education in developing countries. By systematically underestimating grade repetition and overestimating dropout, these statistics provide a distorted image of basic education processes. Obviously, this fact seriously compromises the potential usefulness of planning and evaluation studies based on official education statistics in the developing countries.

After briefly reviewing these issues, we will present a mathematical model known as PROFLUXO, which provides an alternative procedure for analyzing student flows in a graded education system. Based upon demographic data, this methodology circumvents use of official statistics, producing independent estimates of grade transition rates and other indices of education system performance.

In Brazil, Teixeira de Freitas pioneered studies in this area after noting that official school statistics were incompatible with demographic data.[1] During the 1940's, he observed that the number of allegedly "new" students reported in school statistics greatly surpassed the number of individuals in an age cohort entering

[1] See the series of articles published by M.A. Teixeira de Freitas in Revista Brasileira de Estatística (Rio de Janeiro), beginning with "Dispersão Demográfica e Escolaridade", RBE, v.1, n.3 (1940), p. 497-527, and concluding with "A Escolaridade Média no Ensino Primário Brasileiro", RBE, v.8, n.30/31 (1947), p. 295-474.

school.[2] Since the school system is relatively stable from one year to the next, the number of really new students could not possibly exceed the number of individuals in an age cohort entering school. To do so over a period of years would imply that school intake consistently surpasses demographic possibilities.

Obviously, official school statistics overestimated the number of really new students entering the Brazilian education system during this period. Since the sum of new students, repeaters and dropouts must necessarily equal enrollments, it was also obvious that official data underestimated the number of repeaters or dropouts. Teixeira de Freitas concluded, in the course of a series of analyses, that the number of repeaters must be greatly underestimated and proceeded to develop a method for correcting errors in official data.[3]

In 1975, Schiefelbein noted that these same problems appear in the official statistics of virtually all of the Latin American countries, with the possible exception of Uruguay. He, too, shows that repetition is consistently underestimated in each of these cases and presents certain methods to correct official data.[4]

Thonstad, in an UNESCO manual of methodology for projecting school enrollments in developing countries, acknowledges that there are possibly large margins of error in official data on new entrants, which could be due to the under-reporting of repetition in grade one. He reports "seemingly strange intake rates," inconsistent with demographic possibilities, for four Latin American countries, three countries in Asia and two African countries.[5]

These analyses suggest that problems with official data are essentially methodological in nature, related to the source of the information, that is, the reporting provided by individual schools.

---

[2] An age cohort consists of all individuals born in the same year. Ignoring mortality, an age cohort is composed of all individuals of a given age.

[3] See "Retificação necessária e método empregado" in M.A. Teixeira de Freitas, "A Escolaridade Média no Ensino Primário Brasileiro", RBE, v.8, n.30/31 (1947), p. 295-474.

[4] Ernesto Schiefelbein, "Repeating: An Overlooked Problem of Latin American Education," Comparative Education Review, v.19, n.3 (1975), p. 468-487.

[5] Tore Thonstad, Analysing and Projecting School Enrolment in Developing Countries: A Manual of Methodology, Statistical Reports and Studies, n.24 ([Paris]: UNESCO, [1980]), p. 34, esp. Table 3.1, "Intake Rates to Primary Education for Selected Countries. Boys."

Systematic distortions in reporting may be due to any of the following reasons:

- Repetition may be considered a social stigma. When a teacher asks a student, who has just arrived at a new school, if he is repeating the grade, there is a natural tendency to say "No", especially in the first grade, where evidence of prior enrollment is not required. Some children, who are reported as "new" students in a given grade, actually failed the grade, providing a motive for transferring from one school to another.

- Students who have already passed a given grade, upon transferring to another school, may be forced to repeat the same grade at the new school. They are not included in the total of repeating students at the new school since they passed the grade the previous year.

- The true number of repeating students may be underestimated because a high rate of repetition produces a negative image of the teacher or school. Repetition may be intentionally under-reported to avoid external evaluations or administrative interventions.

- Enrollment reporting may be intentionally inflated to justify the number of teachers or the very existence of the school. At the end of the school year, the artificial excess in enrollments is eliminated by over-reporting dropouts.

In any event, it should be noted that the sum of dropouts reported by individual schools does not necessarily represent the number of students leaving the school system in a given year. Students who apparently drop out of one school may subsequently enroll at another. The limited vision of individual schools is thus inevitably reflected in school statistics.

Collectively, these problems seriously compromise the usefulness of school statistics, producing a distorted image of how the school system functions as a whole. Certain methods have been developed in attempting to correct these errors.[6] However, the success of these methods depends upon accepting at least one of the grade transition rates reported in official statistics as a starting point for correcting the other rates, a procedure which is likely to be risky at best.

---

[6] See Ernesto Schiefelbein, op. cit., and especially ibid., "Statistical Report on Repetition in Latin America" (mimeo), report prepared for the UNESCO Division of Statistics on Education, Office of Statistics (Paris: UNESCO, 1980).

Studying these problems over the last four years, the authors of this report developed PROFLUXO, a mathematical model of student flows through a graded education system, based on a methodology developed initially by Fletcher.[7] PROFLUXO belongs to a class of formal mathematical models known as flow models, where enrollments in a given year are based upon enrollments in the grade or level below in the previous year and upon coefficients describing student flows between grades and levels from one year to the next.

PROFLUXO is used to develop a grade transition matrix, containing coefficients describing age-specific intake, promotion, repetition, and dropout in each grade. In this sense, PROFLUXO follows in the tradition of a series of mathematical models used in analyzing and projecting school enrollments, widely disseminated by UNESCO.

The originality of PROFLUXO stems from the fact that demographic data of the type commonly found in national censuses are used as a basis for calculating all grade transition rates. This offers several advantages over conventional methodologies based upon school statistics.

First, by dispensing use of data supplied by ministries of education, PROFLUXO circumvents the widely acknowledged deficiencies commonly found in official school statistics: in particular, the systematic under-reporting of repetition and the over-reporting of dropout, but also the fact that school statistics themselves do not provide information on the school-age population out of school.

PROFLUXO estimates repetition, promotion and dropout from demographic data based on statistical inferences, using questionnaire items that in no way compromise the self-esteem or interests of informants. PROFLUXO is not based on the declarations of students, teachers or school principals.

---

[7] Philip R. Fletcher, A Mathematical Model of School Trajectory, Repetition and the Performance of First Level Schooling in Brazil (Brasília, DF: CNRH, 1985). See also, ibid., "A Repetência no Ensino de 1° Grau: um Problema Negligenciado da Educação Brasileira," Revista Brasileira de Administração da Educação (Porto Alegre), v.3, n.1 (jan./jun. 1985), p. 10-41. The authors would like to express their appreciation for the support and encouragement received from Cláudio de Moura Castro while developing the first PROFLUXO prototype.

Second, PROFLUXO assures that school intake rates are consistent with demographic possibilities. This is easily obtained since only demographic data are used in generating the model.

Third, as a general mathematical model, PROFLUXO can be applied to a great variety of sub-groups in the population or to the entire population of a country. In particular, PROFLUXO can be applied to any population group identified by a (relatively) permanent characteristic, including sex, region, urban/rural situation, township of residence, socio-economic group or even parents' occupational category.[8]

This is an important advantage over methodologies based upon school statistics, which do not report income level and consequently do not permit analyses of education system performance by income group. PROFLUXO permits analysis of the equity of education system opportunities offered different groups in the population.

Fourth, PROFLUXO is, at least in principle, easily transportable. Based upon age and grade declarations, together with optional classificatory variables (sex, region, income group, etc.), the model can be applied to virtually any census or large scale sample survey in any country in any year.[9] Thus, PROFLUXO should be of potential interest to education researchers and planners in a great variety of countries, where the model could also be used to analyze longitudinal trends in education system performance.

Fifth, it should be recognized that PROFLUXO shares a number of important advantages common to all formal mathematical models:

The planning and administration of education systems requires handling large quantities of information interrelated in a very complex fashion. PROFLUXO enables the systematic, consistent and

---

[8] To do so, it must be reasonable to assume that "migration" from one category to another is either limited or inconsequential from an educational point of view. It should be noted that school systems under different types of jurisdiction (private, local, state or federal) cannot be analyzed with PROFLUXO because students often attend different kinds of schools over a period of years and therefore "migrate" from one jurisdictional type of school to another.

[9] Economical in terms of the basic variables used by the model (namely, age and grade), PROFLUXO requires a large number of observations in each population sub-group analyzed. Experience shows that 500 observations between 5 and 29 years of age are essential in each unit of analysis but that 1000 observations in each unit are more satisfactory.

rapid analysis of this information, especially when implemented on a computer.

PROFLUXO permits detailed quantitative analyses that provide an understanding of formal relationships existing within the education system and sheds light on the ways this system interacts with its social environment. This methodology, through specification of logical relationships among different components of the system, reveals errors and inconsistencies in statistical data and erroneous interpretations that otherwise would be extremely difficult to detect.

PROFLUXO, together with demographic projections and cost data, can be used to predict future demand for basic inputs (financial resources, teachers, schools, textbooks, etc.) required to meet the educational needs of the population. As a general mathematical model, PROFLUXO permits simulation of the potential effects of changes introduced in the education system by altering basic parameters of the grade transition matrix. Thus, one could study the probable social consequences, over the medium and long run, of the implementation of any of a variety of education policies bearing on promotion and retention practices.

Finally, we hope that PROFLUXO will stimulate further education research and indicate needs for systematic collection of additional information. PROFLUXO reveals important quantitative aspects of the education system but is less useful in analyzing its qualitative features due to the common absence of cognitive data in national censuses and most surveys.[10] If such information were available, application of the basic principals of PROFLUXO would be extremely useful in analyzing the rate and extent of learning in the population.

The following pages of this report describe PROFLUXO, first as a basic mathematical model of student flows through a stable education system. The basic model is then generalized to incorporate change in the education system over time and to enable analysis of different subgroups in the population. Throughout, specific applications of the model to 1982 Brazilian National

---

[10] However, most censuses include information on literacy, a basic cognitive outcome of schooling. The basic principles of PROFLUXO could be applied to this data in order to calculate the average age at which literacy is attained in the school system. This age could then be compared with the average age of first grade promotion, in order to determine what de facto promotion criteria are used in the first grade (promotion criteria short of basic literacy, equal to literacy or in excess of literacy). See Philip R. Fletcher and Cláudio de Moura Castro, <u>Os Alunos e as Escolas no Brasil de Hoje</u> (Brasília, DF: [IPLAN/IPEA], 1986), esp. p. 51 and ff.

Household Sample Survey (PNAD-1982) data are used to illustrate certain features of the model.

Software for generating the basic parameters of PROFLUXO is currently based on a series of SAS and PL/1 programs operating on IBM mainframe computers. This software is not supplied with this report. We hope that a unified software package will eventually become available so that PROFLUXO parameters can be readily generated for any census or large scale sample survey representing any country in any year. These parameters can then be used to generate any of a large variety of indicators describing education system performance.

The Basic PROFLUXO Model

For individuals five years of age or older, household surveys and censuses normally report the level and grade of schooling attended or the last level and grade concluded at the time of investigation. On this basis, it is possible to identify individuals in any of three general circumstances:

n_i := individuals of age i who have never gone to school;

m_{ki} := individuals of age i enrolled in grade k; and

d_{ki} := individuals of age i having left school after passing grade k;[11]

where, for example, in the Brazilian case,

i : 5 .. 29; and

k : 1 .. 10.

When PROFLUXO is applied to Brazilian first level schooling, i can be restricted to ages 5 through 29 years, since virtually everyone will have concluded first level schooling by 29 years of age.

Again, in the Brazilian case, to represent the eight grades of first level schooling, k can take on values between 1 and 8; k=9 is reserved for the first grade of the next level of schooling; and k=10 is used for all grades and levels beyond this point.

Summing expansion factors for each observation found in the sample, it is possible to estimate the population in any of these three general circumstances. N then becomes a vector representing all who have never gone to school, where each cell refers to a different age. M is an enrollment matrix and D is a matrix of those who have completed their schooling; the cells of each of these matrices represent different age-grade combinations.

If the population is growing, then there will generally be more individuals of age 5 than of age 29. Since population growth is exogenous to the model, this unwanted factor is eliminated through normalization, dividing the total in each cell of age i by the total population of age i.

[11] The expressions "passing a grade" and "graduating from a grade" will be used interchangeably in this report to denote those who have successfully completed a grade.

After normalization, each cell in the vector N or matrices M and D will represent the proportion of an age group found in each situation. Thus, N, M, and D share the same metric, where units are expressed as proportions of an age group.

This normalization is necessary for a basic hypothesis underlying the model. This hypothesis assumes that, in a formal education system, any individual who has reached grade k has attended and passed all of the previous grades.[12]

Thus, we can write the following Equations for each of the k=1 through k=9 grades:

$$A_{ki} = \sum_{j=k}^{10} D_{ji} + \sum_{j=k+1}^{10} M_{ji} \tag{1}$$

$$I_{ki} = M_{ki} + A_{ki} \tag{2}$$

and thus,

$$A_{ki} = D_{ki} + \sum_{j=k+1}^{10} I_{ji} \tag{3}$$

where,

A_{ki} := the proportion of individuals of age i having passed grade k; and

I_{ki} := the proportion of individuals of age i having already enrolled in grade k.[13]

---

[12] Grade "skipping" is ignored, which is an acceptable simplification, since its incidence is usually low.

[13] There is no need to make explicit use of vector N in this formulation since N_i is implied in $1 - I_{1i}$, i.e., those age i who have never entered first grade.

It should be noted that the proportions A_{ki} are reliable since individuals should pass a grade only once. On the other hand, the proportions I_{ki} tend to underestimate those of age i who have entered grade k. Since $I_{ki} = M_{ki} + A_{ki}$, any person having entered without passing grade k, and who is not currently enrolled in school, is excluded from I_{ki}.[14] In any event, it should be noted that the proportions entering grade k by age i can never be <u>less</u> than I_{ki}.

With this exception, I and A are accumulative, representing all individuals of age i having respectively entered or passed grade k.

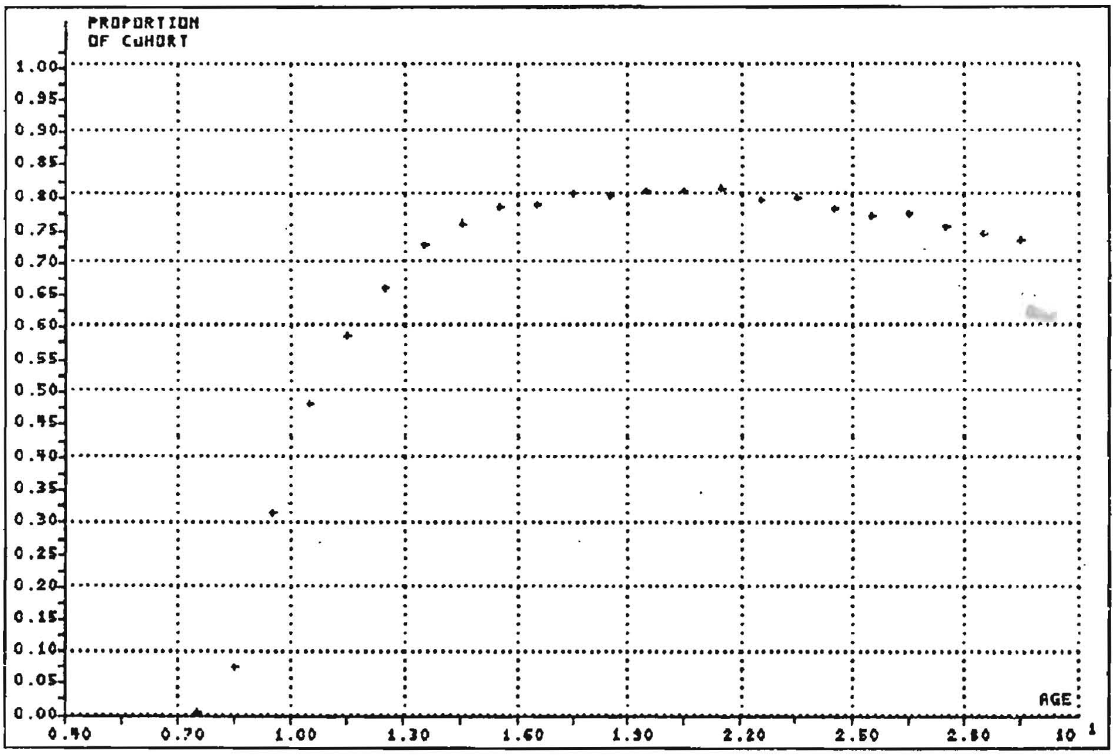

Figure 1

Population Entering Grade k in Relation to Age i:
Plot of Original Data I_{ki}

---

[14] Surveys can avoid this by asking people if they attempted any schooling beyond the last grade successfully completed.

For any grade k, the proportions in I and A can be plotted in relation to age i. Figure 1 shows the original data points I_{ki} for the Brazilian case, representing the proportions in each age group i having entered grade k.[15]

In Figure 1, the sequence of points I_{ki} rise to a maximum, following an imaginary curve that is roughly sigmoid in shape.

The curve rising to a maximum suggests that the situation of the population of age i with respect to grade k in the following years i+1, i+2, i+3... can be approximated by examining the proportions I_{ki+1}, I_{ki+2}, I_{ki+3}....

Where education systems are stable, the grade distribution by age of a population is that of any of the different cohorts constituting the population. In this case, a sequence of age groups can be used to represent a cohort currently i years of age. The reconstructed age cohort, derived from crossection data, will represent a true age cohort based on longitudinal data when conditions are stable.[16]

By using age as a proxy for time, an acceptable approximation to the longitudinal flow of an age cohort entering grade k can be obtained. In this case, the sequence of points I_{ki}, rising to a maximum, represents <u>the flow of new students entering grade k</u>. Moreover, Figure 1 suggests that there are, indeed, precisely defined age patterns of grade entrance in Brazil.

The current grade intake rate, the proportion of an age cohort that presently enters grade k, is approximated by the maximum of $I_{k.}$.[17] This proportion represents new students of all ages entering grade k in a given year.

Following the maximum, the proportions I_{ki} decline slightly in value. This decline can be attributed to increasing intake

---

[15] Figure 1 is based on data obtained for the third grade of Brazilian first level schooling in 1982. Age is represented in the Figure as a continuous variable, whereas in the original data, age is a discrete variable representing an interval between i' and i' plus twelve months. To represent the midpoint of this interval, Figure 1 shows data for age i, where i = i' + 0.5.

[16] The assumption of education system stability over time, where grade intake rates are constant and unchanging, is introduced in order to facilitate presentation of the basic PROFLUXO model. In the next section of this report, the basic model will be generalized to explicitly incorporate education change over time.

[17] In the Brazilian case, this maximum usually occurs somewhere between 18 and 20 years of age.

rates over successive years as the education system extends its coverage of the population.

The points in Figure 1 that appear after the maximum is reached represent the grade participation rates of older age cohorts.[18] Thus, in the very strictest sense, the assumption of a "stable education system" in the Brazilian case is violated since participation rates are obviously rising over time (i.e., decline with successively older age cohorts). Each age cohort represented in Figure 1 is following a slightly different trajectory and will ultimately rise to a different grade participation rate.

The lower participation rates of older age groups imply that the proportions shown in Figure 1 cannot be interpreted as a longitudinal function of time. Obviously, an age cohort cannot lose education that has already been attained. In order to deal with this difficulty, a function with a horizontal asymptote is fitted to the original data rising to the maximum.

For any grade **k**, this function is:

$$P_k(i) \quad = \quad 2A \;/\; (\;1 + e^{ai-b}\;) \tag{4}$$

where,

$P_k(i)$:= the proportion of an age cohort entering grade k by age i;

A, a and b := parameters of the function $P_k(i)$ fitted to the data of I_{ki} up to the observed maximum;[19] and

i := age in years.

Figure 2 shows the function $P_k(i)$ fitted to the data of I_{ki} first presented in Figure 1.

---

[18] The grade participation rate represents the proportion of a given age cohort sooner or later entering grade k. There is a subtle difference between grade intake rates and the participation rates of age cohorts when education systems are changing over time. This distinction will be elaborated in the next section of this report. In this section, it is assumed that education system performance is constant over time; in this special case, grade intake and participation rates are identical.

[19] This function was proposed by Sergio Costa Ribeiro in 1986 and was subsequently fitted to the original data using an algorithm developed for the project by João José Farias Neto.

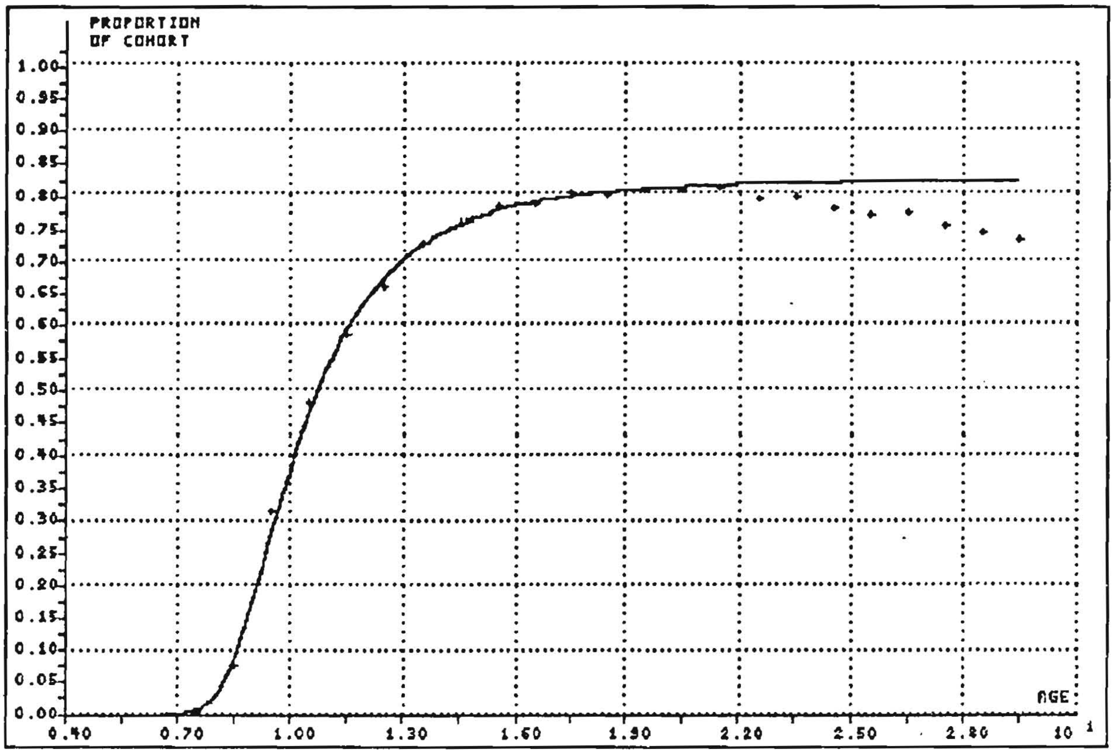

Figure 2

Population Entering Grade k in Relation to Age: The Function $P_k(i)$ Fitted to the Data of I_{ki}

The advantage of using an analytical function to represent entering students is not only to provide the best possible representation of the originally observed data but also to permit calculation of a series of indicators of education system performance.

In particular, the <u>first derivative</u> of $P_k(i)$, shown in Figure 3, represents the age frequency distribution of students entering grade k. This distribution can be used to calculate the mean, median and modal age of students entering the grade.[20]

This discussion can be repeated substituting A_{ki} for I_{ki}. In this case, A_{ki} represents the proportion of an age group having

---

[20] The first derivative of the function $P_k(i)$, presented in Equation (4), is calculated as:

$$P_k'(i) = \frac{2Aab\,i^{-(b+1)}e^{ai^{-b}}}{e^{ai^{-b}}} . \qquad (5)$$

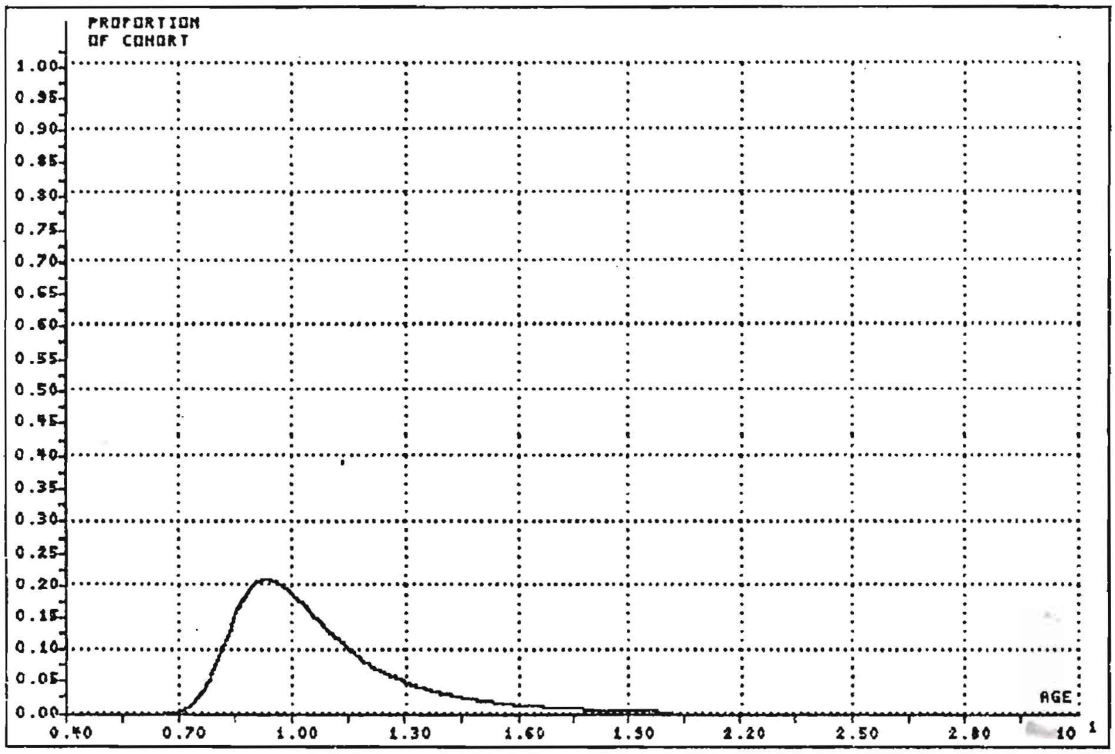

Figure 3

Frequency of Students Entering Grade k in Relation to Age:
First Derivative of the Function $P_k(i)$ Fitted to the Data of I_{ki}

passed grade k.

Once again, the sequence of points A_{ki} rise to a maximum, following a curve that is essentially sigmoid in shape.

Using age as a proxy for time, under stable conditions, an acceptable approximation to the longitudinal flow of an age cohort passing grade k can be obtained. In this case, the sequence of points A_{ki}, rising to a maximum, represents the flow of students passing grade k. Once again, there are, indeed, precisely defined age patterns of grade completion in Brazil.

The current graduation rate, the proportion of an age cohort passing grade k in a given year, is approximated by the maximum of $A_{k.}$.

Following the maximum, the proportions A_{ki} decline slightly. This decline is attributed to increasing graduation rates over successive years as the education system extends its coverage of the population.

The points in A_{ki} that appear after the maximum is reached represent the proportion in each age cohort completing the grade in previous years.

The lower graduation rates of older age groups imply that the proportions found in A_{ki} cannot be interpreted as a longitudinal function of time. In order to deal with this difficulty, a function with a horizontal asymptote is fitted to the original data rising to the maximum.

Finally, the same function $P_k(i)$ defined in (4) is fitted to the data of A_{ki} up to the maximum in order to represent graduation as a function of time.

In this sense, there are really two functions P for each grade. Functions P_I represent an age cohort entering specific grades and P_A represent graduation.

Once again, the first derivative of P_A represents the age frequency distribution of students passing grade k. This distribution can be used to calculate the mean, median and modal age of students graduating from grade k.

The standard deviation of the difference between the predicted values of P and the originally observed values of I or A is approximately 0.01 of an age cohort in more than 500 cases analyzed with Brazilian data, suggesting that the general function P is well suited to represent distributions of either I or A up to the observed maximums.

The the maximum of $I_{k.}$ is of considerable theoretical interest since it represents the number of new students of all ages entering grade k in a given year. One alternative is to use the value of A as an acceptable approximation, since this is the assymptote of Equation (4). However, if there are optimistic sampling errors around the maximum of $I_{k.}$, the parameter A will tend to float upwards, making it less reliable as an estimate of the maximum.

A more conservative, and generally safer, procedure is to fit a curve to the points beyond the maximum of $I_{k.}$ and find the point where this new curve intersects the function $P_k(i)$. The point of intersection should be a good estimate of the maximum since it is based on a larger number of data points.

When dealing with proportions, such as that represented by the maximum, it is advisable to use a function such as the logistic, which has the desirable property of being limited to the extremes of 0 and 1, respectively. Proportions beyond the maximum of $I_{k.}$ can thus be readily represented by the logistic function

$$L_k(i) = 1 \;/\; (\; 1 + e^{c+di} \;) \tag{6}$$

where,

$L_k(i)$	:=	the proportion having entered grade k of the age group currently of age i;
c and d	:=	parameters of the function $L_k(i)$ fitted to the data of I_{ki} beginning with the observed maximum; and
i	:=	age in years.

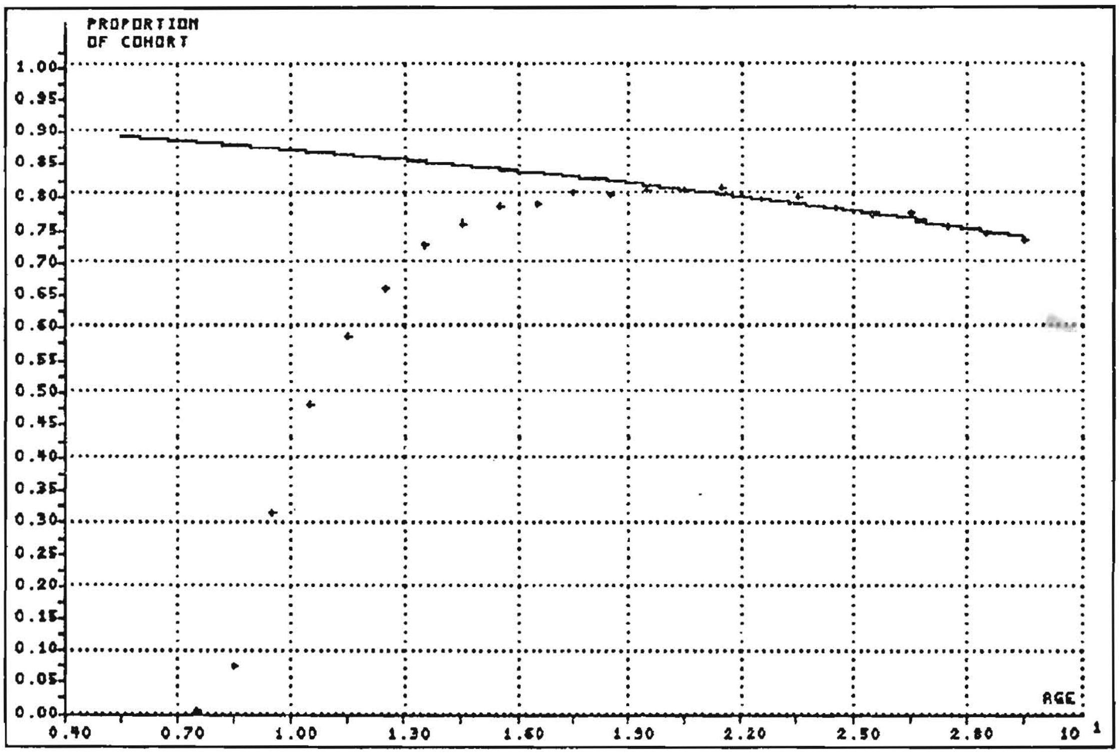

Figure 4

Trends in Participation Rates for Grade k in Relation to Age: The Function $L_k(i)$ Fitted to the Data of I_{ki}

Figure 4 shows the function $L_k(i)$ fitted to the data of I_{ki}, beginning with the observed maximum, as first presented in Figure 1.[21]

Using age as a proxy for time, a rough approximation to the change in grade k participation rates over time is obtained. In this case, the sequence of points I_{ki}, following the observed maximum, represents the declining coverage of grade k as we go backwards in time.

The intersection of $P_k(i)$ and $L_k(i)$ at the maximum is used as an estimate of the current intake rate in grade k, as illustrated in Figure 5.

---

[21] Since the observed maximum is usually found between 18 and 20 years of age in the Brazilian case, the function L has been fitted to at least 10 of the original data points, through age 29.

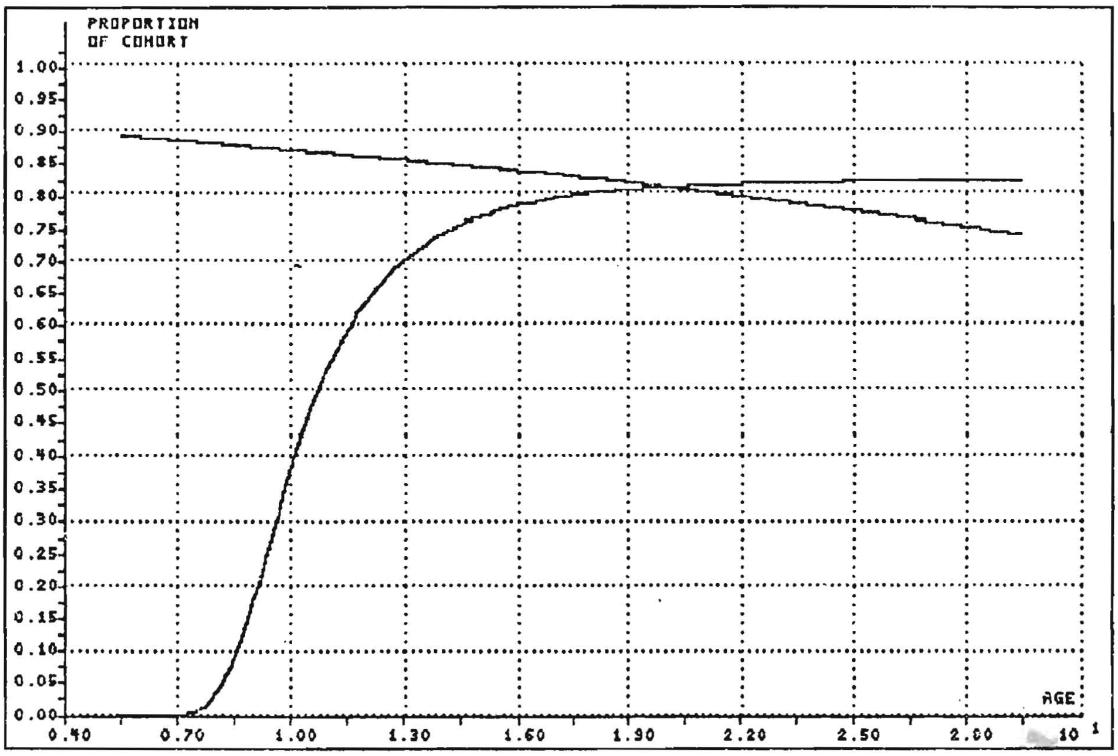

Figure 5

The Current Intake Rate in Grade k, Represented by the Intersection of Functions $P_k(i)$ and $L_k(i)$ Fitted to the Data of I_{ki}

The same function $L_k(i)$ defined in (5) can be fitted to the data of A_{ki}, beginning with the observed maximum, in order to represent graduation rates in grade k as a function of time.

In this sense, there are really two functions L for each grade. Functions L_I represent participation rates and L_A represent graduation rates.

The standard deviation of the difference between the predicted values of L and the originally observed values of I or A is approximately 0.01 of an age cohort in more than 500 cases analyzed with Brazilian data, suggesting that the general function L is well suited to represent distributions of either I or A after the observed maximums.[22]

---

[22] Functions L_A for the different grades are not sufficiently stable to permit projecting graduation rates into the distant future. Occasionally, L_{k+1t+n} will exceed L_{kt+n}, implying that more students graduate from grade k+1 than from grade k. In other

The intersection of P_A and L_A is used as an estimate of the current graduation rate.

Functions P and L can be used to represent virtually all of the originally observed data within a tolerable margin of error. Moreover, all of the elements of the grade transition matrix can be derived from the same functions.

Figure 6 helps visualize this process.

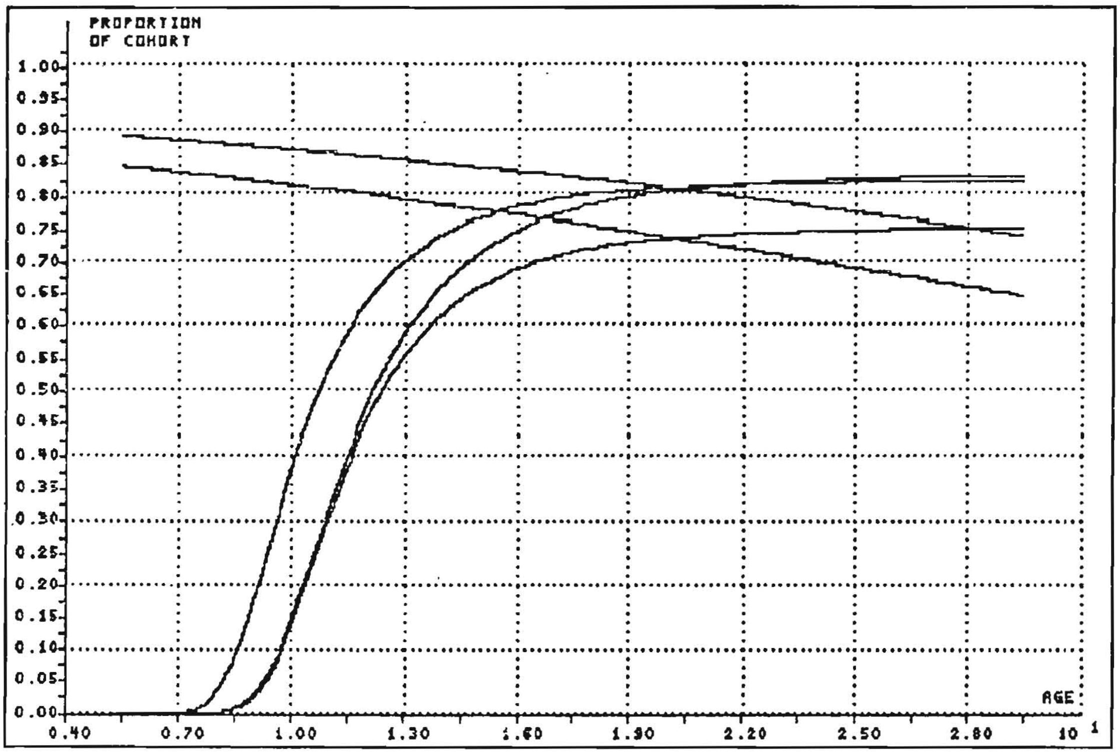

Figure 6

Using Functions P_I, L_I and P_A to Calculate Components of the Grade Transition Matrix

From Figure 6, we have:

I_k := New students of all ages entering grade k, represented by the proportion found at the

---

words, the lines represented by functions L_{k+1} and L_k occasionally will cross at some point in the distant future, represented by t+n. Current versions of PROFLUXO make no attempt to project grade participation and graduation rates into the future.

intersection of $P_k(i)$ and $L_k(i)$ for entering students;[23]

M_k := Enrollments in grade k, represented by the area between the two curves $P_k(i)$ for entering students and $P_k(i)$ for graduates;[24]

R_k := Repeating students enrolled in grade k, represented by the difference $M_k - I_k$;[25] and

D_k := Dropouts leaving school each year after having entered grade k, represented by the difference $I_k - I_{k+1}$.

Each of these components is expressed in the same metric, where units represent proportions of an age cohort.

Student flows in a stable education system comprised of eight grades can then be represented by a grade transition matrix, such as that presented in Table I.[26]

Under stable conditions, the grade transition matrix row totals

---

[23] Alternatively, the proportion in each age cohort currently entering grade k. For k=1, I_k represents new students entering grade one, i.e., the rate of access to schooling. For k=9, I_k represents students enrolling in the first grade of second level schooling.

[24] Alternatively, the average number of years of instruction currently received by an age cohort in grade k or the equivalent number of cohorts represented among total enrollments in grade k. M_T represents total enrollments in first level schooling or, alternatively, <u>the total number of years of first level schooling received by an age cohort</u>. Over-age enrollments can be excluded by ignoring the area between the two curves beyond the intersection of $P_k(i)$ and $L_k(i)$ for graduates.

[25] Alternatively, the average number of years that an age cohort spends repeating grade k.

[26] In a "stable" education system, from one year to the next, grade participation rates are assumed to be constant and enrollment growth is assumed to be proportional to population growth. Suitable refinements can be introduced in the model if this is not the case, however, their effect on grade transition rates is normally small since enrollment growth tends to accompany changes in participation rates. Current versions of PROFLUXO ignore such contingencies.

$$M_k = R_k + I_{k+1} + D_k \tag{7}$$

are equal to the column totals

$$M_k = R_k + I_k \,. \tag{8}$$

Table I

A Grade Transition Matrix for the Eight Grades of Brazilian First Level Schooling, Showing Components

Grade Yr t	1	2	3	4	5	6	7	8	P	D	T
1	R_1	I_2								D_1	M_1
2		R_2	I_3							D_2	M_2
3			R_3	I_4						D_3	M_3
4				R_4	I_5					D_4	M_4
5					R_5	I_6				D_5	M_5
6						R_6	I_7			D_6	M_6
7							R_7	I_8		D_7	M_7
8								R_8	I_9	D_8	M_8
N	I_1										
T	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8			M_T

The grade transition matrix presented in Table I reveals at a glance what proportion of a age cohort currently enters each grade, what proportion abandons school at each grade level, how many age cohorts are represented among grade enrollments, and what fraction of an age cohort is currently repeating any given grade.

This information is useful in describing how an age cohort currently obtains its formal education. Thus, when expressed as proportions of an age cohort, we obtain a formal accounting of enrollment flows in each grade of the education system:

$$M_{kt+1} = M_{kt} + I_{kt+1} - I_{k+1t+1} - D_{kt+1} \qquad (9)$$

PROFLUXO assures that this accounting is consistent with demographic realities.

Grade transition rates, expressed as proportions of grade enrollments, can be easily calculated from the matrix presented in Table I, dividing each matrix element by the corresponding row total M_k. The resulting promotion, repetition and dropout rates are useful in describing how enrollments are expected to behave from one year to the next.

Using grade transition rates, a formal accounting of enrollment flows in each grade of the education system is obtained as follows:

$$M_{kt+1} = M_{k-1t}(1 - d_{k-1} - r_{k-1}) + M_{kt}r_k \qquad (10)$$

where,

M_{kt} := Enrollments in grade k in year t;

r_k := Repetition rate in grade k;[27] and

d_k := Dropout rate in grade k;[28]

In this instance, PROFLUXO avoids the under-reporting of repetition and over-reporting of dropout rates commonly found in official statistics.

It should be noted that the repetition rate is directly related to the period of average duration of instruction. The average duration of instruction v_k can be defined as total enrollments divided by the average number of entering and leaving students:

---

[27] Under stable conditions, the repetition rate is equal to the proportion of repeating students among total enrollments.

[28] Note that the promotion rate

$$p_k = 1 - d_k - r_k . \qquad (11)$$

However, in this formulation, "promotion" represents students in grade k who will not only pass the grade but subsequently enroll in grade k+1.

$$v_k = \frac{M_k}{1/2(I_k + D_k + I_{k+1})} . \quad (12)$$

Under stable conditions, the number of entering students equals the number of leaving students, i.e., $I_k = D_k + I_{k+1}$. In this case, Equation (12) can be simplified to:

$$v_k = \frac{M_k}{I_k} , \quad (13)$$

which is simply the total enrollment divided by the number of entering students. The proportion of new students among total enrollments is therefore the inverse of the ratio presented in Equation (13). The proportion of repeating students r_k among total enrollments can then be represented by the remainder:

$$r_k = 1 - \frac{1}{v_k} = 1 - \frac{I_k}{M_k} , \quad (14)$$

and under stable conditions, this is equal to the repetition rate.

The average duration of instruction v_k can also be calculated directly. We begin by using the <u>first derivatives</u> of P_I and P_A, shown in Figure 7, to calculate the mean ages of students as they enter and graduate from grade k.

The first derivatives of P_I and P_A, represent the frequency distributions of students as they respectively enter and pass grade k. The arithmetic mean provides a convenient measure of the central tendency in each of these distributions. The difference between the average age entering and the average age at graduation will then represent the average duration of instruction in grade k:[29]

$$v_k = \bar{i}_A - \bar{i}_I \quad (15)$$

It is thus possible to calculate the proportion of repeating students r_k among total enrollments directly by:

---

[29] For this to be so, it is important that the area under the two curves be equal. This is accomplished in practice by using the graduation rate A_k as the upper limit of integration for both curves.

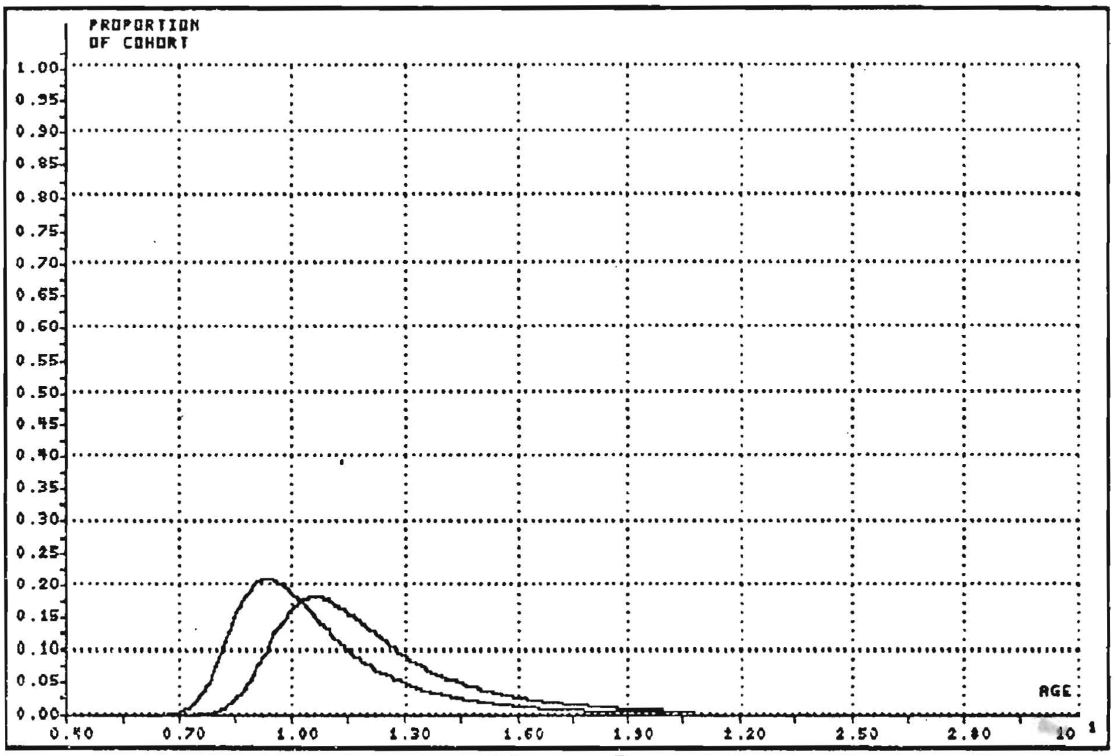

Figure 7

Frequencies of Students Entering and Graduating From Grade **k** in Relation to Age: First Derivatives of the Functions P_I and P_A

$$r_k = 1 - 1 / (\bar{i}_A - \bar{i}_I) \qquad (16)$$

and once again, under stable conditions, this is equal to the repetition rate.

Extension of the Basic Model to Incorporate Longitudinal Change in Education System Performance

In the previous section it was shown that, under allegedly stable conditions, the function

$$P_k(i) = 2A / (1 + e^{ai-b}) \qquad (4)$$

provides an appropriate representation of the proportion in each age group i having entered or graduated from grade **k** in a given year, while vectors $I_{k.}$ or $A_{k.}$, respectively, are rising to a maximum.

"Stable conditions" mean that education system performance is assumed to remain constant from year to year. Under these conditions, each of the age cohorts represented by vector $I_{k.}$ or $A_{k.}$ would be following the same trajectory. The situation of the population of age i with respect to grade k in the following years i+1, i+2, i+3... could thus be approximated by examining the proportions I_{ki+1}, I_{ki+2}, I_{ki+3}... or A_{ki+1}, A_{ki+2}, A_{ki+3}....

However, Figure 4, in the previous section, shows that the assumption of a stable education system in the Brazilian case (and probably most others) is violated since participation rates are rising over time (i.e., decline with successively older age cohorts). This implies that each age cohort represented by vector $I_{k.}$ or $A_{k.}$ is following a slightly different trajectory and will ultimately rise to a different grade participation rate.

When participation rates are rising over time, the asymptotic proportion **A**, encountered in the numerator of Equation (4), can no longer be considered a constant, but rather, is itself a function varying over time.

Exactly <u>what</u> function should substitute parameter **A** in Equation (4) is not immediately apparent. However, when dealing with proportions, such as that represented by parameter A, it would seem appropriate to use a function such as the logistic, which has the desirable property of being limited to the extremes of 0 and 1, respectively.

This reminds us of

$$L_k(i) = 1 / (1 + e^{c+di}) \qquad (6)$$

which was used previously to estimate the proportion of older age cohorts currently of age i having entered (or graduated from) grade **k**.

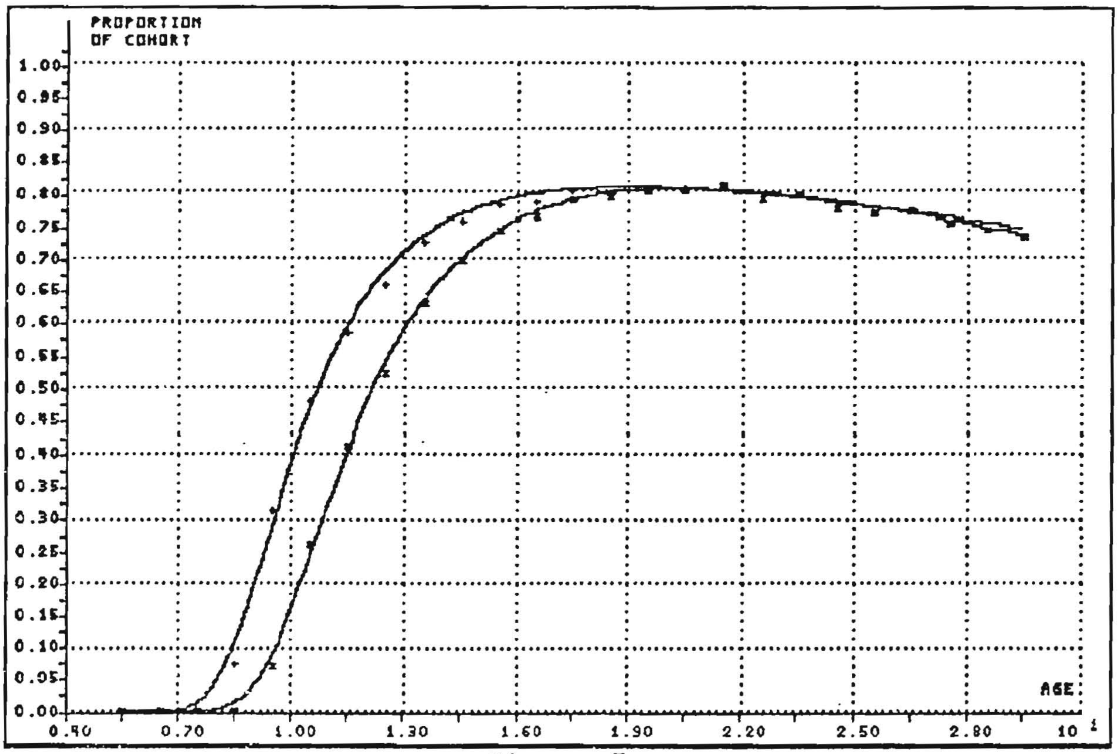

Figure 7

Using Function $M_k(i)$ to Represent Vectors $I_{k.}$ and $A_{k.}$, Respectively: Plot of Observed and Predicted Values

Replacing the parameter A in the numerator of Equation (4) with Equation (6), the proportions found in vectors $I_{k.}$ and $A_{k.}$ can thus be represented by

$$M_k(i) = 2 \, \frac{1 / (1 + e^{c+di})}{(1 + e^{ai-b})}, \tag{17a}$$

where i represents age.

When fitting the parameters of Functions (4) and (6), it was necessary to determine, a priori, the maximum of the observed data. This is risky, since the maximum of vectors $I_{k.}$ or $A_{k.}$, respectively, tend to reflect opportunistic sampling errors. By contrast, Equation (17a) requires no such a priori decision. The maximum of $M_k(i)$ is found at the top of each curve, where the function's first derivative is equal to zero.

When the function is applied to proportions entering grade **k**, the root of the first derivative of $M_k(i)$ will thus represent the intake rate in grade k, i.e., the number of new students among total enrollments.

Figure 7 shows Function $M_k(i)$, represented by solid lines, adjusted to the vectors $I_{k.}$ (+) and $A_{k.}$ (x), respectively.[30] Function $M_k(i)$ does not appear to represent the empirical data quite as well as Function $P_k(i)$ as participation rates are rising to a maximum. $M_k(i)$ tends to float above the empirical data as students begin to enter the grade and as participation rates approach the maximum (compare Figure 5). Biases of this nature are evident when examining a number of different grades in the Brazilian case.

Nevertheless, the standard deviation of error of estimation in the case we are examining, between 5 and 29 years of age, is only 1.0 percentage point in the case of entering students and 0.8 percentage point in the case of graduating students.[31] Moreover, the maximum of $M_k(i)$, used in calculating grade transition matrices, appears to represent the empirical data quite well.

Finally, it should be noted that a slight modification can be introduced in Equation (17a) so that change in education system performance over time will be explicitly represented. With this modification, a new function, $M_k(i,t)$ can be written as follows:

$$M_k(i,t) = 2 \frac{1 / (1 + e^{c+d(i-t)})}{(1 + e^{ai-b})}, \qquad (17b)$$

where i represents age and t represents time.

Up to this point, it has been assumed that t=0, i.e., that education system performance during the year of data collection is well represented by the data itself. By altering the value of **t**, Equation (17b) enables simulation of education system change over time, in particular, the rise in participation rates noted previously.

Figure 8a shows the sensitivity of Equation (17b) in relation to time. The Figure shows that time affects only the height of the curve but not is form, as the asymptote of the function changes

---

[30] Figure 7 is based on data obtained for the third grade of Brazilian first level schooling in 1982.

[31] For the nine grades of Brazilian schooling examined, the standard deviation of error in estimation is 0.8 +/- 0.2 percentage points for entering students and 0.8 +/- 0.3 percentage points for graduating students.

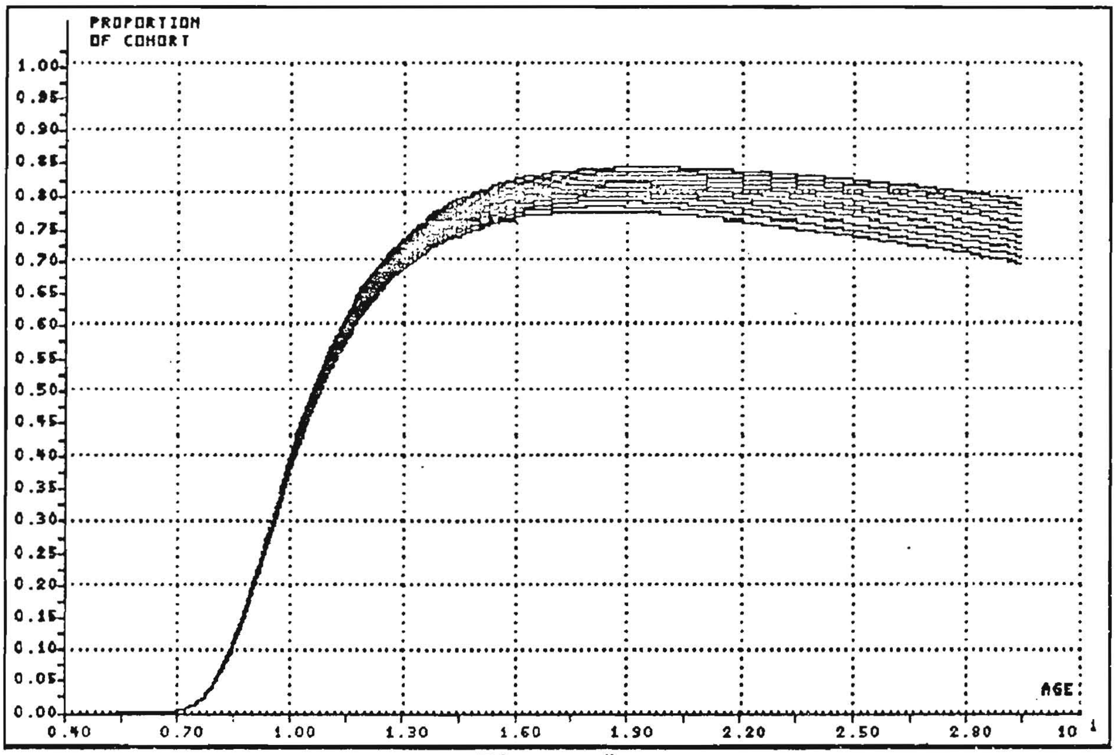

Figure 8a

Sensitivity of Function $M_k(i)$ with Respect to Changes in Time: Graphic Representation of a Wave Rising In Height over Time

with time. Seen from this perspective, the sequence of curves appear to represent a kind of stationary wave that is rising in height.

In Figure 8a, t changes by units between t=-5 and t=5. Based on data from 1982, this represents the period between 1977 and 1987. The Figure shows that grade intake rates, represented by the maximum of each curve, were rising rapidly in the third grade of Brazilian first level schooling during this period.

A slightly different perspective of the same phenomena is presented in Figure 8b. The curves are the same as those presented in the previous Figure, however, they have been positioned at age i' = i-t so that time advances from right to left. The allusion created by Figure 8b is that of a wave rolling forward and rising in height as time advances.

The perspective provided by Figure 8b highlights an important relationship found at the top of the succession of curves. The maximum of any curve should not be confused with its asymptote. The maximum represents the <u>intake rate in grade k</u>, whereas the

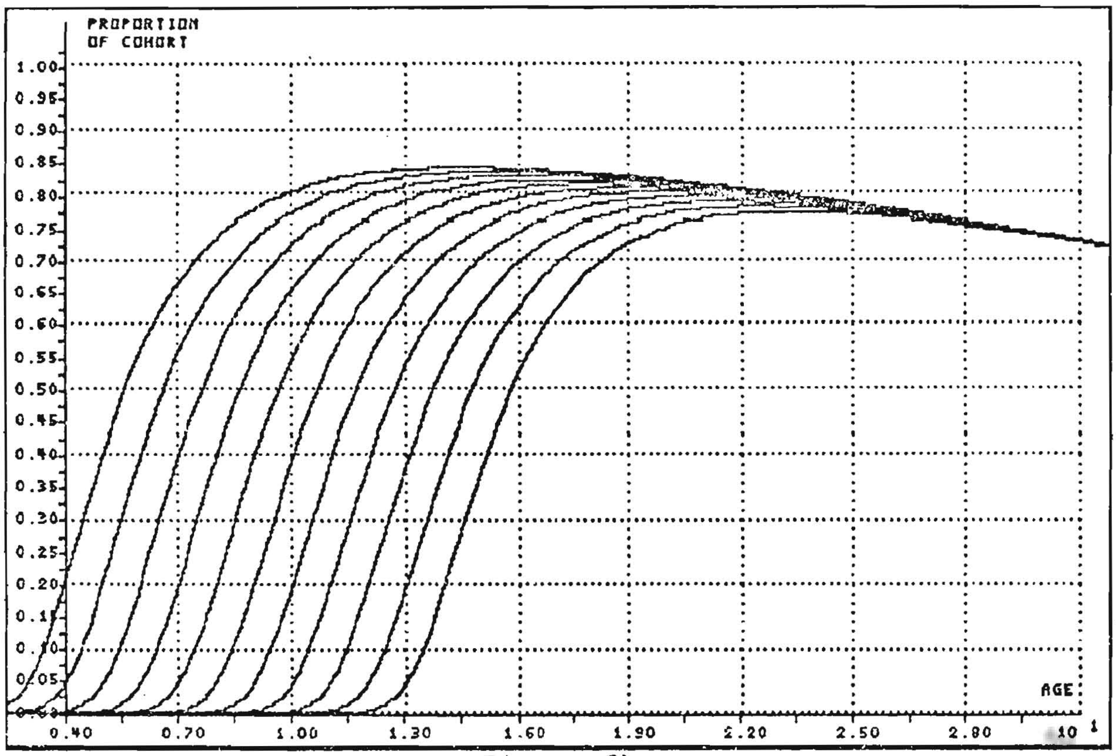

Figure 8b

Sensibility of Function $M_i(i)$ in Relation to Time t: Graphic Representation of a Wave Rolling Forward Through Time

asymptote represents the grade participation rate of an age cohort. The maximum is used to describe an education system in a given year; the asymptote is used to describe education received by different age groups in the population over a period of years.

This distinction is more readily visible in Figure 9. Superimposed on top of the original curve are two nearly straight lines. The upper line is the asymptote of Function $M_k(i,t)$, varying with time, and represents grade participation rates to be achieved by different age cohorts in the population.

Figure 9 shows that, whereas only 77% of the cohort 28 years of age in 1982 obtained a third grade education, secular trends suggest that fully 91% of the cohort 5 years of age in 1982 will eventually obtain a third grade education. Thus, over a period of 23 years, third grade participation rates in Brazil will tend to rise about 14 percentage points.

The second nearly straight line is the maximum of Function $M_k(i,t)$, varying with time, and represents intake rates in the third grade of Brazilian first level schooling. In this example,

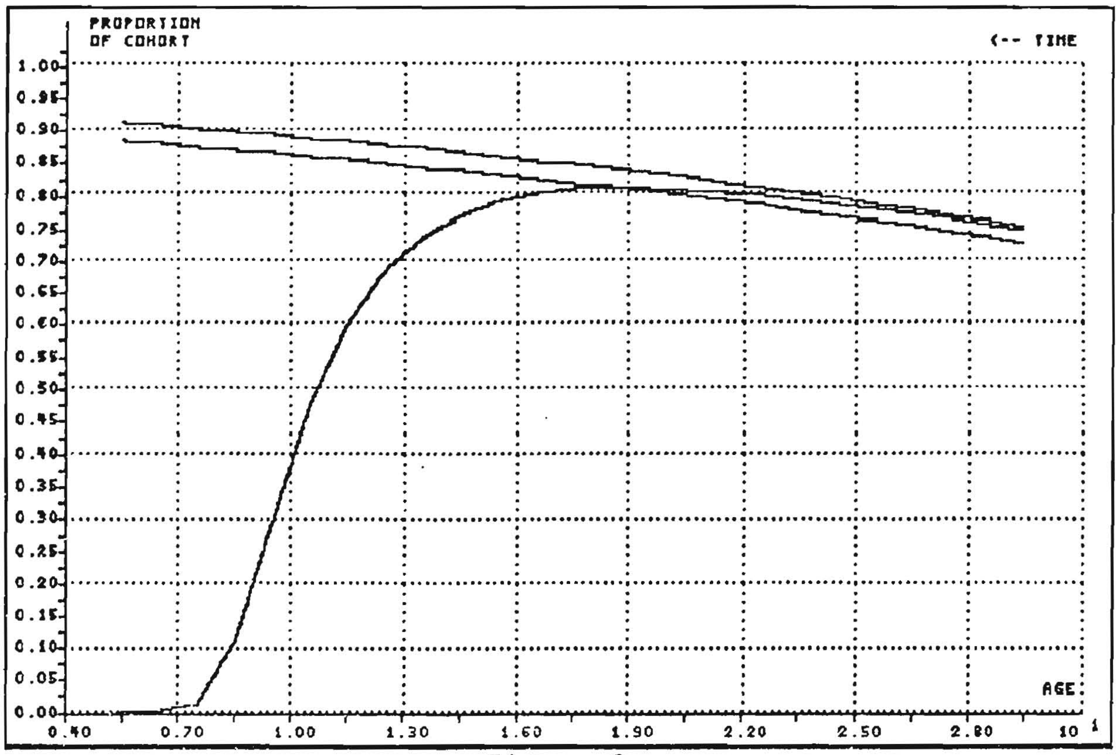

Figure 9

Changing Intake and Participation Rates: A Comparison of the Maximum and Asymptote of Function $M_k(i)$ over Time

the equivalent of about 81% of an age cohort enrolled as new students in 1982. The age distribution of new students is extremely varied; however, most new students will have entered the third grade by 19 years of age.

Even so, Figures 8b and 9 tell us that this is not the end of third grade education for the cohort 19 years of age in 1982. Enrollments are growing from year to year and thus a few more people in this cohort will enroll during the following years. However, the important point is that these additional enrollments come in following years and thus should not be counted among new enrollments in 1982.

This implies that the maximum of Function $M_k(i,t)$ should be used when calculating grade transition matrices in year t. The difference between Function $M_k(i,t)$ for entering and graduating students, respectively, can be used to estimate enrollments in year t and t+1, respectively. The other elements of the grade transition matrix can then be easily calculated, following procedures discussed in the previous section of this report.

Equation (17b) thus permits a dynamic simulation of education performance over time.

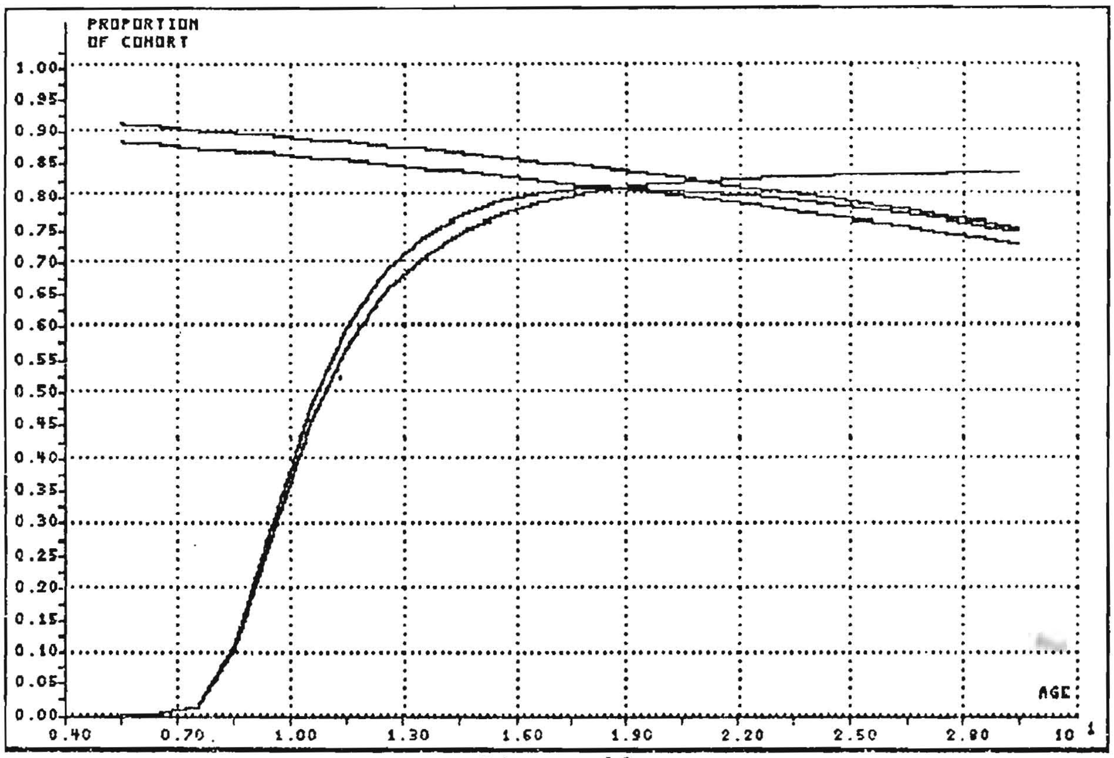

Figure 10

Reconstructed School History of a Single Age Cohort:
The Cohort 19 Years of Age in 1982 Entering the Third Grade
of Brazilian First Level Schooling

On the other hand, if the objective is to describe the school history of a single age cohort, it is necessary to remove the effects of changing participation rates from the numerator of Function $M_k(i,t)$. When this is accomplished, the asymptotic function appearing in the numerator of Equation (17b) is once again replaced with a constant.

Supposing that the cohort of interest is currently of age i, two conditions are to be imposed. Once these conditions are accepted, there is, in principle, only one possible solution to this problem.

First, any adjustment to be made eliminating the effects of changing participation rates must necessarily have the same value $M_k(i,t)$ at age i in year t.

Second, the parameter associated with time in the numerator of Equation (17b) should be set to zero. The effects of time on the value of the numerator of the Function are thus eliminated.

In this case, Equation (17b) is reduced to

$$M_k(i) = 2 \frac{1 / (1 + e^c)}{(1 + e^{ai-b})} . \qquad (18)$$

Having selected the cohort of age **i**, we solve for the value of **c**. Giving the name **A** to the term $1 / (1 + e^c)$ in the numerator of (18), then

$$A = \frac{M_k(i)(1 + e^{ai-b})}{2} . \qquad (19)$$

Finally,

$$c = \ln ((1 - A) / A) . \qquad (20)$$

Figure 10 reproduces the curves of the previous Figure, with the addition of a new curve representing the reconstructed school history of a single age cohort. This example shows the cohort 19 years of age in 1982 entering the third grade of Brazilian first level schooling.

Generalizing the Basic Model in Order to Represent Different Subgroups in the Population

Since the PROFLUXO model is based on demographic data, it is well suited for showing how the education system affects different subgroups in the population. Analysis of different groups is important when considering questions of equity.

One way of studying this issue is to produce a series of separate analyses. Each analysis involves a separate "unit of analysis," comprised of people having the same characteristics. Parameters are generated and grade transition matrices are calculated for each unit of analysis. Comparison of the results may reveal important differences in education opportunities. Formally, the logic reminds us of contingency table analysis.

However, partitioning the original data set into separate units of analysis may become impractical when dealing with survey samples of limited size. Depending on the distribution of the population among the different categories of interest, there simply may not be a sufficient number of observations in certain units of analysis.

Based on age-grade relationships, the PROFLUXO model uses only small number of variables; however, stability in the proportions estimated with the model require a fairly large number of observations--probably at least one hundred--in each age group. This requirement becomes increasingly difficult to satisfy when dealing with many units of analysis drawn from the same sample. The problem resembles that encountered with contingency tables containing low cell frequencies.

A second problem arises when the classificatory variable of interest is continuous. This occurs with virtually every measure of socio-economic status, a variable that is likely to be extremely useful in studying questions of equity. Separate analyses of percentiles, quintiles or even deciles of the national distribution of income would be extremely onerous and unwieldy.

These problems suggest that the approach involving separate units of analysis makes inefficient use of the data. An alternative that comes to mind involves multivariate techniques. In this case, each of the parameters used in PROFLUXO equations would become a function of the different characteristics of the population.

This possibility is explored in this section of the report using a quantitative measure of socio-economic status scaled to represent the different percentiles of Brazilian income per capita. In principle, the same technique could be extended to include any number of qualitative characteristics of the population represented by a series of binary variables.

But first a word of caution. The PROFLUXO model is based on the principle of simulated age cohorts. In order to accompany the same category of persons through successive years of age, it is necessary to assume that there is a relatively low incidence of "migration" of persons across categories. The problem of rural-urban migration illustrates this problem. If large numbers of undereducated persons migrate to the urban areas in early adulthood, this will produce a distorted image of urban education when the urban data is processed by PROFLUXO.[32]

Thus, in general, PROFLUXO is restricted to the analysis of permanent or relatively permanent characteristics of the population. Where there is a high incidence of student transfers, it would be extremely hazardous to analyze individual schools or even jurisdictional categories of schools with this methodology. In particular, analysis of the performance of private and public school systems would seem inappropriate.

When dealing with a measure of socio-economic status, such as a scale based on household characteristics or income per capita, it should be noted that families and individuals tend to improve their living conditions with advancing age. Persons at a certain percentile in the distribution of income will, however, tend to maintain their relative position over time. Thus, the most appropriate use of socio-economic measures involves normalization in each age group prior to analysis in order to preserve the notion of a cohort.

To show how the basic model can be generalized to n characteristics of the population, we will explore the relationship between socio-economic status and education system performance. Figure 11 shows how participation rates in Brazilian secondary education vary as a function of socio-economic status in the urban southeast, and most developed, region of the country.

The Figure shows that participation rates across socio-economic categories are well represented by a logistic function

---

[32] This problem can be partially controlled by restricting the urban sample to persons who are long term residents of the urban areas, provided that the original data contain information on period of residency. Out-migration is generally a more serious problem due to the lack of information.

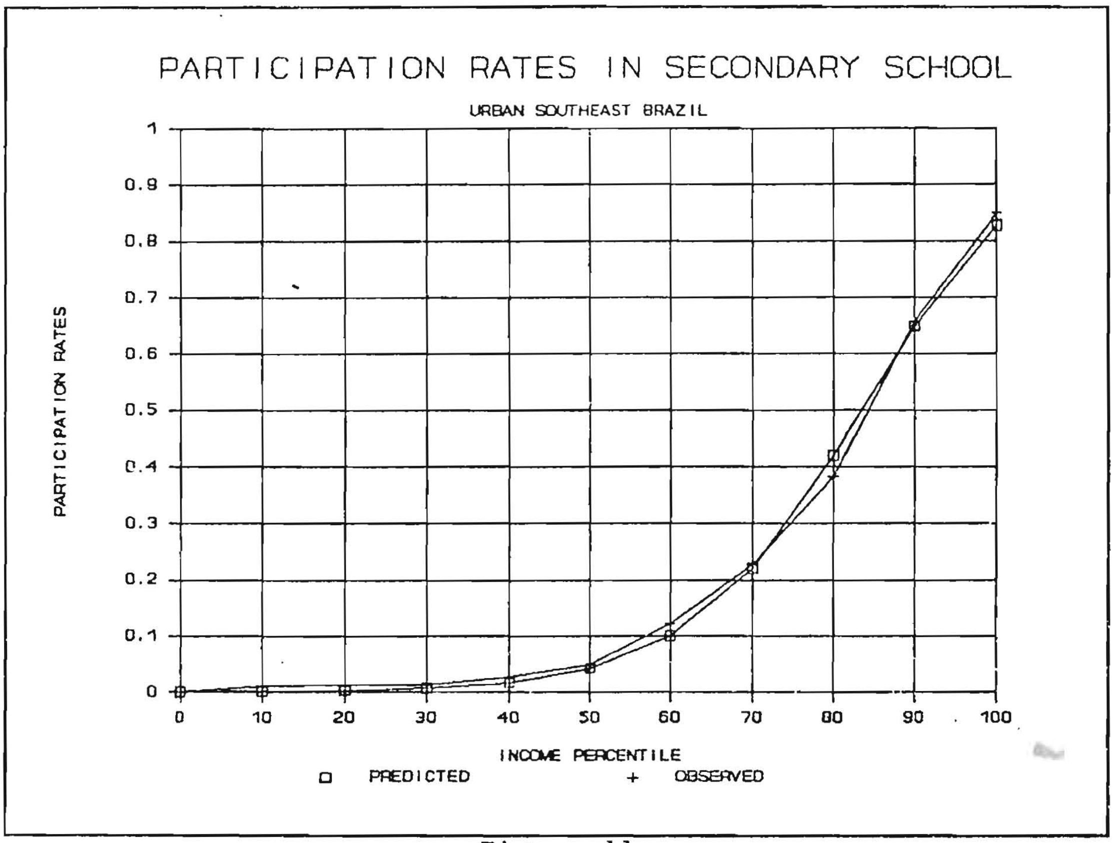

Figure 11

Probability of Entering Secondary School by Income Percentile
Predicted and Observed Values

fitted to the empirical data.[33] Indeed, the efficiency of the logistic function in predicting participation rates has been verified at a variety of different grade levels in a number of areas throughout Brazil. This suggests that the logistic function appearing in the numerator of Equation (17b) should be altered to include a parameter for a measure of socio-economic status.

The logistic in the numerator would then appear as follows:

$$1 \,/\, (\, 1 + e^{c+d(\, i-t\,)+fs}\,)\,, \qquad (21a)$$

where f is the new parameter and s is an appropriate measure of socio-economic status. In this representation, the log likelihood

---

[33] The authors express their appreciation for the assistance provided by João Batista Ferreira Gomes Neto in demonstrating this relationship.

ratio of participation rates is a linear function of time and socio-economic status.

Equation (21a) allows participation rates to rise in the dimension of time and also over successive levels of socio-economic status. However, it also seems likely that the rate of increase over time depends upon the socio-economic status of students. Different rates of increase at different levels of income are taken into account by introducing a multiplicative term which is simply the product of i-t and s:

$$1 \,/\, (1 + e^{c+d(i-t)+fs+g(i-t)s}) \,. \qquad (21b)$$

The parameters a and b in the denominator of Equation (17b) move the rising part of the curve left and right and determine the steepness of ascent. Stated somewhat differently, they determine the mean and standard deviation of the age distribution of students entering or graduating from the grade.

Preliminary analyses with grouped data showed that the values of a and b vary with the socio-economic status of students. Compared with more affluent students, lower income groups enter a given grade at a later age and the age distribution of new students at lower income levels is less homogeneous.

The addition of two new parameters, p and q, to the denominator of Equation (17b) allow for these contingencies:

$$M_k(i,t,s) = 2 \, \frac{1 \,/\, (1 + e^{c+d(i-t)})}{(1 + e^{(a+ps)i-(b+qs)})} \,. \qquad (22a)$$

Substituting the numerator of Equation (22b) with (21b), we obtain the final form of Function $M_k(i,t,s)$:

$$M_k(i,t,s) = 2 \, \frac{1 \,/\, (1 + e^{c+d(i-t)+fs+g(i-t)s})}{(1 + e^{(a+ps)i-(b+qs)})} \,. \qquad (22b)$$

Figure 12 shows the efficiency of Equation (22b) when adjusted to third grade participation rates representing three socio-economic levels representing the 74^{th}, 36^{th}, and 12^{th} income percentiles, respectively.

These three levels offer only three degrees of liberty in the socio-economic dimension; a logistic function, such as that found in the numerator of Equation (22b) will adjust three points without error. The question raised below is whether the mean and standard deviation of the age distribution entering the grade can be adequately represented with the modification introduced in the

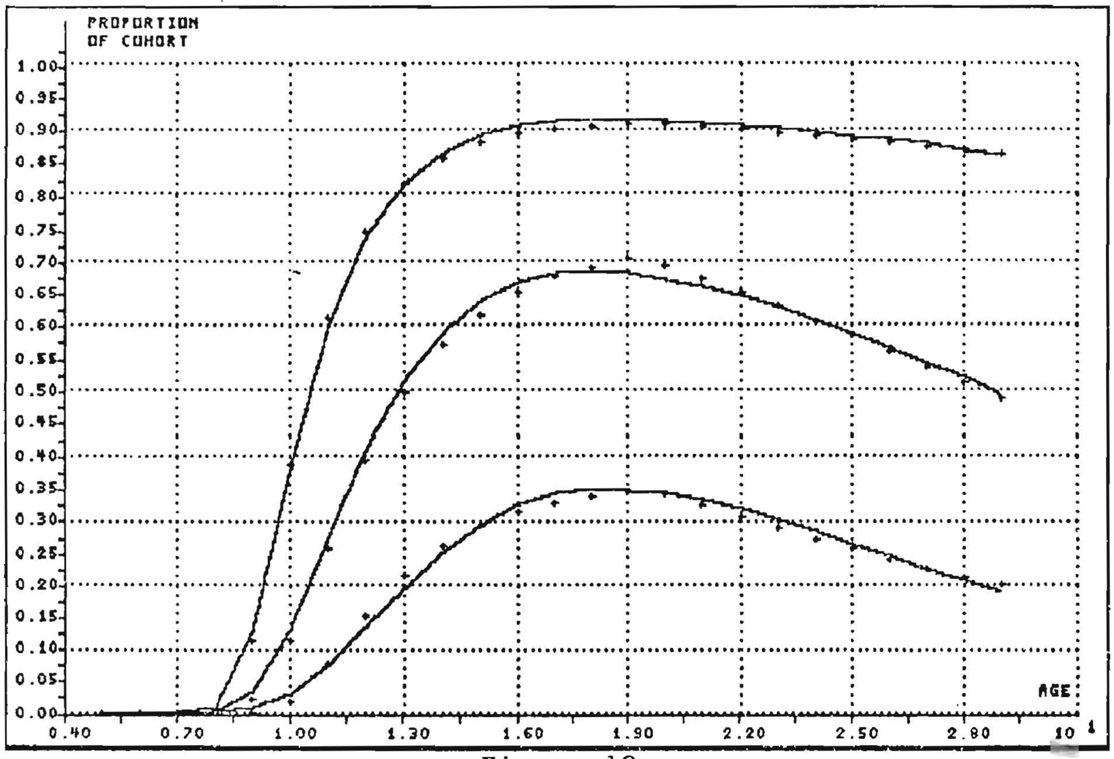

Figure 12

Function $M_k(i,t,s)$ Fitted to Data Representing Three Socio-Economic Levels: Observed and Predicted Values

denominator of Equation (22b).

For the set of data points presented in Figure 12, the standard deviation of the residuals is only 1,0 percentage point. In 24 fittings, the standard deviation remained below 1,0 percentage point in 22 cases (92%). The worst fitting obtained involved errors of 1,6 percentage points. This suggests that Equation (22b) is appropriate for representing participation rates in their temporal and socio-economic dimensions.

By varying the value of i in Equation (22b), grade participation rates can be obtained for different ages. By changing the value of t, grade participation rates can be projected backwards and forwards in time. By altering the value of s in the same equation, the same proportions are obtained for different income percentiles. One function serves many useful purposes.

In principle, combinations of different values for each variable produce estimates for a practically endless variety of conditions. In practice, however, inconsistencies across grades appear when extrapolating well beyond the limits of the original data. For example, the participation rate in grade k+1 may exceed that of grade k when t=50, for example.

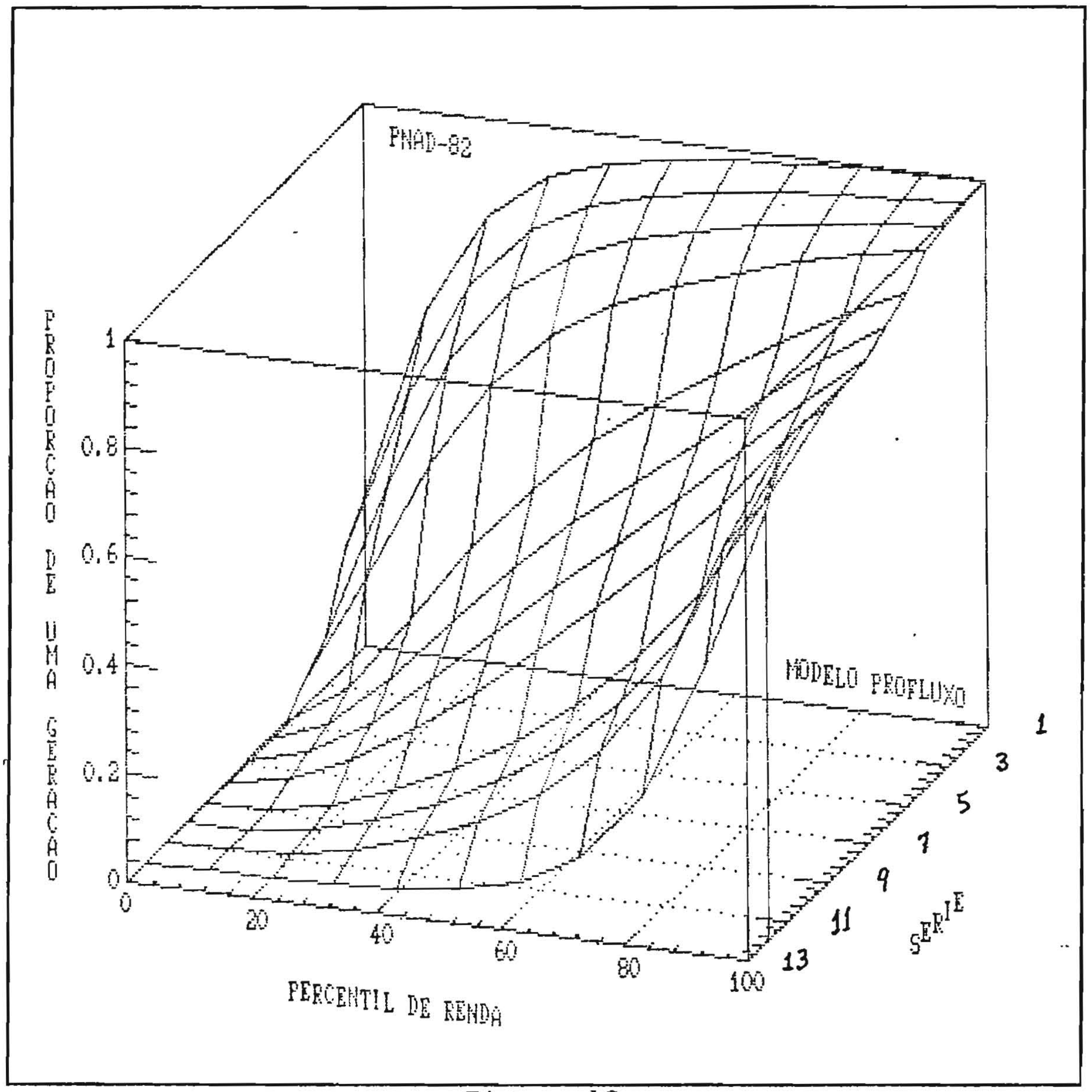

Figure 13

Grade Participation Rates in Brazilian Education, by Grade and Income Percentile

Nevertheless Equation (22b) does appear to provide considerable descriptive detail. Figure 13 shows participation rates by income percentile for practically the entire education system. Twelve "grades" are represented. The first eight grades comprise first level schooling. The "ninth" through "eleventh" comprise secondary education. In this representation, the "twelfth" grade coincides with the first year of university education.[34] No attempt has been made to estimate participation rates beyond the first grade of higher education.

Lines running left to right show participation rates in a given grade according to income level. The lines running from the back of the Figure towards the front show participation rates in different grades for people of the same income level.

The Figure shows that the first grade coverage of the population has been extended to nearly all of the population in the same way that university education has been attained by the most wealthy. Between these extremes, between the back and sides of the Figure, there is a trough that is only partially bridged by the fifth, sixth, seventh and eighth grades.

The first few grades of schooling represent an insuperable obstacle for those below the 20^{th} income percentile. By contrast, the fifth, sixth, seventh and eighth grades eliminate students representing a wide variety of socio-economic levels. At the secondary and especially the higher education levels socio-economic status again plays an important role in determining who gets access to further schooling.

Virtually no one below the 30^{th} income percentile enters secondary schooling, just as decidedly few below the 60^{th} gain access to university education. The vast majority of students entering higher education belong to the wealthiest 20% of the population.

The first derivatives of the lines running left to right in the Figure represent income elasticities of demand for the different grade levels of schooling. These elasticities can be calculated precisely using Function (22b); here, we can only highlight general aspects.

The first few grades of schooling appear to be extremely important as incomes rise among the poorest segments of the

---

[34] Terms in parentheses are inconsistent with Brazilian nomenclature. First level schooling is comprised of eight grades. Second level schooling covers another three grades. Third level schooling comprises still other grade levels, which vary in number, depending on the course of study. Numbering grades sequentially merely facilitates the graphical presentation.

population. Secondary and higher education become extremely important when incomes rise among the richest 20% of the population.

The intermediate grades of schooling presumably become important somewhere around the 50^{th} percentile, but this is hardly discernable. The surprising thing about these intermediate grades is that there does not appear to be a precisely defined income level at which the need for additional schooling acquires an overwhelming importance.

Nevertheless, it seems quite evident that policy decisions favoring one level of schooling over another benefit certain social groups at the cost of others. The redistributive benefits of higher expenditures for university programs in the Brazilian case appear quite limited. It could just as easily be argued that an increase in first grade enrollments would benefit upper income groups.

The first derivatives of the lines running from back to front in the Figure represent the potential demand for schooling at different grade levels. Where grade participation rates fall precipitously from one grade to the next at a given socio-economic level, there is the greatest potential for enrollment growth. Figure 12 shows that, in the Brazilian case, this potential is greatest at the fifth grade level.

The importance of Function $M_k(i,t,s)$ is thus not merely descriptive. Grade transition matrices, showing how current generations obtain their education, show a number of important aspects that are useful in monitoring education system performance. Projections forward in time would appear to be useful in programming the need for future resources. Analysis of educational opportunities made available to different subgroups in the population raise potential issues of equity.