Education and training systems are complex organizations of facilities, instructors, equipment, routines, and students. Because of this complexity, mathematical models can be used effectively to explain the behavior of educational systems and to assist in making decisions relative to better design and operation, and improved cost-effectiveness, capacity, and quality. This document, therefore, presents an assessment of the present state of the art of modeling educational systems. Important to all areas of educational research, the document is written in relatively nontechnical terms and is intended primarily for administrators who are assumed not to have strong mathematical backgrounds. A technical appendix is included for those readers who wish to pursue the actual construction of models in greater depth. (Author)
MATHEMATICAL MODELS IN EDUCATION AND TRAINING

Allen Hammond

This research is supported by the United States Air Force under Project Rand—Contract No. F44620-6-C-0045—monitored by the Directorate of Operational Requirements and Development Plans, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this study should not to be interpreted as representing the official opinion or policy of Rand or of the United States Air Force.
PREFACE

The study reported in this Memorandum is part of the initial phase of Rand's work in Air Force Technical Training. A primary objective of that work is to determine ways in which technology can be used to aid in the design of instruction for formal technical training. Education and training systems are complex organizations of facilities, instructors, equipment, routines, and students. Because of this complexity, mathematical models—both existing and future—can be used effectively to explain the behavior of educational systems and to assist in making decisions relative to better design and operation, and improved cost-effectiveness, capacity, and quality.

This Memorandum presents an assessment of the present state of the art of modeling educational systems. The existing models have been developed outside of the military; however, the modeling concepts are applicable and of importance to all areas of educational research. Therefore, this study should be of use to those concerned with policy and planning at DCS/Technical Training and the Training Development Directorate, Headquarters Air Training Command.

The text of this Memorandum is written as a general introduction to the field, in relatively nontechnical terms, and is intended primarily for administrators who are assumed not to have strong mathematical backgrounds. A technical appendix is included for those readers who wish to pursue the actual construction of models in greater depth.
Present-day mathematical models of educational systems can provide useful answers to limited but important quantitative questions concerning budgeting, resource allocation, and enrollment planning. Such models are designed, at the conceptual level, by determining the major features of the system, outlining their interrelationships with a flow chart, and choosing the variables to be used. If, in addition, specific assumptions embodying the educational "physics" of the model or empirical relationships are included, the model can then be solved; that is, each dependent variable can be stated as a function of the independent variables, and the consequences of the assumptions can be determined. The model structures that result can be characterized by their scope and complexity, by the degree of aggregation of the variables employed, by the model inputs and outputs, and by the purpose for which the model is to be used.

A number of representative existing models of various types are discussed. Input-output models can be a convenient way to examine large amounts of data on enrollments and student flows, but these models are limited in that a current cross-sectional analysis is generally used to predict the future time series of the variables. Input-output models may find wide application for analyzing systems with relatively static structures, however, such as training institutions for specific purposes.

Manpower planning models seem to be less useful than many other models. Because these models do not provide explicit allocations of educational resources and because they do not describe actual student flows, they are perhaps too simplified for the problem they attempt to solve. Optimization models have the advantage of making explicit the basic choices of a resource-allocation problem, when the desired benefits can be quantitatively described. Since such models yield priorities and plans as output, they are more likely to stimulate discussion at the policy level. Simulation models will be of considerable assistance in management and short-term planning for educational systems, but they run the risk of foundering in a wealth of detail. It may well be that optimization and simulation models can serve in complementary ways in
educational planning, one operating on the policy level and the other on the detailed operational level. Finally, where specific educational mechanisms can be identified, relatively simple models can be extremely effective. The usefulness of simple models for flows of students and teachers can be extended further by including cost factors and other simple economic variables, but without attempting to model all aspects of an educational system.

More research is needed to increase our understanding of the dynamics of educational systems. The appropriate mathematical basis for the research suggested here would be very simple; stochastic models for probabilistic problems and simple difference and differential equations for deterministic problems, coupled with optimization or simulation techniques where appropriate, should be adequate for most modeling of educational systems in the near future. These techniques and the mathematical structure of educational-system models are discussed in the Appendix. A selected bibliography is also included.

The author is a Consultant to The Rand Corporation.
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREFACE</td>
<td>iii</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>v</td>
</tr>
<tr>
<td><strong>Section</strong></td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. THE EDUCATIONAL-SYSTEM MODEL: DESIGN AND STRUCTURE</td>
<td>3</td>
</tr>
<tr>
<td>Model Design</td>
<td>3</td>
</tr>
<tr>
<td>Model Structures</td>
<td>6</td>
</tr>
<tr>
<td>III. SELECTED EDUCATIONAL-SYSTEM MODELS: A GUIDE TO THE LITERATURE</td>
<td>13</td>
</tr>
<tr>
<td>IV. CONCLUDING REMARKS</td>
<td>20</td>
</tr>
<tr>
<td>Appendix: MATHEMATICAL STRUCTURE OF EDUCATIONAL-SYSTEM MODELS</td>
<td>23</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>27</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>29</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

It is only in the past few years that researchers and analysts have begun to develop models of educational systems—that is, quantitative, systematic descriptions of the operation of educational systems and the behavior of their component parts. The growing body of literature on the quantitative characteristics, or system character, of educational and instructional systems reflects a change in the focus of educational research, which was previously concentrated almost exclusively on individual instructional processes. Not that administrators and educational planners have not had to concern themselves with enrollment figures, costs, and other quantitative variables in the past; their decisions, however, have all too often been based on guesses or the crudest of estimates.

Present-day models are a long way from being able to reliably relate the variables in educational systems (students, subject matter, teaching methods, teachers) to the immediate outputs of education (the learning of facts, skills, and attitudes), because the "physics" of learning remain unknown. The model-makers can, however, provide some impetus for research into basic questions, such as how learning is brought about by the instructional process, and how that process should be organized to serve the needs of the nation (or state, or institution) and those of the students. Although few explicit answers to such basic questions are available, models can often yield quite definite answers for a host of subsidiary questions. Our primary interest here, therefore, will be in these subsidiary questions.

Many of the educational-system models that have been published do not accurately represent the behavior of educational systems, or they represent that behavior only in sharply limited ways. Nonetheless, such models are the beginning of a more consistent, analytic approach to educational planning than has heretofore existed, and they can be of great help in answering certain kinds of questions if their uses and limitations are understood. To the administrator faced with his yearly budget crisis, for example, a model may provide a more accurate estimate of the next year's enrollment; for the national planner
in a developing country, a model may enable a more effective allocation of scarce educational resources, such as teachers. While problems of budgeting and resource allocation are subordinate to more basic educational concerns, they are important issues in an educational system and both are characteristic of types of problems that are amenable to modeling at present.
MODEL DESIGN

Models can range in complexity from the simplest drawing to a full computer simulation of a complex system. All models seek to idealize reality as a structure comprising sets of elements and relationships among them. Model building is the selection and definition of the elements and their interconnections. For example, a model-maker might choose to view a school as a series of grade levels through which groups of students move, with rules specifying how many are to proceed to the next grade level. Or he might choose to divide a school into subject areas, such as language arts, social studies, and physical education, with rules which relate the number of teachers, budget dollars, and student hours allocated to each area.

It is important to emphasize the arbitrary character of both the choice of the major features for the model and the nature of the formal relations between these features, or variables. As the example above shows, which variables will be useful depends entirely on the purpose of the model and on the type of questions it will be used to address. Similarly, it is not necessary—and, in fact, in educational models is rarely true—that the formal statements or functional relationships assumed between the variables really express directly the "physics" of the process, the true causes and effects; all that is necessary is that these relationships give the correct empirical answer.

A specific example will help to illustrate this characterization of an educational-system model and will serve as a useful framework for some additional terms and concepts to be discussed later. Let us assume that we are a federal administrator concerned with the progress of graduate schools throughout the country, and in particular with the production of Ph.D.'s. We wish to study the factors that influence the functioning of graduate schools, with special attention to those factors over which we might have some control; and we wish to be able to estimate or predict the numbers of graduate students and Ph.D.'s in future years.
The first step is to divide the process with which we are concerned into internal and external features. This division may be indicated symbolically on a flow chart, as shown in Fig. 1. Those features pertaining to the internal workings of graduate schools are enclosed in a box labeled graduate schools; the external features are divided between outputs and inputs, signified by arrows. The inputs are further divided into "federal" and "other," since we are specifically concerned with the federal influences.

![Flow chart for illustrative educational-system model](image)

Next we identify the variables (the items whose numerical value provides a measure of a quantity relevant to the educational process) that we will use to characterize the state of each part of the process. In educational models, the choice of variables is often strongly limited by the types of data that are available. For our example, we might identify the following as useful variables:

**For the model as a whole**

- \( k \) = time, in units of academic years

**Items within the graduate schools**

- \( G \) = the number of graduate students
- \( A \) = the number of professors in graduate schools
- \( B \) = the budget of graduate schools, in dollars
Inputs to the graduate schools

\[ F = \text{total federal funds allocated to graduate schools} \]
\[ U = \text{the number of individuals entering graduate schools} \]

Output from the graduate schools

\[ D = \text{the number of Ph.D.'s awarded} \]

Here we have chosen time, \( k \), as the independent variable, that is, the one which can assume any given value; this is a variable over which we have control. The other variables are dependent variables, since their value depends on which year \( k \) we are talking about. We express this mathematically by saying that \( G \) and the other dependent variables are functions of \( k \); we represent this relationship by, for example, \( G(k) \) or \( G_k \).

It should be evident that the choice of variables in this example is somewhat arbitrary and would vary widely with the specific purpose of the model and with the tastes of the model-maker. Although the most common independent variable in models of educational systems is time, other possibilities include variables representing subject matter, student ability, teacher ability, money (or budget levels), and even school-bus routes. Common dependent variables are numbers of students, numbers of teachers, material-resource variables (including money, classroom space, supplies), and national economic indicators such as gross national product. The variety in the exact definitions of the variables used is almost as great as the number of models extant. When dealing with simple models of complicated phenomena—as is always the case in modeling educational systems—there is no unique formulation of a model, no necessarily best choice of variables and assumptions. The test of a model is its accuracy in reproducing or predicting behavior and its usefulness to the problem at hand.

In trying to model the complexities of educational systems, it is well worth while to master a very limited model, whose assumptions and limitations can be easily held in mind, before attempting to include more variables. In the hypothetical example given above, the real output of graduate education to the society may include the production of
research, advanced training even for those who do not complete the doctorate, and assistance for undergraduate training—much more than is measured by the number of Ph.D.'s produced. However, these additional features are not easy to characterize or measure and could greatly complicate the model. The basic rule is always to try the simplest thing first.

MODEL STRUCTURES

The types of models with which we are concerned may be divided into three classes: conceptual models, mathematical models, and gaming models. A conceptual model is one that establishes an idealized framework of the process under study and identifies the variables but does not include assumptions and specific statements on how the variables are related; it contains no mathematics. Such models do not produce operationally useful results, so there is no way to compare the idealization to reality. A mathematical model includes, in addition to the above, specific assumptions connecting the variables to each other, stated in mathematical form. These assumptions embody either the educational "physics" of the model or, more commonly, arbitrary relationships for which there is empirical justification. Mathematical models, properly formulated, can be solved, meaning that each dependent variable can be stated as a function of the independent variable, and thus the consequences of the assumptions can be determined. A gaming model is a mathematical model in which some of the variables are human beings playing decision-making roles. A gaming model is usually "solved" by computer simulation or other means, as part of a total environment to aid decision-makers. This type of model, often used in military and business contexts, has had as yet little application in educational systems; therefore we will not discuss gaming models further except to point out their potential usefulness.

The illustrative graduate-school model we have been discussing is an example of a conceptual model. Though numerous in the literature, models of this class are useful only as a first stage in the construction of a mathematical model. To illustrate the evolution, let us

*There must be as many assumptions as dependent variables.
continue with our example. Using the variables previously defined and the flow chart given in Fig. 1, we make assumptions relating the variables, thus constructing a mathematical model:

1. The number of graduate students grows at a constant rate, \( a \), given by
   \[ C_{k+1} - C_k = a. \]

2. A constant fraction, \( d \), of the number of graduate students graduates every year with a Ph.D.; thus
   \[ D_k = d \cdot C_k. \]

3. The number of faculty is proportional to the graduate-school budget, i.e.,
   \[ A_k = e \cdot B_k. \]

4. The graduate-school budget is proportional to the amount of federal aid, i.e.,
   \[ B_k = f \cdot F_k. \]

5. Federal aid to graduate schools grows linearly with time, thus
   \[ F_k = F_0 + g \cdot k. \]

Here we have left out any explicit mention of \( U \), the number of persons entering graduate schools, and instead we have made an assumption about the overall growth rate of the graduate school. In formulating these equations we have also introduced some parameters into the model: \( a, d, e, f, \) and \( g \). These parameters are constants that express the relationships between the variables and are usually determined from data on the actual operation of the system being modeled. For example, \( d \), which can be viewed as a passage fraction for the graduate school "class," would be given the value of the ratio \( D/C \) for the most recent year for which data on degrees and graduate students were available.

We are primarily interested in predicting the number of Ph.D.'s that will be granted. For this number the solution of the model gives
Fig. 2.--Solutions of the model

the formula $D_k = D_0 + a \cdot d \cdot k$, which predicts a linear growth (one described by a straight line) in degrees awarded. In fact, the assumptions in this model were chosen so that all the variables would grow linearly (see Fig. 2), giving the same predictions as would linear projections from past data. Linear projections are one of the easiest, if crudest, methods of estimating future values, and are still widely used; our model shows the kinds of assumptions that are implicit in these estimates. Since such estimates have erred seriously on the low side when compared to the actual data for the past twenty years, it is instructive to consider an alternate set of assumptions, which lead to rather different predictions:

1. The number of graduate students in year $k + 1$ is the number in the previous year plus the number of incoming students, minus the number that graduated, minus that fraction of students, $b$, who left without receiving a degree:

$$G_{k+1} = G_k + U_{k+1} - D_k - b \cdot G_k.$$
6. The number of students entering graduate school increases like \( a^k \), where \( a > 1 \):

\[ U_k = U_0 \cdot a^k. \]

Assumptions 2 through 5 are unchanged from the previous model.

In this model the input of new students is assumed known and is modeled by an exponentially increasing function of time. The equation modeling the number of graduate students expresses the principle of conservation of students (the increase equals the number entering less the number leaving). These equations can also be solved to give, for the number of Ph.D.'s, the formula

\[ D_k = D_o \left(1 - \frac{d \cdot a \cdot U_o}{(a + b + d - 1)D_o}\right) (1 - b - d)^k + \frac{d \cdot U_o \cdot a^{k+1}}{(a + b + d - 1)}. \]

This formula, more complicated than the corresponding result of the earlier model, has two parts: The first becomes rapidly smaller as \( k \) increases, while the second increases in proportion to the growth in the number of incoming students. The present (revised) model, therefore, predicts exponential growth (after an initial equilibration period) in the number of Ph.D.'s (see Fig. 3); this results in many more Ph.D.'s than would be predicted by the previous model. Many of the models in the literature that are concerned with numbers of students use similar assumptions, including "conservation of students," constant "passage" and "dropout" fractions.
(d and b in this model), and exponential growth of numbers of incoming students.

We could construct other sets of assumptions for this illustrative mathematical model; instead we recommend to the reader a study of the literature,* a guide to which is provided in Section III of this Memorandum. Before discussing particular models, however, we shall introduce several additional concepts that are useful in distinguishing one model from another.

The scope and complexity of a model are measures of how much it tries to include. The most ambitious models have dealt with an educational system imbedded in a national economy. Others have dealt with an entire educational system, including interlocking levels of education but treating the system as an isolated phenomenon. Still others are concerned only with a small part of an educational system, such as the logistics of bussing students from one school to another. The scope of a model relates to the size of the system being modeled, usually in terms of the numbers of people involved.

Regardless of its scope, a model may include many types of phenomena and many kinds of variables, or it may deal with only a few variables. The complexity of a model indicates the number of relationships and the variety of phenomena explicitly included in the model.

Related to the scope of a model is the degree of aggregation of its variables; a variable with a high degree of aggregation combines in one symbol many items of the same type which may be physically distinct or independent. For example, a variable that represents the total number of graduate students in a country, even though those students are enrolled in many different states, has a high degree of aggregation, while a variable which represents the first-year graduate students in a particular subject at a particular school has a much lower degree of aggregation. In using highly aggregated variables, the model-maker incurs additional problems with data collection and interpretation of results but gains the advantage of increasing generality. Although it may not

---

*See for example Bolt's model for predicting the production of Ph.D.'s,(1) which leads to more accurate predictions than either of the two models discussed above.
be intuitively clear, we feel that the use of highly aggregated variables often results in more accurate models, at least for the types of assumptions and the relatively simple models of educational systems that have been attempted to date.

The inputs required by a model are those pieces of information that the model-maker must supply from empirical data sources, such as parameter values and initial values of the variables. Inputs in this context are to be distinguished from input variables, which refer to particular external variables on a flow chart. If the inputs required by a model are not obtainable from empirical data, then the model is of little use. Similarly, the outputs of a model are those pieces of information that are provided by the solution of a model, namely predictions of the values of the dependent variables.

The goals of a model relate to how and in what way its output will be useful. For example, a normative model purports to describe some optimizing system, to show what might or ought to happen; a descriptive model purports to describe an existing system, to predict what will actually happen. The examples discussed earlier are descriptive models. Both types of models, common in analyses of educational systems, have their uses; it is difficult to check the assumptions in a normative model, however, since they do not have to correspond to actual systems, and hence it is difficult to guard against unrealistic or meaningless cases in such models.

Finally, the mathematical techniques used to express the assumptions of the model should be appropriate for predictive purposes. The most common techniques for modeling educational systems have been the probabilistic tools of stochastic processes and the deterministic tools of difference equations and linear programming. Specifically excluded here are the statistical methods of regression or correlation analysis.

---

*This opinion is essentially based on the assumption that by dealing with aggregates, most of the local, small-scale variability will cancel out; while this is not always true, much of the success of input-output economic analysis depends upon the stability of aggregate economic quantities.

**These techniques are discussed in the context of specific examples in the Appendix.
Statistical studies are often useful for preliminary examination of large amounts of data and may lead to valuable insights or indicators. The implied predictive character of these studies is that a change in input variables will shift performance to a different position on the distribution of the output variable, but it is usually difficult to distinguish between causal and merely associative relationships. Hence such statistical studies, while they may be a useful basis on which to construct a model, do not in themselves constitute predictive models.

*See Ref. 2 for an example of such a statistical study applied to educational systems.
III. SELECTED EDUCATIONAL-SYSTEM MODELS: A GUIDE TO THE LITERATURE

Thus far we have discussed the educational-system model in terms of scope, complexity, degree of aggregation, inputs, outputs, goals, and types—conceptual and mathematical models, normative and descriptive models. In this section, we shall examine some existing models which we believe to be representative of the literature. These models are often mixtures of the various types defined in Section II, so although the above list may serve as an implicit set of dimensions for describing and differentiating among them, no attempt is made here to use these factors in establishing a formal classification scheme. Instead, we shall be concerned with the context in which each model was developed and the problem to which each is addressed. Mathematical details of the models are not included here but are presented in the context of specific examples in the Appendix.

Thonstad's mathematical model of the Norwegian educational system exemplifies a class of models that deals primarily with the flow of students through a system. The major goal of the model is to examine the long-range implications of present educational policies in terms of the numbers of students attaining various levels of education. Although wide in scope, this model is mathematically simple and explicit; it is descriptive in purpose and requires as input data on the flows of students between different levels or activities within the educational system. The model yields predictions of the numbers of students to be expected in each activity in future years, the average number of years of education remaining for a student in a given activity, and the percentage of students finishing each level of schooling.

Although Thonstad includes 60 different categories of educational activity in this model, the only phenomena modeled are the flows of students from one activity to another. No economic limitations or constraints based on school capacity, availability of teachers, or other resource limitations are included. The variables are highly aggregated, representing

*Activities include primary schools, college-preparatory secondary schools, vocational secondary schools, technical institutes, universities, and similar educational divisions.
all students in a particular activity or level of education, regardless of subject matter, student ability, or the particular school attended. Hence Thonstad's model is not a complex model in the sense defined earlier. While this severely limits the range of questions that the model can help answer, it does not necessarily imply that the predictions will be any less accurate than those of a more complex model; in fact, the reverse will generally be true, given the current level of understanding of educational systems and educational processes.

A more basic difficulty with Thonstad's model and with other input-output models, whether formulated in terms of stochastic processes, as Thonstad's is, or in terms of input-output tables like those often used in economic analysis, is that they do not model the specific educational mechanisms that cause change. Such models can be very accurate in describing a system that is essentially static or changing smoothly or slowly, but they are very inaccurate for systems where rapid changes can occur or where influences other than the demand of students for more education play a substantial role. Thonstad, for example, found that his model gave relatively good predictions for primary and secondary schools in Norway, which are part of compulsory education, yet gave much poorer predictions for higher education, where enrollments were limited by the capacity of the system, which depended in turn on administrative decisions. These comments also apply to models similar to Thonstad's, such as Zabronski and Zinter's student-teacher population growth model, and Gani's model for projecting enrollments.

The model of doctoral feedback into higher education developed by Bolt is even simpler than those discussed above, but it surmounts two of the difficulties of those models: It is reasonably accurate.
for higher education, and it includes a specific educational mechanism. Bolt's model seeks to describe the relationship between the number of degrees awarded and the numbers of persons involved in training graduate students. The model is national in scope and uses very highly aggregated variables; it requires as input information on the flows of students leaving graduate school and can be used to predict the numbers of degrees awarded and the numbers of faculty required in future years. Although the model is primarily descriptive, it is also used in a normative fashion to suggest suitable policy alternatives for arriving at planning objectives. It does not include any economic considerations and is limited to situations where the demand for more education is larger than the supply (that is, where the supply of graduate students is not input-limited). Nevertheless, it would seem to be an improvement over input-output models like Thonsted's for studying higher educational systems, both in terms of accuracy and in terms of identifying the significant policy variables, because of its use of a specific mechanism.

In contrast to these descriptive models, Stone's model of the educational system(6) is purely normative. Based on an analysis of the number of trained persons of each type needed in the economy at a future date and on the growth of demand for education due to population increases, Stone has attempted to calculate how the English educational system should change to accommodate these demands. The model is partly mathematical and partly only conceptual (that is, not yet finished) in form, and national in scope. It would require a great deal of information about future industrial manpower demands and student flows as input; the output would be a plan for the long-range development of the educational system. Despite the economic basis for this study, the model itself is relatively simple, dealing only with flows of students and excluding economic variables.

The underlying assumption in Stone's model, which is also implicit in many other manpower planning models, is that students can be induced to seek whatever kind of education is envisioned for them in the plan, through scholarships and admission policies. At the level of higher
education, this assumption is probably not very realistic. In the United States, for example, the fraction of all Ph.D.'s awarded in science and engineering has not changed appreciably during this century (except during World War I and II), despite the massive amounts of federal money that have been put into academic science since 1945, particularly since the beginning of the Sputnik era. It has been possible to increase the production of trained persons in science only as fast as the educational system could be expanded as a whole. Whether the encouragement of students to seek particular training can be more successful in areas such as vocational training, or in countries with centrally controlled economies and educational systems, is still unclear.

A similar but more elaborate planning model than Stone's is that of Tinbergen and Bos for planning the educational requirements of economic development. The goal of this model is to project the flows of trained manpower that will be required by the economic growth of the country and rapid expansion of the educational system. It is a very simplified model of the relations between production, labor force, and the educational system, using highly aggregated variables, such that all types of training and all types of labor skills are lumped together. The outputs of the model are plans for the expansion of the educational system; these results are obtained by a mathematical method that is extremely crude, particularly compared to the optimization methods used in models discussed below. (However, this in itself does not necessarily imply that Tinbergen and Bos's model is less accurate or applicable.)

This model is entirely normative and hence has no descriptive base for the assumptions it makes about student flows. For this reason, the results of the model are suspect and there is little to guarantee their applicability. (The comments made above in connection with Stone's model apply here also.) Furthermore, the method used in Tinbergen and Bos's model (and in many manpower planning models) to obtain a plan for future development of the educational system does not allow the particular plan adopted to be compared with alternative policies, nor does it provide a feeling for the tradeoffs involved in choosing a particular plan. In this respect, Tinbergen and Bos seem to have oversimplified their approach; or at least they have not chosen an appropriate simple model.
A rather different method is used by Bowles (8) to develop planning-policy proposals for the efficient allocation of resources in education. Bowles attempts to apply optimization techniques such as linear programming to educational problems. His model seeks to determine what amount of a country's resources should be devoted to educational development and how such resources should be distributed within the educational system. The model is national in scope, including students at all levels of the educational system, and also including (in a parameterized way) the labor force of the entire country. It is both a descriptive and a normative model, since the flows of students and teachers are modeled descriptively in the input-output manner, but the model as a whole produces normative proposals for the distribution of resources. It is a complex model in that it includes flows of students among parts of the educational system, the cost relations of that system to the national economy, and the economic-benefit relations of such education to the national economy. The inputs required include descriptive data for the student-flow model, as well as information on cost factors and estimates of income by educational level.

Bowles's model attacks the same problem as that of Tinbergen and Bos. However the planning projections of the latter model are of little use, as mentioned above, since they are not based on a description of the educational system as it actually operates, whereas the normative results of Bowles' model take into consideration a description of the actual system and its constraints. In this respect optimization models offer an improvement over manpower planning models, although their applicability depends upon an adequate characterization of the benefits that are to be maximized. Both Adelman's linear-programming model of educational planning: (a case study of Argentina) (9) and Schliefelbein's multiperiod linear-programming model for forecasting the quantitative results of alternative national educational policies (10) are similar in most respects to Bowles' model.

Most of the models introduced so far have been national in scope; in contrast, Koenig's systems approach to higher education (11) and Judy's model for resource allocation in universities (12) both deal with the more restricted scope of a single university. Koenig's model is a very
complex mathematical model which considers many relationships and interactions, including flows of students and faculty, economic and resource limitations, controls of student flows such as fellowship and grant money, and the demand for research supervision as well as for teaching. The model is primarily descriptive but can be used repetitively to compare the consequences of competing policies and hence to select the most desirable. The model is not applied to a specific situation, since some of the required input data are unavailable; like most models that ultimately involve computer simulation, an enormous data base is required, so that the inputs necessary to run the model are substantial.

Judy's model is similar to Koenig's in most respects. However, Judy places greater emphasis on categorizing expenses by function rather than by department, in accordance with the administrative management philosophy known as planned program budgeting (PPB). *

The wealth of practical detail in these and other simulation models is both their great advantage and their major limitation: Without a well-developed information-gathering system to provide accurate and regularly updated parameters and input data, such models are comparatively useless. Furthermore, it is important to distinguish between a wealth of bookkeeping detail and a real knowledge of the mechanisms relating inputs to outputs in a complex system, which can only come from such models after considerable confidence has been gained as to their accuracy. To make effective use of a simulation model for planning purposes, a planner must have in mind specific policies that he wants to evaluate, whereas with optimization models, normative proposals are an output of the model rather than an input. The desired goals and the specific priorities involved in accomplishing them must originate with the planner; if it does not occur to him to test alternative goals and priorities, the simulation model can be of little help.

Other models of similar scope and aggregation include Nordell's dynamic input-output model of the California educational system (13) and Weathersby's university cost simulation model. (14)

Effective models of educational systems need not be large or of

* This model has been applied to the University of Toronto.
wide scope, although the majority of the models available so far have these characteristics. Brooks, for example, has developed and applied mathematical models of a training program for automotive mechanics. (15) The model is normative, since it simulates the effects of alternative policies on a proposed system rather than describing the operation of an existing training system. The model deals with the flows of students and the numbers of instructors and training resources needed. The variables are not highly aggregated, and the model is relatively simple in the kinds of interactions it tries to represent. It yields predictions of how fast trained mechanics could be produced, although, since the parameter values assumed as input are hypothetical, the results have no import for any particular system.

Another limited-scope model is that of Fulkerson et al., (16) which is concerned with the bussing of students in Los Angeles so that existing schools may be adequately used. The model takes as input a given distribution of schools and pupils, then finds the bus routes over which a given amount of pupils can be transported in the shortest time or for the lowest cost. And finally, we have Bruno's models for optimizing various objective functions of foundation-type state support programs, (17) which deal with the design of funding programs to achieve desired statewide goals by encouraging particular policies at the district level.
IV. CONCLUDING REMARKS

The educational-system models discussed in Section III do not begin to exhaust the rapidly growing literature, but they do give an indication of the types of models that have been constructed, the variety of purposes for which models have been used, and the difficulties of adequately modeling educational systems. None of the models examined here are entirely adequate, but some have less serious faults than others and some are more useful than others in particular situations.

Input-output models such as Thonstad's can be a convenient way to examine large amounts of data on enrollments and student flows. The difficulty with such models is that a current cross-sectional analysis is generally used to predict the future time series of the variables, masking the causal mechanisms of the system. Any changes in conditions that affect the system more than marginally necessitate a change in the model parameters, and such models are limited to slowly changing situations or short time periods, if they are to provide accurate results. Nonetheless, given the complexity of educational systems and the necessity of attempting systematic planning even with sketchy data, such models may be very useful if their limitations are understood and allowed for. In particular, input-output models may find wide application in analyzing systems with relatively static structures, such as training institutions for specific purposes (Air Force technical training schools, for example).

Where educational mechanisms can be identified, such as in Bolt's model, relatively simple models can be extremely effective. The usefulness of simple models for flows of students and faculty can be extended further by including simple economic variables, such as cost factors, but without attempting to model all aspects of a system.

Manpower planning models such as Stone's or that of Tinbergen and Bos seem to be less useful than many other models, despite their popularity in Europe and in some developing countries. Because these models do not describe actual student flows, and because they do not provide explicit allocations of educational resources, they are perhaps too simplified for the problem they are attempting to solve.
Optimization models such as that of Bowles have the advantage of making explicit the basic choices present in any resource-allocation problem. Furthermore, such models yield priorities and plans as output, so they are likely to stimulate discussion and thought at the policy level; this may well be one of the most important results of modeling efforts. However, optimization models are limited to situations in which the benefits of a given educational policy can be quantitatively characterized. To date, such modeling has been performed only in terms of national economics. It would seem useful to attempt similar models using direct measures of student achievement.

It seems clear that simulation models such as those of Judy and of Koenig can be improved as more data on a particular system become available, and that these models will be of considerable assistance in management and short-term planning of education systems. The danger to be avoided is that of foundering in the wealth of detail, both in verifying a model, and in applying its results. It may well be that optimization models and complex simulation models can serve in complimentary ways in educational planning, one operating on the policy level and the other on the detailed operational level.

Finally, a variety of simple models may be useful tools in solving subsidiary problems of all kinds in educational systems. This does not mean that a model is always the preferable tool, but rather, that model building often encourages more systematic examination of relationships and assumptions and enables more accurate estimates to be made.

Further research is urgently needed to increase our understanding of the dynamics of educational systems; on the basis of such additional knowledge, it would be possible to model systems in increasing detail, with some confidence. Attempts to model complex systems at the present time have, in fact, had an important impact as an aid to data collection, in pointing out what kinds of data are needed, and in encouraging a systematic approach to data collection. It was not, for example, until modeling efforts began that the importance of detailed information on dropout and failure rates was recognized, and universities are only now beginning to collect such information.

The appropriate mathematical basis for most of the research suggested here would be very simple; stochastic models for probabilistic
problems and simple difference and differential equations for deterministic models, coupled with optimization techniques where appropriate, should be adequate for most modeling of educational systems in the near future. Where great detail, or ease of data manipulation is desired, computer simulation offers a convenient tool for implementing and supplementing the above techniques. This point of view--that the present level of mathematical complexity is more than adequate if appropriately used--is complementary to our belief that relatively simple models are the best tools for attacking complex problems whose quantitative relationships are poorly understood.
Appendix

MATHEMATICAL STRUCTURE OF EDUCATIONAL-SYSTEM MODELS

The mathematical structure of the models examined has been referred to only briefly in the text of this Memorandum, although this is an important element in their construction and use. In this appendix the mathematical structure and mathematical tools commonly used in models of educational systems are discussed in more detail, with emphasis on the implications of particular tools for the type of phenomenon being modeled. We have assumed here that the reader has some familiarity with simple stochastic processes (especially Markov chains), first-order difference-equation systems, and optimization methods such as linear programming.

To an outside observer, the movement of people through an educational system may well seem stochastic in character. It is often assumed that the probability of movement from one educational activity to another depends only on the present activity of the student, and that the future state of the system may thus be predicted by knowledge of the present state and the transition probabilities; this assumption characterizes the Markov process.

Thonstad's model\(^{(3)}\) is developed as a discrete Markov chain, such that the constant probability that a student in activity \(i\) at time \(t\) will be in activity \(j\) at time \(t + 1\) is given by \(c_{ij}\), thus defining a transition matrix:

\[
C = \begin{bmatrix}
    c_{11} & \cdots & c_{1n} \\
    \vdots & & \vdots \\
    c_{i1} & \cdots & c_{in} \\
    \vdots & & \vdots \\
    c_{n1} & \cdots & c_{nn}
\end{bmatrix}
\]  

(1)

The elements \(c_{ij}\) are nonnegative (since they are probabilities), and the row sums must equal unity \(\sum_{i=1}^{N} c_{ij} = 1\), which can be viewed as expressing conservation of people (i.e., a student must be in one of the states of the system). The Markov assumption embodied in Eq. (1)--that the future state of the system depends only on the state of the
system in the most recent time period—means that the expected number of students leaving a given activity for another \( (P_i \cdot c_{ij}) \) is a constant fraction \( (c_{ij}) \) of the number of students in the original activity \( (P_i) \); this is equivalent to the input-output method of forecasting, which assumes that the flows of interest are proportional to the relevant stock variables, such as the numbers of students.

A Markov chain has the property that the state vector (in this case the number of students in each activity) approaches an equilibrium value, so that a purely stochastic model is inappropriate for modeling a growth process such as an expanding educational system. Thonstad avoids this difficulty by using a deterministic approach in the application of his model, interpreting the elements of Eq. (1) as fixed transition ratios rather than probabilities, so that he arrives at a set of difference equations for the number of students \( P_s \) in activity \( s \),

\[
P_s(t) = \sum_{h=1}^{N} q_{hs} P_h(t-1) + Y_s(t), \quad s = 1 \ldots N, \quad (2)
\]

where \( q_{hs} \) are the transition ratios and \( Y_s(t) \) represents the new entrants from outside the system in year \( t \). Equation (2) effectively predicts the expected value of the state vector \( P \) due to the stochastic process modeled by the transition ratios \( q_{hs} \) and the forcing function \( Y \) of new enrollments (assumed known empirically).

Most deterministic difference-equation models of educational systems have the form of Eq. (2); that is, they are linear, first-order, constant-coefficient difference-equation systems. Bolt's model, \(^{(1)}\) for example, reduces to a two-equation system of this form without a forcing function; and the model of Tinbergen and Bos \(^{(7)}\) is a six-equation system with a forcing function, which in this case is the total volume of production of a country.

Linearity means that \( P_s(t) \) is a function of \( P_s(t-1) \), rather than \( [P_s(t-1)]^2 \) or \( [P_s(t-1)]^{2.5} \), or whatever. Since we do not know the functional dependence (we know no "physics" of education from which to derive the mechanisms of educational systems), we have no basis for choosing a model any more complicated than a linear one. However, linear
first-order difference systems, like first-order differential systems, have the property of either growing exponentially without limit or decaying to zero (or to purely forced behavior if there is a forcing function) after a sufficient period of time, so that linear models may lead to substantial error if applied over too long a time period. Nonlinear systems and real systems of all kinds often have the property of limiting their growth, and in general they display more complex behavior than linear systems. Except where specific nonlinear mechanisms can be identified, however, linear models are adequate if only small changes are considered.

First-order models of the movement of students through an educational system neglect the past history of the students, a simplification usually necessary because of the lack of detailed information. Assumption of a first-order model is implicit in a Markov chain and in input-output tables, which are sometimes used to describe the flow of students or resources in a system; in both cases, the use of constant coefficients means that a current cross-sectional analysis of system operation is used to predict the future time series of the variables. This difficulty can be overcome somewhat when the model is used in connection with information systems, as is the case in the models proposed by Koenig et al. (11) and Judy, (12) so that the coefficients are updated every year.

Bowles' model (6) uses the simplest form of optimization theory, linear programming. A benefit function (describing the economic benefits of education) is formulated in terms of the present value of the estimated stream of lifetime earnings, \( Y_j \), associated with a level of education, \( j \), the present value of the foregone earnings with (lower) educational level, \( j' \), and the present value of the costs to the society of that education, \( C_j \). The benefit function that Bowles uses is

\[
Z = \sum_{j} \sum_{p} (Y_{j} - Y_{j}' - C_{j})^p \cdot X_{j}^p, \quad \text{all } j \text{ and all } p, \quad (3)
\]

where \( X_{j}^p \) is the planned enrollment in year \( p \) in level \( j \). This function is maximized over the planning period, subject to constraints on the values of the \( X_{j}^p \). The constraints are of the form

\[
\sum_{j} X_{j}^p = \text{total enrollment}
\]
\[ x_{ij}^p = \min_{t,i} \left( \frac{x_{ij}^t}{a_{ij}^t} \right), \quad \text{for all } t \text{ and } i, \quad (4) \]

where \( x_{ij}^t \) is the amount of input \( i \) devoted to activity \( j \) in year \( t \), and \( a_{ij}^t \) is the minimum amount of input \( i \) required to maintain one student in activity \( j \) in year \( t \). Equation (4) requires that the planned enrollments do not exceed the resources allocated to the educational system. Once policies concerning \( x_{ij}^t \) are adopted and information regarding the other parameters is available, the programming problem can be solved to give the "optimal" planned enrollments, \( x_{ij}^p \).

Just how optimal the results of a programming model are depends to a great extent on the benefit function adopted. In educational systems, it is particularly difficult to find quantitative measures of desired benefits for which data are available, and most models so far have made use of indirect measures (e.g., lifetime earnings as a measure of the benefits of education). These difficulties are balanced by the advantages of explicitly modeling the tradeoffs among competing needs in a resource-allocation problem by optimization methods.

The mathematical techniques discussed here are those that have been most widely used to date in published models of educational systems; they are convenient and suitable tools for many modeling situations, particularly those involving simple models. A host of additional techniques exist that could also be applied to modeling educational systems; thus this appendix is by no means exhaustive and should be viewed only as a starting point for modeling efforts.
REFERENCES


BIBLIOGRAPHY


Koulourianos, D., Educational Planning for Economic Growth, Center for Research in Management Science, University of California, Berkeley, TR #23, 1967.


Schliefelbein, Ernesto F., "Design of a Multiperiod Linear Programming Model for Forecasting the Quantitative Results of Alternative National Educational Policies with Special Application to the Chilean Case," Doctoral Qualifying Paper, Harvard University, 1967.


Stanford School Scheduling System--S4, School Planning Laboratory, School of Education, Stanford University, Stanford, California.


