I. PHILOSOPHICAL APPROACH

A. Problem Definition

Stated in its simplest form, the question addressed here is how to allocate instructional resources to achieve a desired objective. Broadly interpreted, this question could include the total educational resources of society and all possible learning situations. In practical terms, however, the setting is restricted to the structural educational system, because this is the only context in which decisions on the allocation of instructional resources may be implemented.

When the question of allocating resources is examined in this setting, attention is usually focused on a well-defined subcomponent of the problem. Once the characteristics of one of these subcomponents are understood, their implications may be extended to a larger context. However, in general, the characteristics of many subcomponents must be synthesized before solutions can be derived for the problem of resource allocation.

In the school setting, the principal resources to be allocated are the human resources of teachers and students. When the teaching function is augmented by nonhuman resources, such as computer-aided instruction, then the total instructional resources must be considered. The time spent by the students also must be included because there is frequently a trade-off between instructional resources to be allocated and speed of learning.
There are two basic questions in any resource allocation problem: (a) what are the alternatives and their implications, and (b) which alternative is preferred? The first question concerns the "system" and includes such questions as what is feasible, what happens if, and what is the cost? The second question has to do with the goals, objectives, and preferences of the decision-maker or the collection of people he represents. These are very difficult questions to answer; but they must be answered, at least implicitly, every time an allocation decision is made. In this chapter we review the development and application of mathematical models that help the decision-maker directly with the first question and indirectly with the question of identifying objectives and preferences.

B. Empirical Approach versus Modeling Approach

The core of any decision problem is the determination of the implications or outcomes of each alternative, that is, the determination of the answers to what happens if? The questions of feasibility and cost are ancillary to this central problem and are relatively uncomplicated. For example, consider the problem of determining optimal class size. For a particular situation, the question of feasibility might involve simply the availability of physical facilities and instructional resources. Analysis of the question of cost also would be reasonably straightforward. However, it would be very difficult to determine and quantify the expected results with sufficient accuracy to permit assessment of the cost-effective trade-off. It is the quantitative analysis of the core of the decision problem that can be approached with empirical or modeling techniques.

In the empirical approach, the input variables (class size, for example) and the output variables (amount learned, for example) are defined for the particular problem at hand, and then empirical data relating to these variables are collected and analyzed. From the analysis, it is hoped that a causal relationship can be determined and quantified. This relationship then serves to predict the output from the system for the range of alternatives under consideration. Once the expected output has been quantified and once the costs of the alternatives have been determined, the decision problem is reduced to an evaluation of preferences.

The empirical approach has a natural appeal for several reasons. First, perhaps, is its simplicity. If a particular system has only a few variables that are amenable to quantification, then, given sufficient data, the relationships between them can be determined. The second reason for its appeal is that no a priori knowledge of the relationships among variables is necessary; the data simply speak for themselves. A third reason is that data analysis can never really be avoided completely, whatever approach is employed. Thus, if the problems of data collection, verification, and analysis must be encountered regardless, it may appear expeditious to rely on data analysis alone.

There are, however, many problems with the application of the empirical approach, especially to situations that are as complicated as those that comprise the educational system. It is extremely difficult to formulate the system within the framework of control theory. Often surrogate variables must be used that are suitably quantified. For example, teaching effectiveness can be expressed in terms of student learning, years-of-experience, etc. These problems involve statistical sampling of survey and interview techniques.

In addition to definitional and measurement problems, controlling multiple variables and long time constants in a system of many variables, the relationships are impossible to extract empirically because of the presence of unquantifiable variables. Moreover, the fact that some time constants introduce complications in time series analysis is required. Time series or "longitudinal" analysis is required when the objective is to study the development of the system, whether it be an experimental or a naturally occurring system. For example, in education, the effect of the social environment on the development of the student is extremely important. The measurement of the effect of the social environment is critical to determining the expected output of any given system, whether it be an experimental or a naturally occurring system.

The second method of analyzing the system is formulated within the framework of control theory. Mathemati-
in any resource allocation problem: (a) what are the alternatives, and (b) which alternative is preferred? The first question includes such questions as what is the purpose of the decision-maker or the collection of data? These are very difficult questions to answer, but implicitly, every time an allocation decision is made, the development and application of mathematical techniques must be faced. The second question has to do with the development of a behavioral framework for the decision-maker directly with the first question and identifying objectives and preferences.

C. Mathematical Models and Optimization Theory

In its most abstract form, the modeling approach offers the power of mathematical analysis with the capability of examining a wide range of alternatives or parameter values. The models that result from fitting mathematical equations to empirical data also may be amenable to mathematical analysis; but often, because of their complexity, they require the power of computers to analyze the effects of various alternatives and parameter values. It is, of course, possible to combine the abstract model form with extensive data analysis. Indeed, the optimal balance of model abstraction and data analysis is the goal of any model builder. This balance depends upon many factors, including the purpose of the model, the availability of appropriate data, and the characteristics of the decision-maker as well as the analyst. A good model is characterized by providing sufficient detail for the decision-maker while retaining no more complexity than is required to portray adequately relationships within the real environment.
of the fundamental variables of the system. For an alternative under consideration, the model determines all the implications or outcomes over time resulting from the implementation of that particular alternative or policy.

Once the implications of each alternative are known and the costs have been evaluated, preferences can be assigned to the various alternatives. In the framework of control and optimization theory, these alternatives for resource allocation are associated with settings of the control variables. The preferences over all possible alternatives are specified by an objective function that measures the trade-off between benefits and costs, which are defined in the model by the values of the control variables and the state variables. The control and state variables define, generally speaking, the inputs and outcomes of a system, respectively. The problem of optimal resource allocation is thus the problem of choosing feasible control variable settings that maximize (or minimize) the objective function.

The central dynamic behavior that must be modeled when considering problems of resource allocation in the educational setting is the interaction between the instructor—whether it be teacher, computer-assisted instruction or programmed instruction—and the individual learner. The effects of the environment (for example, the classroom) also are important. Models of these interactions are essential in order to predict the outcomes of alternative instructional policies. Once the cost components of the various alternatives have been evaluated, the optimization problem may take one of three forms. If the quantity of resources is fixed, then benefits can be maximized subject to this resource constraint. If there is a minimum level of performance to be achieved, then the appropriate objective is to minimize cost subject to this performance level. Finally, if performance and cost are both flexible and if the trade-off of benefit and cost can be quantified in an objective function, then both the optimal quantity of resources and the level of performance can be determined.

II. PREVIOUS RESEARCH

A. Overview

The applications of learning models and optimization theory to problems of instruction fall into two categories: (a) individual learner oriented and (b) group of learners (classroom) oriented. In category a applications, instruction is given to one learner completely independently of other learners. These applications are typical of computer-assisted instruction and programmed instruction and also include the one-teacher/one-student situation. Within this category, many situations can be adequately described by an appropriate existing model from mathematical-learning theory. In such cases, as outlined below, the results of applying mathematical models have been encouraging. In other more complex situations, existing models must be modified to describe the instructor/learner interaction.

In category b applications, instruction is given to learners. This characteristic is typical of classroom situations that includes other forms of instruction, such as full-time classroom teaching. In these situations, more learners may be receiving instruction being from one instructor. In contrast to category a situations, existing models of instruction are not sufficient for resource allocation. A comparable theory for the group of learners must therefore include model development and analysis.

Most applications, whether in category a or b, follow a common course of action.

Step 1 is to isolate a particular learning situation. This step is classified as category a or b, the nature of the material to be learned is specified.

Step 2 is to acquire a suitable model for the learning. This step may be as simple as the set of procedures known from mathematical-learning theory, as mentioned in the previous section, or it may require development of a new model for the particular situation.

Step 3 is to define an appropriate criterion for evalu­ating solution possibilities, taking account of benefits and costs of the resource allocation.

Step 4 is to perform the optimization analysis to find the optimal solution. These characteristics may involve solution to key variables of the model and the results are compared to those of other solutions. In some situations the optimization problem is difficult or impossible to solve. In this case, variables are identified whose results represent improvements or solutions.

B. Individual Learner Setting

1. Quantitative Approach for Automatic Devices

An important application of mathematical learning theory was Smallwood’s (1962) development of teaching machines. Smallwood’s goal was to produce teaching machines that would emulate the two-tutor human tutor: (a) the ability to adjust instruction according to student response history, not only to the benefit of the student and (b) the ability to adapt instruction based on the student’s response history within this framework must be able to determine the best decision for each student in the instructional process.
the system. For an alternative under consideration, the implications or outcomes over time result in particular alternative or policy.

Alternative are known and the costs have been assigned to the various alternatives. In the optimization theory, these alternatives for resource settings of the control variables. The preferences, specified by an objective function that measures costs, which are defined in the model by the and the state variables. The control and state settings, the inputs and outcomes of a system, optimal resource allocation is thus the problem of settings that maximize (or minimize) the objectives.

that must be modeled when considering problems for educational setting is the interaction between teacher, computer-assisted instruction or individual learner. The effects of the environment are important. Models of these interactions are outcomes of alternative instructional policies.

various alternatives have been evaluated, the one of three forms. If the quantity of resources is optimized subject to this resource constraint. If there to be achieved, then the appropriate objective performance level. Finally, if performance and trade-off of benefit and cost can be quantified in the optimal quantity of resources and the level of situations, existing models must be modified or new models must be developed to describe the instructor/learner interaction.

In category b applications, instruction is given simultaneously to two or more learners. This characteristic is typical of classroom-oriented instruction and also includes other forms of instruction, such as films and mass media, where two or more learners may be receiving instruction but there is no feedback from learner to instructor. In contrast to category a situations, where mathematical-learning theory provides suitable models of instructor/learner interaction, there is no comparable theory for the group of learners environment. Category b applications must therefore include model development as well as mathematical analysis.

Most applications, whether in category a or b, follow a 4-step procedure.

Step 1 is to isolate a particular learning situation. In this step, the learning situation is classified as category a or b, the method of instruction is defined, and the material to be learned is specified.

Step 2 is to acquire a suitable model to describe how instruction affects learning. This step may be as simple as the selection of an appropriate model from mathematical-learning theory, as mentioned above, or as difficult as the development of a new model for the particular situation.

Step 3 is to define an appropriate criterion for comparing the various instruction possibilities, taking account of benefits and costs as determined by the model.

Step 4 is to perform the optimization and analyze the characteristics of the optimal solution. These characteristics may include the sensitivity of the optimal solution to key variables of the model and the comparison of its results relative to those of other solutions. In some situations the optimization problem may be very difficult or impossible to solve. In this case, various suboptimal solutions may be identified whose results represent improvements over those of previous solutions.

B. Individual Learner Setting

1. Quantitative Approach for Automated Teaching Devices

An important application of mathematical modeling and optimization theory was Smallwood’s (1962) development of a decision structure for teaching machines. Smallwood’s goal was to produce a framework for the design of teaching machines that would emulate the two most important qualities of a good human tutor: (a) the ability to adjust instruction to the advantage of the learner and (b) the ability to adapt instruction based on the learner’s own experience. The decision system within this framework must therefore make use of the learner’s response history, not only to the benefit of the current learner, but also for future learners.
The learning situation considered by Smallwood has three basic elements: (a) an ordered set of concepts that are to be taught, (b) a set of test questions for each concept to measure the learner's understanding, and (c) an array of blocks of material that may be presented to teach the concepts. Two additional elements are required to complete the framework for the design of a teaching machine: (d) a model with which to estimate the probability that a learner with a particular concept to measure the learner's understanding, and (e) a criterion for choosing which block to present to a learner at any given time.

Having defined his model requirements in probabilistic terms, Smallwood (1962) considered three modeling approaches: correlation, Bayesian, and intuition. He discarded the correlation model approach as not useful in this context. Then he developed Bayesian models, based on the techniques of maximum likelihood and Bayesian estimation (these models are too complex to review here). His intuition approach led to a relatively simple quantitative model based on four desired properties: representation of question difficulty and learner ability, together with model simplicity and experimental performance. The model is

\[ P = \begin{cases} \frac{bc}{a} & b \leq a \\ \frac{1 - (1-b)(1-c)}{(1-a)} & b > a \end{cases} \]

where \( P \) is the probability of a correct response, \( b \) measures the ability of the learner, \( c \) measures (inversely) the difficulty of the question, and \( a \) is an average of the fraction of correct responses. All parameters are between zero and one.

For a particular set of learners and a given set of questions, the parameter \( a \) is fixed. For an average learner \((b = a)\), this model equates the probability of a correct response for a question of difficulty \( c \) to \( c \) itself. For less than average learners \((b < a)\), the probability of a correct response varies linearly with \( c \) but is uniformly lower than for the average learner. Similarly, for better than average learners \((b > a)\), this relationship is again linear but higher than that for the average learner.

As an objective function for determining optimal block presentation strategies, Smallwood (1962) suggested two possibilities with variations. One was an amount-learned criterion, which measured the difference before and after instruction, and the other was a learning-rate criterion, which essentially normalized the first criterion over time. In the optimization process, these criteria are used to choose among alternative blocks for presentation in a local, rather than global sense.

At any branch point in the presentational strategy where more than one block or set of blocks could be presented to the learner, the learner's response history is used to calculate a current estimate for the parameter \( b \). The other parameters are estimated previously making use of all available learner-response histories. Each alternative branch from this point can then be evaluated using the previously mentioned criteria and the best branch for instruction.

A simple teaching machine was constructed to test the decision structure. The experimental evidence distinguished between learners and presented the best blocks of material. It also verified that differences under similar circumstances, indicated.

2. Order of Presentation of Items from a List

The task of learning a list of paired-associates appears in many areas of education, notably in reading (Atkinson, 1972). It is also a learning task for which learning theory has been very successful. Therefore not surprising that the earliest application of optimization techniques have led to finding learning models employed in these studies. The value of a particular criterion is evaluated for three reasons: (a) the application of further critical assessment of the basic learning model, (b) the analytical procedure is transferable to many areas of education, notably in reading (Atkinson, 1972). It is also a learning task for which learning theory has been very successful. Therefore not surprising that the earliest application of optimization techniques have led to finding learning models employed in these studies. The value of a particular criterion is evaluated for three reasons: (a) the application of further critical assessment of the basic learning model, (b) the analytical procedure is transferable to many areas of education, notably in reading (Atkinson, 1972). It is also a learning task for which learning theory has been very successful. Therefore not surprising that the earliest application of optimization techniques have led to finding learning models employed in these studies. The value of a particular criterion is evaluated for three reasons: (a) the application of further critical assessment of the basic learning model, (b) the analytical procedure is transferable to many areas of education, notably in reading (Atkinson, 1972).
alternative branch from this point can then be evaluated using one of the above mentioned criteria and the best branch for immediate gain is chosen for presentation.

A simple teaching machine was constructed based on the concepts of this decision structure. The experimental evidence verified that the machine distinguished between learners and presented them with different combinations of blocks of material. It also verified that different decisions were taken at different times under similar circumstances, indicating that the machine was adaptive.

2. Order of Presentation of Items from a List

The task of learning a list of paired-associate items has practical applications in many areas of education, notably in reading and foreign language instruction (Atkinson, 1972). It is also a learning task for which models of mathematical-learning theory have been very successful at describing empirical data. It is therefore not surprising that the earliest and most encouraging results of the application of optimization techniques have come in this area. Although the learning models employed in these studies are extremely simple, the results are valuable for three reasons: (a) the applications are practical, (b) these results lead to further critical assessment of the basic learning models, and (c) the general analytical procedure is transferable to more complex situations.

The application of mathematical models and optimization theory to the problem of presenting items from a list can be illustrated by three examples from the literature. The first is a short paper by Crothers (1965) that derives an optimal presentation criterion, which essentially normalized the difference before and after instruction, and thus essentially maximized the optimization process. These criteria are used to determine optimal block presentation strategies.

In the second example, a model approach as not useful in this context. Models, based on the techniques of maximum likelihood (these models are too complex to review here), are used to determine the order of presentation. This model equates the probability of a correct response, $b$, to the difficulty of the question, $a$, and $c$ is an average of the parameters $a$ and $b$. All parameters are between zero and one. For a given set of questions, the parameter $a$ is $a = \frac{b - a}{1 - a}$, and $c$ is the difficulty of the question, and $a$ is the average of $b$ and $c$. For less than average difficulty, a correct response varies linearly with $c$ but is less than average difficulty. Similarly, for better than average difficulty, a correct response is again linear but higher than that for the average difficulty.

In the third example, a computational strategy where more than one block is presented to a learner, the learner’s response history is used to determine optimal presentation in a local, rather than global, sense. The parameters $a$, $b$, and $c$ are measures of the ability of the learner, the learner’s response history, and the difficulty of the question, respectively. The expected proportion of correct responses is therefore linear in the parameter $b$. The other parameters are obtained from all available learner-response histories. Each alternative branch from this point can then be evaluated using one of the above mentioned criteria and the best branch for immediate gain is chosen for presentation.
of learning, and his transition from state-to-state is defined by the probabilistic transition matrix

\[
\begin{pmatrix}
1 & 0 & 0 \\
a & 1-a & 0 \\
b & c & 1-b-c
\end{pmatrix}
\]

This model simplifies into the two-element model by setting \( b \) equal to 0 and further into the all-or-none model by dropping the partial-learning state. This model is assumed to describe the learning process for each mode of presentation, so that the response probabilities for each state are identical for all modes, but the parameters \( a, b, \) and \( c \) are different for each mode. For a discussion of these models, see Atkinson, Bower, and Crothers (1965).

The result of the optimization step in this application is contained in two theorems. The first theorem states that the ranking of presentation schedules based on the expected proportion of correct responses (which is the defined objective) is identical to the ranking based on the probability of occupying the learned state. The second theorem states that the ranking of two presentation schedules is preserved if the schedules are either prefixed or suffixed by identical strings of presentations. These theorems are sufficient to conclude that moving one presentation mode to the right of another in a schedule always has the same (qualitative) effect on the terminal proportion correct and, hence, that optimal presentation schedules have all presentations of one mode together.

In the learning situation described by Karush and Dear (1966), there are \( n \) items of equal difficulty to be learned, and the problem is to determine which item out of the \( n \) to present for study at any given time. The strategy for choosing items for presentation is to take into account the learner’s response history up to the current time. The all-or-none model is used to describe the learning process, and it is assumed that the single model parameter has the same value for each item.

In order to formulate an objective function, it is assumed that all presentational strategies have the same cost so that the objective can be defined in terms of the state of learning at the termination of the strategy. Assuming that all items are weighted equally, an expected loss function is defined in terms of the probabilities \( P_k \) that at the terminal node exactly \( k \) items are still unlearned. The expected loss for a particular terminal node is given by

\[
\sum_{k=0}^{n} P_k b_k
\]

where \( b_k \) is the value (weight) of the loss if \( k \) items are still unlearned. The overall expected loss, which is to be minimized, is therefore

\[
\sum_{k} q(h) \sum_{k} P_k(h) b_k
\]

where \( q(h) \) is the probability of occupying terminal \( h \) over all possible terminal nodes. For the objective function above is equivalent to the all items are learned; and for \( b_k = k \) it is the expected sum of the probabilities of being in \( h \) of the results that are derived in the paper are \( b_k \), and so they are quite general.

The optimization is accomplished using the linear programming. The principal result is that, for all the learned state for each item, an optimal strategy for which the current probability of learning an application of the results is for the case when in which case the optimal strategy can be accounted for correct and incorrect responses of the optimal strategy is independent of both the transition and the probability of guessing.

Atkinson and Paulson (1972) reported empirical results that a none-based optimal strategy derived by Karush and Dear (1966) with strategies based on other learning models. The none-based strategy is compared with the optimal model. In the derivation of this latter optimal model parameters are identical for all items. The expected number of correct responses at the expected number of correct responses at the terminal node exactly \( k \) items are still unlearned.

In another experiment, the all-or-none-based strategy are compared with a strategy based on the RTI model. The RTI model is a compromise between the all-or-none and the linear model. Defined in terms of the probability that the learning experience for the all-or-none strategy is emphasizing those items that are performance on the postexperiment test for this strategy confirms that for this particular objective learned rather than learned items during the learning strategy is independent of both the transition and the probability of guessing.
8. Optimization Theory

The optimization is accomplished using the recursive formulation of dynamic programming. The principal result is that, for arbitrary initial probabilities of being in the learned state for each item, an optimal strategy is to present the item for which the current probability of learning is the least. The most practical application of the results is for the case where these initial probabilities are zero, in which case the optimal strategy can be implemented simply by maintaining counts of correct and incorrect responses on each item. Also in this case, the optimal strategy is independent of both model parameters: the probability of transition and the probability of guessing.

Atkinson and Paulson (1972) reported empirical results employing the all-or-none-based optimal strategy derived by Karush and Dear (1966) and compared it with strategies based on other learning models. In one experiment, the all-or-none-based strategy is compared with the optimal strategy derived from the linear model. In the derivation of this latter optimal strategy, it is assumed that the model parameters are identical for all items. For the objective of maximizing the expected number of correct responses at the termination of the experiment, it is shown that all items should be presented the same number of times. Consequently, a random-order strategy is employed in which all items are presented once, then randomly reordered for the next presentation and so on. The experimental results show that during the learning experience the all-or-none-based strategy produces a lower proportion of correct responses than the linear-based (random) strategy, but that on two separate postexperiment tests, the all-or-none-based strategy yields a higher proportion of correct responses. From these results it can be concluded that the lower proportion of correct responses during the learning experience for the all-or-none-based strategy indicates that this strategy is emphasizing those items that are not yet learned. The superior performance on the postexperiment test for this strategy relative to the linear-based strategy confirms that for this particular objective it is preferable to stress unlearned rather than learned items during the learning experience. In this sense, it can be concluded that in this learning situation and for the stated objective, the all-or-none model is superior to the linear model.

In another experiment, the all-or-none-based strategy and the linear-based strategy are compared with a strategy based on the random trial increment (RTI) model. The RTI model is a compromise between the all-or-none and the linear models. Defined in terms of the probability $p$ of an error response, at trial $n$ this
probability changes from \( p(n) \) to \( p(n + 1) \) according to

\[
p(n + 1) = \begin{cases} 
  p(n) \text{ with probability } 1 - c \\
  a p(n) \text{ with probability } c 
\end{cases}
\]

where \( a \) is a parameter between 0 and 1, and \( c \) is a parameter that measures the probability that an "increment" of learning takes place on any trial. This model reduces to the all-or-none model if \( a = 0 \) or to the linear model if \( c = 1 \).

This application of the RTI model differs in two ways from the earlier studies outlined above. First, because of the complexity of the optimization problem, only an approximation to the optimal strategy is used. The items to be presented at any particular session are chosen to maximize the gain on that session only, rather than to analyze all possible future occurrences in the learning encounter. Second, the parameters of the model are not assumed to be the same for all times. These parameters are estimated in a sequential manner, as described in the Atkinson and Paulson paper; as the experiment progresses and more data become available regarding the relative difficulty of learning each item, refined estimates of the parameter values are calculated.

The results of the experiment show that the RTI-based strategy produces a higher proportion of correct responses on posttests than either the all-or-none-based or linear-based strategies. The favorable results are due partly to the more complex model and partly to the parameter differences for each item. This conclusion is supported by the fact that the relative performance of the RTI-based strategy improves with successive groups of learners as better estimates of the item-related parameters are calculated.

3. Interrelated Learning Material

In many learning environments, the amount of material that has been mastered in one area of study affects the learning rate in another distinct but related area, for example, the curriculum subjects of mathematics and engineering. In situations such as this, the material in two related areas may be equally important, and the problem is to allocate instructional resources in such a way that the maximum amount is learned in both areas. In other situations, the material in one area may be a prerequisite for learning in another rather than a goal in itself. Here, even though the objective may be to maximize the amount of material learned in just one area, it may be advantageous in the long run to allocate some instructional resources to the related area. This problem of allocating instructional effort to interrelated areas of learning has been studied by Chant and Atkinson (1973). In this application, a mathematical model of the learning process did not exist, and so one had to be developed before optimization theory could be applied.

The learning experience from which the model was developed was a computer-assisted instructional program for teaching reading (Atkinson, 1974). This program involved two basic interrelated areas (called strands) of reading, one devoted to instruction in sight-word identification and the other to instruction in phonics. It has been observed that the instantaneous learning rate depended on the student's position on the other strand.

In the development of the learning model, information of the two strands was such that the instantaneous learning rate in each strand is a function of the difference in achievement levels on the two strands at time \( t \) and the instructional effort allocated to strand one, expressed in differential equation form as

\[
\begin{align*}
\dot{x}_1(t) &= u(t)f_1(x_1(t) - x_2(t)) \\
\dot{x}_2(t) &= [1 - u(t)]f_2(x_1(t) - x_2(t))
\end{align*}
\]

where \( f_1 \) and \( f_2 \) are the learning-rate characteristics.

In this formulation of the problem, the total time interval is fixed, and the objective is to maximize a weighted sum of the achievement levels on the two strands at the termination of the encounter, to maximize

\[
c_1 x_1(T) + c_2 x_2(T)
\]

where \( c_1 \) and \( c_2 \) are given nonnegative weights, subject to the constraint \( 0 \leq u(t) \leq 1 \) for all \( t \).
HARD C. ATKINSON

\[ J(n + 1) \] according to
\[ \sigma(n) \text{ with probability } 1 - c \]
\[ \sigma_p(n) \text{ with probability } c \]
and 1, and \( c \) is a parameter that measures the
if learning takes place on any trial. This model
if \( a = 0 \) or to the linear model if \( c = 1 \).
Del differs in two ways from the earlier studies
the complexity of the optimization problem,
also strategy is used. The items to be presented
then to maximize the gain on that session only,
future occurrences in the learning encounter.
Del are not assumed to be the same for all times.
in a sequential manner, as described in the
experiment progresses and more data become
ficulty of learning each item, refined estimates
ated.
show that the RTI-based strategy produces a
ences on posttests than either the all-or-none-
the favorable results are due partly to the more
parameter differences for each item. This
that the relative performance of the RTI-based
groups of learners as better estimates of the
ated.
Material
the amount of material that has been mastered
arning rate in another distinct but related area,
cts of mathematics and engineering. In situ-
do related areas may be equally important, and
nal resources in such a way that the maximum
other situations, the material in one area may
nter rather than a goal in itself. Here, even
imize the amount of material learned in just
in the long run to allocate some instructional
problem of allocating instructional effort to
been studied by Chant and Atkinson (1973). In
model of the learning process did not exist, and
imization theory could be applied.
which the model was developed was a
ogram for teaching reading (Atkinson, 1974).
interrelated areas (called strands) of reading,
word identification and the other to instruction
in phonics. It has been observed that the instantaneous learning rate on one strand
 depended on the student’s position on the other strand.
In the development of the learning model, it was assumed that the interdepen-
dence of the two strands was such that the instantaneous learning rate on either
strand is a function of the difference in achievement levels on both strands.
Typical learning-rate characteristics are shown in Figure 1. If the achievement
levels on the two strands at time \( t \) are represented by \( x_1(t) \) and \( x_2(t) \), then the
instantaneous learning rates are the derivatives of \( x_1 \) and \( x_2 \) with respect to time;
these rates are denoted as \( \dot{x}_1 \) and \( \dot{x}_2 \). By defining \( u(t) \) as the relative amount of
instructional effort allocated to strand one, the model of learning can be ex-
pressed in differential equation form as

\[
\dot{x}_1(t) = u(t)f_1(x_1(t) - x_2(t)),
\]
\[
\dot{x}_2(t) = [1 - u(t)]f_2(x_1(t) - x_2(t)),
\]

where \( f_1 \) and \( f_2 \) are the learning-rate characteristic functions depicted in Figure 1.
In this formulation of the problem, the total time, \( T \), of the learning encounter
is fixed, and the objective is to maximize a weighted sum of the achievement levels
on the two strands at the termination of the encounter. The objective is therefore
to maximize

\[
c_1x_1(T) + c_2x_2(T),
\]
where \( c_1 \) and \( c_2 \) are given nonnegative weights. This maximization is with respect
to \( u \) subject to the constraint \( 0 \leq u(t) \leq 1 \) for all \( t \) such that \( 0 \leq t \leq T \).
The optimization is carried out, not for the nonlinear learning-rate characteristic functions of Figure 1, but for linearized approximations to them. From the form of the optimal solutions, it is clear that the analysis applies equally well to the nonlinear functions. The optimization is performed by means of the Pontryagin Maximum Principle. It is shown that the optimal solution is characterized by a "turnpike" path in the $x_1x_2$ plane. On the turnpike path the difference $x_1 - x_2$ between the achievement levels on the two strands remains constant. Optimal trajectories are such that initially all of the instructional effort is allocated to one of the strands until the turnpike path is reached. Then the instructional effort is apportioned so as to maintain a constant difference between strands, that is, so as to remain on the turnpike path. Near the end of the learning encounter, the instructional effort is again allocated to just one strand, depending on the relative values of the weights $c_1$ and $c_2$ of the objective function. Figure 2 shows the turnpike path and typical optimal trajectories starting from two different initial points and terminating according to two different values of objective function weights.

It is also shown that of all the stable paths, the turnpike path is the one on which the average learning rate is maximized. A stable path is the steady state path that is approached if the relative allocation of strands is held constant. It can be shown that the difference between achievement levels on the

C. Group of Learners Setting

1. A Descriptive Model Structure

Carroll (1965) developed a structure for describing a classroom setting. This model involves five variables in a quantitative sense, but one is difficult to quantify. The five variables are not precisely defined, but the variables are well described.

The five variables are aptitude, perseverance, class size, quality of instruction, and opportunity to learn. Carroll suggests that this variable is defined as a reference learning rate for a learner measured by the reciprocal of the time required to achieve a given criterion under optimal learning conditions, as defined by the length of time that the learner is involved. Carroll suggests that this variable changes over time and that it can be affected by external factors. Comprehend-insstruction is assumed to be primitive, and so measures of verbal intelligence are used for identification purposes. It is suggested that this variable changes over time, for example, perseverance. Large extent by the individual's early life environment, quality of instruction, is defined imprecisely and that it can be influenced by external factors. The method of instruction are structured so that method is structured to allow for important joint relationship between quality of instruction and instruction comprehension. It is recognized that in the classroom room opportunity to learn since the class must learn.

Without more explicit elaboration of the remaining variables and in some cases more precise definitions, these variables may be quantitative sense. It has been very useful, however, for identifying major features of the learning process in the classroom.

2. Normative Models

Restle (1964) made an early contribution to the application of optimization theory to the classroom or learning setting. He studied two situations, each of which involve
not for the nonlinear learning-rate characteristics, linearized approximations to them. From the
it is clear that the analysis applies equally well to optimization is performed by means of the Pon
shown that the optimal solution is characterized plane. On the turnpike path the difference \( x_1 - \)
on the two strands remains constant. Optimal all of the instructional effort is allocated to one path is reached. Then the instructional effort is constant difference between strands, that is, so as
ear the end of the learning encounter, the instruc
just one strand, depending on the relative values
jective function. Figure 2 shows the turnpike
xes starting from two different initial points and diffe
ent values of objective function weights.
stable paths, the turnpike path is the one on
maximized. A stable path is the steady state
path that is approached if the relative allocation of instructional effort between strands is held constant. It can be shown that stable paths are such that the difference between achievement levels on the two strands is constant.

C. Group of Learners Setting

1. A Descriptive Model Structure

Carroll (1965) developed a structure for describing learning in the school or classroom setting. This model involves five variables; four are defined in a quantitative sense, but one is difficult to quantify. The relationships among these variables are not precisely defined, but the potential interactions are identified and described.

The five variables are aptitude, perseverance, ability to comprehend instruction, quality of instruction, and opportunity to learn. The aptitude variable is defined as a reference learning rate for a learner for a given task. Aptitude is to be measured by the reciprocal of the time required to master the given task to a given criterion under optimal learning conditions. The perseverance variable is defined by the length of time that the learner is willing to spend learning the task involved. Carroll suggests that this variable will change significantly over time and that it can be affected by external factors. The variable ability-to-comprehend-instruction is assumed to be primarily represented by verbal intelligence, and so measures of verbal intelligence are considered adequate for quantification purposes. It is suggested that this variable will demonstrate less rapid changes over time then, for example, perseverance and that it is determined to a large extent by the individual’s early life environment. Carroll’s fourth variable, quality of instruction, is defined imprecisely as the degree to which content and method of instruction are structured so that material is easily learned. There is an important joint relationship between quality of instruction and ability to comprehend instruction on the learning rate. This relationship is such that low-quality instruction more severely hinders the learner with limited ability to comprehend instruction than the learner with greater ability. The final variable, opportunity to learn, is defined as the time actually allowed for learning in the particular situation. It is recognized that in the classroom not all learners have a continuous opportunity to learn since the class must learn together.

Without more explicit elaboration of the relationships among these variables, and in some cases more precise definitions, this model cannot be used in a quantitative sense. It has been very useful, nevertheless, to help identify the salient features of the learning process in the classroom.

2. Normative Models

Restle (1964) made an early contribution to the application of learning models and optimization theory to the classroom or group of learners setting. He has studied two situations, each of which involves a group of identical learners. In
the first situation, the problem is to determine the optimal class size for a large number of identical learners. The objective function is expressed in cost terms, including both instructor and learner costs, and the amount to be learned is fixed. In the second situation, the problem is to determine the optimal pace of instruction for a curriculum consisting of a sequence of identical items in which further learning progress for any learner is terminated if an item is not mastered. The pace of instruction is determined by the amount of time allocated to each item, assuming equal time for each item and a fixed total amount of time. The objective is to maximize the expected number of items learned by the group or, equivalently, by any learner of the group.

The continuous-time all-or-none model is used to describe the learning process in both situations. This version of the all-or-none model is essentially the same as the discrete (learning trial) version introduced earlier and is defined by the cumulative distribution function

\[ F(t) = 1 - e^{-\lambda t} \]

which gives the probability that learning on an item takes place before time \( t \), where \( \lambda \) is the reciprocal of the mean time until learning occurs.

For the optimal class-size situation, Restle (1964) chooses to minimize the expected total (weighted) time cost of both instructors and learners, subject to the constraint that instruction be given until all learners have mastered the item. Based on the model, the expected time \( M(n) \) for a group of \( n \) learners to learn an item is given by

\[ M(n) = \frac{1}{\lambda} \sum_{k=1}^{n} \frac{1}{k}. \]

Letting \( r \) represent the ratio of the value of instructor time to the value of learner time leads to the expression

\[ N M(n) + r N M(n)/n \]

for expected total time cost in learner-time units where \( N \) is the total number of learners and \( n \) the size of each subgroup (assuming that \( N \) is large enough that the integrality error is negligible). Using a continuous approximation for \( M(n) \), this optimization is easily performed to yield the relationship shown in Figure 3 between optimal class size and \( r \), the relative value of instructor and learner time.

For the situation involving optimal pace of instruction, the total amount of time \( T \) is allocated equally to each item in order to maximize the expected number of items mastered by a learner. If \( t \) units are allocated to each unit, then, based on the model, the mean number of items learned is

\[ e^\lambda - 1 - e^\lambda (1 - e^{-\lambda})^T + 1. \]

Rather than calculate the maximum of this expression with respect to \( t \), Restle (1964) shows the function graphically for various values of the basic parameter \( T \lambda \). With this learning model, \( T \lambda \) represents the total units of time available for instruction on all items. For a short course where \( T \lambda \) is small, the group pace approaches the tutor pace. For a medium-length course where \( T \lambda \) is large, the group pace approaches the tutor pace.

In a paper by Chant and Luenberger (1974), a continuous-time model has been developed that describes certain aspects of learner interaction for an individual learner situation.
Determine the optimal class size for a large objective function is expressed in cost terms, as costs, and the amount to be learned is fixed. The aim is to determine the optimal pace of instruction in a sequence of identical items in which further instruction is terminated if an item is not mastered. The only the amount of time allocated to each item, and a fixed total amount of time. The objective is to maximize the number of items learned by the group or, in the case of all-or-none model is essentially the same as an introduced earlier and is defined by the equation:

\[ M(n) = \frac{1}{\lambda} \sum_{k=1}^{n} \frac{1}{k} \]

where \( M(n) \) is the total number of items learned for a group of \( n \) learners, \( \lambda \) is the rate of learning, and \( k \) is the number of items learned.

Restle (1964) chooses to minimize the total time until all learners have mastered the item. This is given by

\[ t = \frac{1}{\lambda} \frac{1}{n} + rNM(n)/n \]

where \( N \) is the total number of items, \( r \) is the relative value of instructor to learner time, and \( M(n) \) is the total number of items learned. This expression can be approximated by

\[ e^{\lambda t} \frac{(1 - e^{-\lambda t})}{\lambda} \]

for various values of the basic parameter \( \lambda \).

The optimal class size is determined by minimizing the total time until all learners have mastered the item. For a short course where \( \lambda = 3 \), the optimal pace for a group is instruction on 2 items. For a medium-length course of \( \lambda = 12 \), the group should receive instruction on 4 items; and for a long course with \( \lambda = 144 \), the group takes 30 items. Thus, for long sequences of items in which a learner is blocked if he misses only one item, the group pace must be very slow compared to the tutored pace.

In a paper by Chant and Luenberger (1974), a mathematical theory of instruction has been developed that describes certain aspects of the classroom environment. This model is developed in two stages; the first models the instructor/learner interaction for an individual learner situation, and the second extends this
model to a group of learners situation. In the first stage, the principal problem under investigation is the optimal matching of instruction to the characteristics of the learner. In the second stage, the analysis is concerned with the problem of instruction pacing, which is an important question in the classroom situation.

Motivated by a differential equation formulation of the learning curve by Thurstone (1930), Chant and Luenberger (1974) assume that the relationship between learning rate, instructional input, and state of the learner can be represented by

$$p(t) = u(t)g(p(t))$$

where $p(t)$ is the achievement level of the learner at time $t$ relative to total learning. In this equation $p(t)$ represents learning rate, $u(t)$ is an instructional input variable, and $g$—the characteristic learning function—describes how learning rate depends on the achievement level for a particular learner in a particular situation. Restrictions are placed on the function $g$, so that for a constant instructional input $u(t)$ the learning curve has the familiar S-shape.

The instructional input variable $u(t)$ is thought of as a measure of the intensity of instruction in the sense that the larger the value of $u(t)$, the greater the learning rate and the cost of instruction. The relationship between instruction cost and learning rate for a given achievement level forms the basis of the precise definition of $u(t)$ such that the total cost of instruction for $t = 0$ to $t = T$ is

$$\int_0^T \ell[u(t)] dt$$

where $\ell[u(t)]$ defines the rate of expenditure of instructional resources for instruction of intensity $u(t)$, $0 \leq t \leq T$.

In formulating an objective function, both the learner’s achievement level and the cost of the learning encounter are considered. The learner’s achievement level at the end of the encounter is represented by $p(T)$ and the cost of the learner’s time by $bT$. The objective function is defined as the net benefits, that is

$$p(T) - bT - \int_0^T \ell[u(t)] dt.$$  

The relative importance of achievement level and instruction cost is assumed to be included in the loss function $\ell$.

The optimization problem is to choose the instructional input $u(t)$ for $0 \leq t \leq T$ and the duration $T$ of the learning encounter so as to maximize the above objective function. It is shown in the paper that the optimal instructional input function $u$ is constant throughout the learning encounter and is determined by the solution of

$$u \ell'(u) - \ell(u) - b = 0.$$

The optimal value of $T$ is given by the large

$$g[p(T)] = \ell'$$

The result that the optimal instructional input function is quite general in that it does not depend on the characteristic learning function or the particular loss function.

In the development of their group-learning model, Chant and Luenberger (1974) first defined aptitude to characterize the diverse nature of a group of learners under identical situations. One learner is as good as another if he learns the same amount of material as great as another if he learns the same amount in the same time. Aptitude is defined in a relative sense by Carroll’s (1965) mentioned above characteristic learning function $g$ is redefine separated from the other components. The approach becomes

$$p(t) = u(t)ag$$

The above optimization is unchanged with the exception that the instructional input is still constant over time.

The development of the group-learning model is divided into two parts: the optimal pace begins with an analysis of aptitude for an individual learner. To model the body of sequential learning material is divided into basic instructor/learner model outlined above. The sequential nature of the process on each block. The sequential nature of the process on each block where the learning performance on preceding blocks. This interblocks interaction function $h$, which relates the current achievement level on the preceding block to the performance for them can be described with characteristic functions and block-interaction functions. The development of the group-learning model is divided into two parts: the optimal pace begins with an analysis of aptitude for an individual learner. To model the body of sequential learning material is divided into basic instructor/learner model outlined above. The sequential nature of the process on each block. The sequential nature of the process on each block where the learning performance on preceding blocks. This interblocks interaction function $h$, which relates the current achievement level on the preceding block to the performance for them can be described with characteristic functions and block-interaction functions. The development of the group-learning model is divided into two parts: the optimal pace begins with an analysis of aptitude for an individual learner. To model the body of sequential learning material is divided into basic instructor/learner model outlined above. The sequential nature of the process on each block. The sequential nature of the process on each block where the learning performance on preceding blocks. This interblocks interaction function $h$, which relates the current achievement level on the preceding block to the performance for them can be described with characteristic functions and block-interaction functions.
In the first stage, the principal problem of matching instruction to the characteristics of the learner is concerned with the problem of selecting the instructional input, and state of the learner can be represented by

\[ u(t)g[p(t)] \]

The level of the learner at time \( t \) relative to total investment represents learning rate, \( u(t) \). Instructional input, and state of the learner can be represented as

\[ u(t)g[p(t)] \]

The result that the optimal instructional input is constant throughout the learning encounter is quite general in that it does not depend on the particular characteristic learning function or the particular loss function.

In the second stage of their development of a mathematical theory of instruction, Chant and Luenberger (1974) first define a learner aptitude parameter that is used to characterize the diverse nature of a nonhomogeneous group of learners. Aptitude is defined in a relative sense by comparing the learning times of two learners under identical situations. One learner is said to have an aptitude twice as great as another if he learns the same amount in half the time. This definition is similar to Carroll's (1965) mentioned above. Using this concept of aptitude, the characteristic learning function \( g \) is redefined such that the aptitude component is separated from the other components. The basic instructor/learner model now becomes

\[ p(t) = u(t)g[p(t)]. \]

The above optimization is unchanged with this modification, so that the optimal instructional input is still constant over time.

The development of the group-learning model for the purpose of determining the optimal pace begins with an analysis of the relationship between pace and aptitude for an individual learner. To model the effect of instruction pacing, a body of sequential learning material is divided into a sequence of blocks. The basic instructor/learner model outlined above is used to describe the learning process on each block. The sequential nature of the material is captured by specifying how the learner’s performance on one block depends on his achievement on preceding blocks. This interblock dependence is defined by the block interaction function \( h \), which relates the initial state on a block to the final achievement level on the preceding block. For analytical purposes, an infinite sequence of similar blocks is considered. Blocks are similar if the learning performance for them can be described with identical characteristic learning functions and block-interaction functions. The infinite sequence is considered in order to eliminate transient effects and to concentrate on steady state relationships. An infinite sequence of similar blocks is illustrated in Figure 4.

\[ \text{FIGURE 4 Infinite sequence of similar blocks.} \]
The steady state learning behavior of a learner on an infinite sequence of similar blocks is characterized by allocating an equal amount of instructional time to each block and determining the achievement level that the learner approaches on each block as the number of blocks increases towards infinity. The pace of instruction is defined as the amount of time $\tau$ that is spent on each block. In the limit, the initial state on each block is the same, the final achievement level on each block is the same and the pace is such that the learner progresses from this initial state to this final level. This steady state condition is illustrated in Figure 5.

For an individual learner with a particular S-shaped learning curve and block-interaction function, the correspondence between pacing $\tau$ and the steady state final achievement level is defined as the steady state response function $p_s$. With suitable assumptions, it can be shown that $p_s(\tau)$ is 0 for $\tau < \tau_c$, where $\tau_c$ is defined as the critical pace, that $p_s$ is concave and increasing for $\tau > \tau_c$ and has infinite slope at $\tau = \tau_c$.

For determining the optimal pace of instruction, the objective function of steady state achievement level on a block per unit of time on the block is defined. This ratio, called gain and denoted $\gamma$, is given by

$$\gamma(\tau) = \frac{p_s(\tau)}{\tau}.$$  

The maximization of gain implies that

$$p_s(\tau_c) = \tau p_s'(\tau_c).$$

This relationship is illustrated in Figure 6.

The steady state response reference function $p_e$ is defined as the function $p_s$ but for a learner with unity aptitude. In view of the definition of aptitude as the reciprocal of learning time, the response is simply $p_e(\alpha \tau)$.

A nonhomogeneous group of learners in the group with the assumption of characteristic learning functions and block-interaction function for the group, called "group g".

$$\Gamma(\tau) = (1/\tau)$$

where the $N$ learners of the group have a group pace is defined by the maximization. In widely diverse groups, the optimal pace in group has a zero steady state response; the the group because of the fast pace. In a group pace is the same as the one with unity aptitude.

III. AREAS OF FURTHER WORK

This concluding section is intended to application of learning models to problems work. In addition, suggestions are given for the most effective for making these applications.
A learner on an infinite sequence of blocks increases towards infinity. The amount of time $\tau$ that is spent on each block. This steady state condition is illustrated in particular S-shaped learning curve and block-interaction between pacing $\tau$ and the steady state response function $p_{s}$. With known that $p_{s}(\tau)$ is $0$ for $\tau < \tau_{c}$, where $\tau_{c}$ is concave and increasing for $\tau > \tau_{c}$ and has concave and decreasing for $\tau < \tau_{c}$.

The optimal group pace is defined by the maximization of this group gain. It is shown that for widely diverse groups, the optimal pace is such that the lower aptitude part of the group has a zero steady state response; that is, these learners are dropped from the group because of the fast pace. In addition, for homogeneous groups, the optimal group pace is the same as the optimal individual learner pace for that aptitude.

III. AREAS OF FURTHER RESEARCH

This concluding section is intended to highlight a few areas in the field of application of learning models to problems of instruction that require further work. In addition, suggestions are given as to the research directions that may be most effective for making these applications more practical.
A. Problems of Measurement

Problems of measurement exist when we cannot quantify exactly what we want quantified. In order to verify a quantitative model empirically or to apply it in real-world situations, the variables of the model must be measurable. The measurement process can be complicated at either of two levels: the variables of the model may not be satisfactorily quantifiable or, if quantifiable, there may be estimation problems; that is, there may be no satisfactory method of determining a unique value for the defined variable.

To illustrate these two kinds of problems, consider a situation where it is required to have a measurement on the state of a learner with respect to some set of material. At the outset, the first kind of problem is evident since a precise definition of the variable concerned is not available. A satisfactory solution to this problem is perhaps to define a surrogate variable that represents the real variable. In this situation, a proportional measure of the learner’s knowledge of the material as indicated by his score on some test may be an adequate surrogate variable. The second kind of problem has to do with the variability of tests themselves and the learner’s performance on them. Different tests that are intended to measure equivalently the set of material involved will yield different results and the results on a particular test are affected by the testing environment, by guessing and by numerous other factors.

In experimental situations, these problems can be alleviated to a certain extent by careful design. In these situations, the set of material that is to be learned is chosen so that it may be described precisely and simply, for example, in paired-associate learning experiments. This simplifies both the knowledge definition problem as well as the estimation problem. However, in real applications these problems can be severe.

These problems of measurement can be attacked during the formulation and modeling phases of the analysis of problems of instruction. It is of limited use to have a model that cannot be investigated experimentally. It is of no practical value to have an experimentally verified instructional technique that requires such extensive measurement and data analysis that implementation is not cost effective. These measurement problems must be considered during the overall analysis. In some cases, they may be alleviated at implementation by having an estimation model incorporated as part of the technique to be applied.

B. Individual Learner Setting

Optimizing the performance of individual learners is an area that has tremendous potential for impact, even though it has already received some attention. The application of mathematical models and optimization theory to learning problems in computer-aided instruction is likely to prove increasingly useful in the future. Complex models of learning must be developed, and they should be designed for implementation in particular situations. To describe adequately the particular learning phenomena, such complexity is manageable provided the models are implemented in a computer environment. What is needed is a satisfactory approach to the estimation problem to ensure that the ultimate application of the model satisfies the requirements of implementation.

C. Classroom Setting

Developments in the classroom setting will be more complex than those for individual learner setting. They will be developed to cover broad categories of instructional problems. Learning models must be extended and new models developed to account for group-learning phenomena that so far have been neglected. To accomplish this, theoretical and empirical work must be integrated and supplemented each other. Similarly, work in practical instructional strategies must be continually synthesized. One approach is to engage in model-building using an existing existing model or by developing new models. Experimental work must be carried out to verify or refute these models, and the empirical research by imposing a set of constraints on the relationships or conclusions to be tested. Models that comprise an educational system can be made more easily understood.

REFERENCES

in we cannot quantify exactly what we want quantitative model empirically or to apply it in of the model must be measurable. The measured at either of two levels: the variables of the quantifiable or, if quantifiable, there may be may be no satisfactory method of determining able. problems, consider a situation where it is the state of a learner with respect to some set kind of problem is evident since a precise is not available. A satisfactory solution to surrogate variable that represents the realxional measure of the learner’s knowledge of on some test may be an adequate surrogate them to do with the variability of tests nance on them. Different tests that are in set of material involved will yield different test are affected by the testing environment, factors. problems can be alleviated to a certain extent s, the set of material that is to be learned is precisely and simply, for example, in paired- simplifies both the knowledge definition problem. However, in real applications these can be attacked during the formulation and problems of instruction. It is of limited use to nigated experimentally. It is of no practical ffified instructional technique that requires data analysis that implementation is not cost items must be considered during the overall e alleviated at implementation by having an part of the technique to be applied.

Individual learners is an area that has tremendous has already received some attention. The and optimization theory to learning problems to prove increasingly useful in the future.
developed, and they should be designed for implementation in particular situations. These models have to be complex so as to describe adequately the particular learning phenomena in the situation; but such complexity is manageable provided that the models can be adapted for computer implementation. What is needed, then, is a clear understanding of the ultimate application of the model so that its development is guided by the requirements of implementation.

C. Classroom Setting

Developments in the classroom setting are much farther from implementation than those for individual learner setting. For the classroom, general models must be developed that cover broad categories of learning and instruction. Existing models must be extended and new models must be developed to account for group-learning phenomena that so far have been ignored or not even identified. To accomplish this, theoretical and empirical research must complement and supplement each other. Similarly, work by researchers in education and psychology must be continually synthesized. One promising avenue to pursue in this respect would be to engage in model-directed data analysis; that is, either by using an existing model or by developing a model appropriate for the situation to be investigated, data gathering experiments and analyses should be designed and carried out to verify or refute these models. In this approach, the model directs the empirical research by imposing a structure on the system or by proposing relationships or conclusions to be tested. In this way, the complex relationships that comprise an educational system can be more readily isolated and, hence, more easily understood.

REFERENCES


**Author Index**

Numbers in *italics* refer to pages on which the contributor appears.

**A**

Abelson, R. P., 287, 293
Abrahamson, A. A., 194, 234, 242, 243, 244, 270
Abrams, M., 208, 234
Akin, O., 66, 67, 68, 70, 76, 87
Allard, F., 152, 155, 176, 185
Allport, D. A., 100, 137
Anderson, J. A., 6, 10, 16
Anderson, J. R., 5, 6, 9, 16, 51, 59, 60, 61, 62, 80, 88, 208, 225, 226, 228, 229, 243, 266, 267, 274, 287, 293
Anderson, M. L., 258, 269
Anes, J., 119, 137
Arbib, M. A., 207, 230
Arbuckle, T. Y., 105, 141
Armstrong, R., 198, 234
Arnott, P., 195, 196, 232
Arnoult, M. D., 215, 229
Arthur, B., 129, 138
Atkinson, R. C., 57, 80, 84, 85, 86, 87, 88, 210, 232, 278, 294, 303, 304, 305, 306, 317
Atkinson, R. L., 278, 294
Atteave, F., 26, 88, 191, 215, 224, 229
Atwood, M. E., 247, 268, 276, 278, 293
Audley, R. J., 199, 229, 236
Avant, L. L., 119, 137
Averbach, E., 102, 103, 104, 105, 109, 116, 129, 137, 217, 229

**B**

Baddeley, A. D., 109, 140, 160, 185, 210, 233