Reprinted from JOURNAL OF MATHEMATICAL PSYCHOLOGY
All Rights Reserved by Academic Press, New York and London

Vol. 11, No. 2, May 1974
Printed in Belgium

A Mathematical Theory of Instruction: Instructor/Learner Interaction and Instruction Pacing[1]

VERNE G. CHANT

Bureau of Management Consulting, Federal Government, Ottawa, Canada

AND

DAVID G. LUENBERGER

Department of Engineering-Economic Systems, Stanford University, Stanford, California 94305

In this paper, a mathematical theory of instruction applicable in the educational environment is developed from concepts of psychological learning theory. Within the framework of optimization and control theory, the dynamics of the interaction between instructor and learner are modelled, and the trade-off between instruction cost and learner achievement is formulated so that optimal instruction inputs can be determined. One important aspect of the classroom environment that is characterized by the theory is the interaction between an instructor and a group of learners with various learning abilities.

A basic dynamic model that relates learner achievement and instruction cost is developed from learning theory concepts. This model, which applies to the individual learner situation, is analyzed in detail to determine instruction intensity inputs that match the learner's characteristics in order to maximize an objective that measures both achievement and cost.

This basic model is used as a building block to describe how individual learner achievement depends on instruction pacing. To determine optimal instruction pacing the concept of gain, which is essentially learner achievement per unit time, is introduced. In this extended model, instruction pacing is intimately related with the concept of learner aptitude. This relationship leads immediately to the consideration of instruction pacing for a group of learners with various aptitudes and thus optimal instruction pacing is determined for nonhomogeneous groups.

Throughout the development of the theory, hypothetical examples are presented to demonstrate many of the implications of the theory. One of the contributions of the theory is the definition of the concepts of learner aptitude and instruction pacing within a framework that structures the empirical investigation of these concepts by means of experimental research.

[1] This research was partially supported by National Science Foundation Grants GK-29237 and GK-32870. The authors gratefully acknowledge the comments and suggestions offered by Professor R. C. Atkinson of the Department of Psychology, Stanford University, during the course of this research.

Introduction

Only a small part of the total education process takes place in structured environments like schools and classrooms. But it is principally in these situations where decisions can be made regarding resource allocation and instructional techniques that affect the opportunity of students to learn. In general, the decisions that must be made involve allocating instructional resources among students and groups of students and designing instructional techniques to match individual learner characteristics. The objective in making these decisions is to balance the trade-off between learner achievement and instruction cost. In this paper, a mathematical theory of instruction is developed, within the framework of optimization and control theory, that structures this problem and leads to quantitative analysis.

At the heart of this framework is a dynamic model that describes the system involved in terms of its key variables, both state and control (decision) variables. This model is complete in the sense that for any possible set of decisions over time, or control input, the model shows what happens to the state variables. Since the implications of each decision are known, preferences for one control input over another can be assessed that take into account costs as well as benefits. In control theory, these preferences are specified by an objective function which is defined in terms of the state and control variables. The problem of choosing the best control input is then defined as the maximization (or minimization) of this objective function with respect to feasible control inputs.

The dynamic behavior that is central to all problems of instruction is the interaction between the individual learner and the instructor, whether the instructor is represented by a teacher, a teaching machine or programmed instruction. A model of this interaction is essential in order to understand more complex situations in the educational setting. In this paper, a model of this behavior is developed based on mathematical learning theory concepts. The model is then extended to describe the group learner situation.

Basic Model of the Instructor/Learner Interaction

Thurstone's Learning Function

The progress of a learner while learning a specific task or body of material can be described by a learning curve. Thurstone (1930) proposed a differential equation model to explain the learning curve, or learning function as he called it, in certain simple learning situations. Although his model applies only to simple learning situations, it provides motivation for a more general, dynamic model that forms the basis of our analysis. This more general model is not as detailed as typical models of mathematical learning theory which are capable of predicting learner behavior on each trial of an experiment, but describes instead the progress of a learner through a block of learning material in an average, aggregate manner. Thurstone's development of the learning function is reviewed in the following paragraphs.

In Thurstone's approach, it is assumed that in the process of learning, a learner performs several "acts," some of which can be classified as successes and the others as failures. It is assumed further that the learner has a repertoire of acts that is composed of a number, s, of potential successes and a number, e, of potential failures. The state of the learner is indicated by the relative number of potential successes in his repertoire. The phenomenon of learning, then, is reflected by changes in s and e over time. If the state of the learner at time t is denoted by $p(t)$, then, by definition,

$$p(t) = s(t)/[s(t) + e(t)]. \tag{1}$$

The state of the learner can be interpreted as the probability that an initiated act ends in success.

Thurstone's model of the learning process is completed by specifying the manner in which s and e change. It is assumed that acts are initiated at a constant rate, that time is measured in "acts initiated" units, and that learning takes place after a success by increasing $s(t)$ and after a failure by decreasing $e(t)$. Since the probability of an act at time t leading to a success is $p(t)$, the change in s is modelled by setting

$$\dot{s}(t) = kp(t) \tag{2}$$

where $\dot{s}(t)$ is the rate of increase in s per unit time. Similarly, the change in e is defined by

$$\dot{e}(t) = -k[1 - p(t)] \tag{3}$$

where, as a simplification, k is the same constant as in Eq. (2). Thurstone notes that the positive constant k is a measure of learner aptitude, since the rate of decrease in potential failures and the rate of increase in potential successes is proportional to k.

Equations 1–3 completely characterize Thurstone's model. These equations may be combined to eliminate the s and e variables to obtain

$$\dot{p}(t) = (2k/m)\,[p(t)(1 - p(t))]^{3/2} \tag{4}$$

where $p(t)$ is the learning rate and m is a positive constant equal to $s(t)\,e(t)$. The parameter m is a measure of the complexity of the learning task since m varies with the "size" of the repertoire.

The solution of the differential equation (4) is defined as the learning function or, synonomously, the learning curve. From Eq. (4), it is clear that the curve is asymptotic to $p = 0$ and $p = 1$ and is symmetric about $p = \frac{1}{2}$. Figure 1 shows typical learning curves that are solutions to Eq. (4) for different values of the ratio k/m.

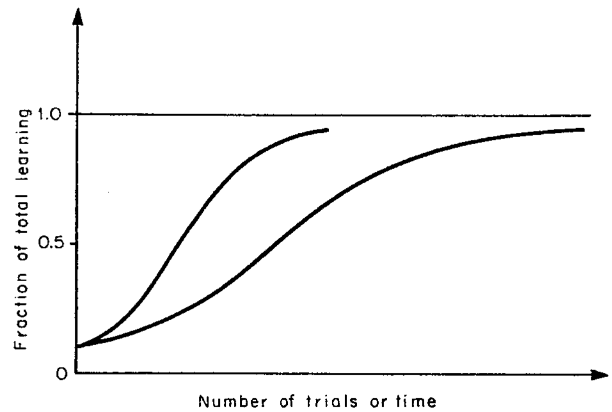

FIG. 1. Typical Thurstone learning curves; solutions to Eq. (4) for different values of k and m.

Basic Model of the Instructor/Learner Interaction

To study the instructor/learner interaction, a variable representing instructional input must be included in the model. Motivated by Thurstone's equation of the learning function, we assume that instructional input interacts with the learner to effectively increase k, the aptitude for learning, or decrease m, the complexity of the learning task. In other words, the effect of increased instructional input is to increase the instantaneous learning rate of the learner. The effect of instruction is represented by the instructional input variable $u(t)$.

As a generalization of Thurstone's model, we assume that the relationship between learning rate and the state of the learner is defined by any smooth function g with certain properties defined below, rather than the particular function given in Eq. (4). The combination of the effect of instructional input and the effect of the state of the learner on the learning rate is given by

$$\dot{p}(t) = u(t)\, g(p(t)), \tag{5}$$

which we call the basic model of the instructor/learner interaction.

We call the function g the *characteristic learning function*. The nature of g depends on the characteristics of the learner and of the material to be learned. In general, we assume that g is continuous and that $g(p)$ approaches zero as p approaches either zero or one. This latter assumption preserves the asymptotic behavior of the learning curve as exhibited by Thurstone's model. This behavior reflects learning situations where the initial learning rate is small when little is known about the material and where the final learning rate is limited as total learning is approached. For the instruction pacing model of the next section, we assume further that g increases monotonically to its maximum and then decreases monotonically to zero. This assumption implies that, for constant instructional input $u(t)$, the learning curve is S-shaped.

One of the objectives in considering this model of the instructor/learner interaction is to be able to prescribe optimal instructional inputs. To do so, it is necessary not only to measure the learning achievement of a particular instructor/learner encounter but also to assess the cost of the encounter. For this reason, the instructional input variable is defined in terms of instruction cost.

The instructional variable $u(t)$ can be 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 of the learner, but also, the greater the cost of the instruction. For a particular instructor/learner situation, there may be many possible instructional methods and many levels of instruction for each method. Instead of considering each method independently, we assume that the characteristics of the instructional methods can be summarized by a single functional relationship that relates learning rate to the rate of expenditure of instructional resources.

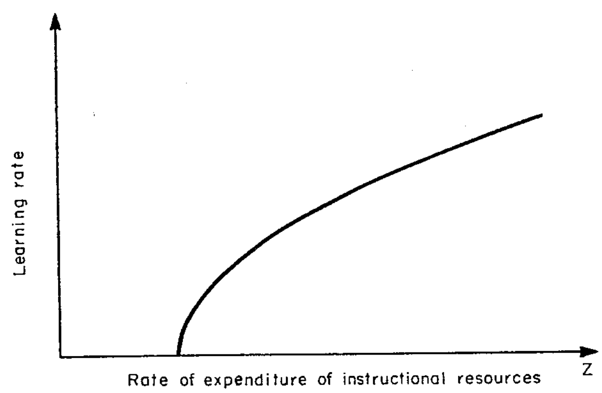

FIG. 2. General form of the relationship between learning rate and instructional cost.

The general form of this relationship is assumed to be of the form illustrated in Fig. 2. In this figure, the instantaneous learning rate, $\dot{p}$, for a fixed state of learning is shown as a function of the rate, z, of expenditure of instructional resources. As shown, there is a minimum rate of expenditure for learning to take place. This minimum rate can be thought of as an allocation of certain fixed instructional costs to this instructor/learner situation, costs that do not depend on the particular method or level of instruction that is actually employed, but which will be allocated elsewhere when this instructor/learner encounter terminates. Above this minimum rate, the raletionship between learning rate and instructional cost is concave, which implies diminishing marginal returns. That is, a small increase in instructional expenditure is not as effective at an already high level of expenditure as it would be at a low level of expenditure.

Since in the basic model of the instructor/learner interaction $u(t)$ affects the learning rate independently of $p(t)$, learning rate may be expressed by

$$\dot{p}(t) = f(z(t))\, g(p(t)) \tag{6}$$

where f denotes the functional relationship illustrated in Fig. 2 expressed in suitable units. Consequently, u is defined by

$$u(t) = f(z(t)) \tag{7}$$

and the instantaneous cost associated with the level $u(t)$ of instructional input is

$$z(t) = f^{-1}(u(t)). \tag{8}$$

The inverse of f is well-defined for all nonzero values of $u(t)$ since it is only reasonable to assume that f is strictly increasing for z greater than the minimum rate.

The loss function l, a function of $u(t)$, is defined as the inverse of f for nonzero arguments. The value of l for $u(t) = 0$ is defined to be zero. Since $z(t)$ is a rate of expenditure, the total cost associated with the input $u(t)$ for a learning encounter from $t = 0$ to $t = T$ is given by the integral

$$\int_0^T l(u(t))\, dt. \tag{9}$$

Optimal Instructional Inputs

To define an optimal instructional input, there must be a preference structure for selecting one input over another. As a general objective, the preference is for high learning achievement and low instructional cost. In the particular case of an instructor/learner encounter with input $u(t)$ lasting for T time units, a reasonable measure of learner performance is the value of the state variable $p(T)$ at the end of the encounter. The corresponding instructional cost is given by Eq. (9). A suitable measure of the trade-off between achievement and cost is then given by the difference

$$J = p(T) - \int_0^T l(u(t))\, dt \tag{10}$$

which is labelled J and called a *performance index*. The relative importance of achievement and cost in the performance index can, without loss of generality, be assumed to be included by the appropriate scaling of the loss function l.

An optimal instructional input can now be defined. For the case of an instructor/learner encounter over the time period $0 \leqslant t \leqslant T$, the optimal instructional input, $u(t)$, is that which maximizes the performance index, Eq. (10). The final state, $p(T)$, of the learner is given by the basic model of the instructor/learner interaction defined by Eq. (5) for a particular characteristic learning function and initial learner state.

Mathematically, the optimal instructional input is defined as the solution to the maximization problem:

$$\begin{aligned}&\text{maximize } J,\\ &\text{subject to } \dot{p}(t) = u(t)\,g(p(t)) \qquad p(0) = p_0\,.\end{aligned} \tag{11}$$

This maximization problem can be solved in several ways. In this presentation, the differential of J with respect to the function u is calculated, and the necessary condition that the differential be zero is used to determine the optimal u. To this end, we note that the differential equation (5) can be written as

$$G(p(t)) = \int_0^t u(\tau)\,d\tau \tag{12}$$

where G represents the definite integral

$$G(p) = \int_{p_0}^{p} g(x)^{-1}\,dx \tag{13}$$

and p_0 is the value of p at $t = 0$. $G(p)$ is strictly increasing, since $g(p)$ is strictly positive in the open interval (0, 1). It is only meaningful for $g(p)$ to be positive, since a negative value would imply a negative learning rate and a zero value at other than $p = 0$ or $p = 1$ would imply an intermediate steady state achievement level. The level of achievement at the end of the learning encounter, $p(T)$, can be written as

$$p(T) = G^{-1}\left(\int_0^T u(\tau)\,d\tau\right). \tag{14}$$

Thus the constrained maximization problem (11) becomes the unconstrained problem:

$$\text{maximize } J(u) = G^{-1}\left[\int_0^T u(\tau)\,d\tau\right] - \int_0^T l(u(\tau))\,d\tau. \tag{15}$$

The differential of J evaluated at u with increment h is denoted by $\delta J(u; h)$, where u and h are both vectors in a suitable (infinite dimensional) vector space. If h is considered to be a change in u, the differential $\delta J(u; h)$ is the linear approximation to the corresponding change in J. The differential can be expressed as [Luenberger (1969)]

$$\delta J(u; h) = G^{-1'}\left[\int_0^T u(\tau)\,d\tau\right]\int_0^T h(\tau)\,d\tau - \int_0^T l'(u(\tau))\,h(\tau)\,d\tau \tag{16}$$

where the prime denotes the ordinary one dimensional derivative with respect to the function's argument. The derivative of the inverse of G can be found in a straightforward manner to be

$$G^{-1'}\left[\int_0^T u(\tau)\,d\tau\right] = g(p(T)). \tag{17}$$

Thus, the differential (16) becomes

$$\delta J(u; h) = g(p(T)) \int_0^T h(\tau)\, d\tau - \int_0^T l'(u(\tau))\, h(\tau)\, d\tau \tag{18}$$

which, by combining terms, can be expressed as

$$\delta J(u; h) = \int_0^T [g(p(T)) - l'(u(\tau))]\, h(\tau)\, d\tau. \tag{19}$$

For u to be the optimal instructional input, the differential (19) must be zero for all possible increments h. It follows that

$$g(p(T)) - l'(u(t)) = 0 \tag{20}$$

for all t, and consequently, $l'(u(t))$ is constant for all t. It has already been assumed that l is convex for nonzero u. If this assumption is strengthened to strict convexity, then l' is strictly increasing. The solution of (20) is, therefore, a unique, constant value for the instructional input $u(t)$ for all t, $0 \leqslant t \leqslant T$.

We have seen that the maximization of the objective of (11) results in an instructional input that does not vary in intensity throughout the instructor/learner encounter. This important result is quite general in that it does not depend on the particular characteristic learning function g or the particular loss function l. This, of course, simplifies analysis of the model and allows for further study of its implications and development toward a richer theory of instruction. The fact that the optimal input is constant has the practical implication that it is not necessary to know the instantaneous value of the state of the learner in order to determine the optimal input.

We have just found the optimal instructional input for an instructor/learner interaction for a fixed encounter time, T. We now turn to the problem of determining the optimal value of T. For this purpose, the performance index (10) is augmented to account for the value of the learner's time. Learner time cost becomes an important factor when the length of the learning encounter is being considered. Learner cost is included as a linear term in the performance index with a weighting coefficient b, which is a constant reflecting the importance of learner cost relative to instructional cost. The new performance index is, then,

$$J = p(T) - \int_0^T l(u(t))\, dt - bT. \tag{21}$$

For fixed T, the added term does not change the optimization with respect to u, so that the previous optimization still applies. That is, the optimal instructional input, u, is a constant function. The value of this constant function will also be represented by u, for notational simplicity. Whether u represents the function or its value will be clear from the context. Equation 21 can be rewritten by using Eq. (14) to replace $p(T)$ and by performing the integrations involved. That is,

$$J = G^{-1}(uT) - Tl(u) - bT. \tag{22}$$

A necessary condition for T to maximize the performance index J, is that the partial derivative of J with respect to T be zero. This derivative is

$$\partial J/\partial T = uG^{-1\prime}(uT) - l(u) - b. \tag{23}$$

Replacing the derivative of the inverse of G by Eq. (17), the necessary condition becomes

$$ug(p(T)) - l(u) - b = 0. \tag{24}$$

The intuitive interpretation of Eq. (24) is simply that the rate of increase of achievement must be equal to the rate of increase of cost at the optimal final time T. The first term of Eq. (24) is the learning rate while the second and third terms are the rate of increase of instructional cost and learner cost, respectively.

Using the necessary condition for optimal u from Eq. (20), Eq. (24) can be rewritten as

$$ul'(u) - l(u) - b = 0. \tag{25}$$

Since l is strictly convex and increasing, Eq. (25) yields a unique value for u. This value of u can be used in Eq. (20) to calculate the optimal T. Equation (20), however, will in general, under our assumptions on g, yield two values for T. It is shown below that the optimal T is the larger of these two values.

To verify that a solution to the necessary conditions represents a maximum, it is sufficient to show that the matrix H of second partial derivatives of J with respect to u and T is negative definite at this solution. In the present case, the appropriate matrix is

$$H = \begin{bmatrix} T(Tr - l'') & uTr \\ uTr & u^2 r \end{bmatrix}, \tag{26}$$

where

$$\begin{aligned} r &= G^{-1\prime\prime}(uT) \\ &= g'(p(T))\, g(p(T)). \end{aligned} \tag{27}$$

It can seen that if r is negative then so are the two diagonal terms of H, since l'' is positive by the convexity of l. The negative definiteness of H follows then if its determinant is positive, which is clear since

$$\det(H) = -u^2 T l'' r. \tag{28}$$

In examining the sign of r, note that $g(p(T))$ is positive, so that the sign of r is the same as the sign of $g'(p(T))$. For an S-shaped learning curve, the derivative g' of g is positive for values of p up to the point of inflection of the learning curve and negative for larger values of p (and, consequently, larger values of T). Thus, the sufficient condition for a maximum is satisfied if the final value of learner achievement, $p(T)$, is past the point of inflection of the learning curve. For positive r, H is clearly not negative definite. The requirement that the larger value of T be chosen is intuitive, since up to the point of inflection the learning rate is steadily increasing, and so the maximization of achievement would certainly imply that instruction should be continued.

The above determination of the optimal instructional input u and learning encounter time T assumed that some instruction would be given. It is possible, however, that whatever the level and duration of instruction, that cost exceeds achievement. If that is the case, the optimal solution is to give no instruction, that is, $u(t) = 0$ for all t.

Definition of Aptitude

When exposed to essentially identical instructional environments, different learners learn at different rates. Now, certainly, a so-called fast learner is not going to learn faster than a slow learner in all situations, but characteristically, on the average, certain learners learn faster than others under similar circumstances. It is this characteristic, defined as learning aptitude, or simply aptitude, that becomes so important in our later model when instruction pacing for a diverse group is analyzed.

Aptitude is defined in a relative manner 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 of aptitude is the one Carroll (1963, 1965) uses in his model of school learning. Thurstone's parameter k is directly proportional to aptitude in this sense.

Incorporating this concept of aptitude into the dynamic model presented above is straightforward once it is realized that the definition requires learning rate to be proportional to aptitude. Thus, if a learner has aptitude represented by a, then

$$\dot{p}(t) = u(t)\, a g(p(t)) \tag{29}$$

so that the aptitude component of the characteristic learning function is separated from its other components. Henceforth, $g(p)$ represents learning characteristics other than learner aptitude. The solution of the differential equation (29) is the same as that given in Eq. (14) with $u(t)$ replaced by $au(t)$. For constant (optimal) instructional inputs, this solution can be written

$$p(t) = G^{-1}(aut) \tag{30}$$

which emphasizes the equivalence of higher aptitude and shorter learning time.

Individual Learner Example

The following example is given to illustrate some of the implications of this basic model for an individual learner. For purposes of example, the variable component of the loss function is taken to be quadratic. In particular,

$$l(u) = \begin{cases} u^2 - c, & u > 0; \\ 0, & u = 0. \end{cases} \tag{31}$$

The quadratic term implies that, in order to double the learning rate of the learner, the variable cost of the instructional input must be increased four times. The optimal instructional input is found from (25) to be

$$u = (b + c)^{1/2}. \tag{32}$$

For the characteristic learning function, $g(p)$, the same symmetric form as that derived by Thurstone [Eq. (4)] is used, but, for simplicity, the 3/2 power is dropped. Thus,

$$g(p) = p(1 - p) \tag{33}$$

and

$$G(p) = \int_{p_0}^{p} g(x)^{-1}\, dx = \log\{[(1 - p_0)/(p_0)]\,[(p)/(1 - p)]\}. \tag{34}$$

Equation (20), in this case, is [noting that in Eq. (20), the aptitude component is included in g]

$$2u = ap(T)(1 - p(T)) \tag{35}$$

which yields two values for $p(T)$. Choosing the larger value, and using Eq. (30) with Eq. (34), gives

$$\log\{[(1 - p_0)/(p_0)]\,[(p(T))/(1 - p(T))]\} = auT \tag{36}$$

from which T can be determined. Finally, the value of the performance index is, according to Eq. (21),

$$J = p(T) - 2(b + c)T. \tag{37}$$

Figure 3 shows the learner response to the optimal instructional input for the arbitrary parameter values $a = 1.0$, $b = c = 0.0005$, and $p_0 = 0.1$. Figure 4 shows the dependence of the final level of learner achievement $p(T)$ and the performance index J on learner aptitude a. For very small values of aptitude the optimal instructional input is zero. With no instruction, the performance index is p_0, as shown in Fig. 4 for values of a less than a_m. Figure 5 shows that the optimal duration of the learning encounter decreases as the learner aptitude increases.

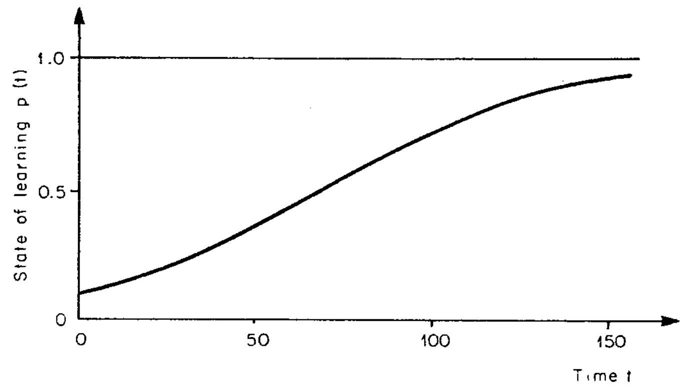

FIG. 3. Learner response to optimal instructional input.

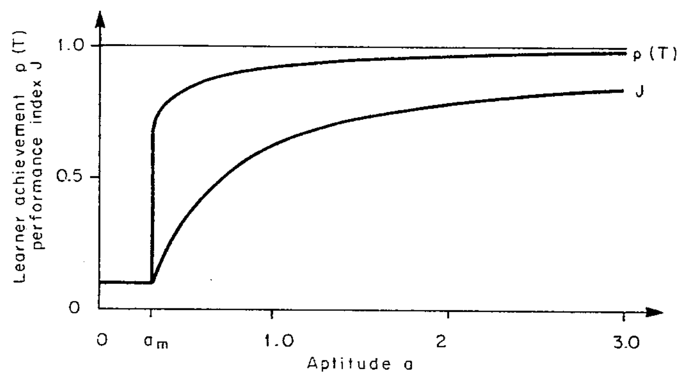

FIG. 4. Learner achievement and performance index versus aptitude for an individual learner.

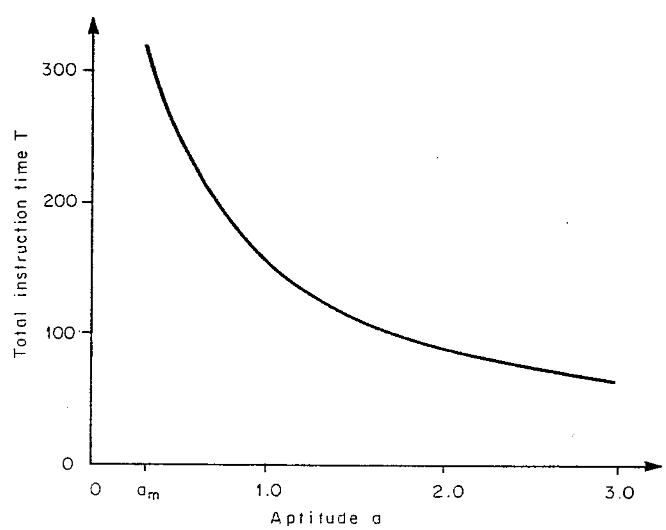

FIG. 5. Total instruction time versus aptitude for an individual learner.

Instruction Pacing for a Nonhomogeneous Group

Instruction Pacing

Instruction pacing refers to the speed with which material is presented to a learner. When only one learner receives instruction, the pace of instruction may be designed especially for that learner. When the instruction is given to a group of learners with various learning aptitudes, however, the optimal pace of instruction is more difficult to determine. If instruction is given at a comfortable pace for one of the slower learners, then the fast learners are likely to be bored. If instruction is paced for the fast learners, then the slow learners may become lost and gain nothing from subsequent instruction. The optimal pace is somewhere between the two extremes, a pace that balances overall group achievement and instruction time.

Intensity and pacing are two principal components of the many attributes of instruction. In the basic model of the instructor/learner interaction discussed in the previous section, the instructional variable u is a measure of the intensity of instruction and the key concept of the model is that of matching instructional input to the learning curve of the learner. In this section, pacing of instruction is analyzed by embedding this basic model in a comprehensive model of the instructor/learner interaction. To simplify the analysis, the effect of instruction pacing is studied independently of the intensity component. The joint effect of these two components could then be analyzed parametrically.

Instruction pacing is analyzed by first modelling the effect of instruction pacing on an individual learner. Analysis of this model leads to the determination of the optimal pace of instruction for an individual learner. Then the model is extended to apply to a nonhomogeneous group (a group of learners with various aptitudes) from which the optimal group pace is determined.

Pacing 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 model developed in the previous section is then 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. If too little time is spent on each block (too fast a pace of instruction), then the learner's performance steadily decreases.

The learning performance on block i is described by the basic model

$$\dot{p}_i(t) = uag_i p_i(t)) \tag{38}$$

once the initial state, $p_i(t_i)$, at the initial time, t_i, is specified. The dependence between blocks is modelled by specifying how the initial state on a block depends on the final achievement level on the preceding block. In this model, achievement on earlier blocks in the sequence affects block i only to the extent that it affects the final achievement level on block $i-1$. The *block interaction function* h_i, for block i, is defined as the relationship between the initial state on block i, $p_i(t_i)$, and the final achievement level on block $i-1$, $p_{i-1}(T_{i-1})$ where T_{i-1} represents the final time on block $i-1$; that is,

$$p_i(t_i) = h_i(p_{i-1}(T_{i-1})). \tag{39}$$

The block interaction function is a mapping from the interval [0, 1] into [0, 1] and is assumed to be nondecreasing. Four possible forms are illustrated in Fig. 6 where, in general, a higher achievement level on one block yields a higher initial state on the subsequent block. The piecewise linear example shown in Fig. 6, which gives a zero initial state for all final states below a certain cut-off point, has the property that [in view of the assumption that $g(p) \to 0$ as $p \to 0$] no further progress in the sequence of blocks is possible if the preceding block is not learned to a level above this cut-off point.

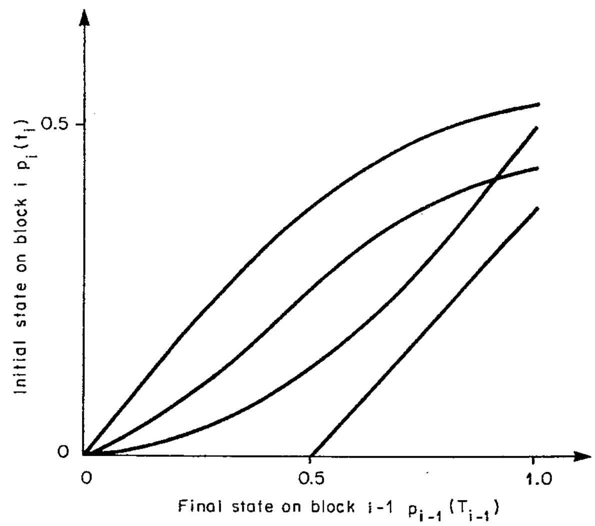

FIG. 6. Examples of possible block interaction functions.

For a particular sequence of blocks with known characteristics, it would be possible to consider the problem of allocating a given total amount of instructional time among the blocks so as to maximize some overall measure of learner achievement. It would be difficult, however, to draw general conclusions from such a particular case. The concept of steady state learner achievement on an infinite sequence of similar blocks is introduced, therefore, in order to analyze the general effects of instruction pacing.

This analysis begins with a number of definitions: Two blocks of a sequence are *similar* if a learner's performance on each block is described by the same characteristic learning function and the same block interaction function. An infinite sequence of similar blocks is defined by specifying a learning curve and a block interaction function. Such a sequence is illustrated in Fig. 7. The *pace* of instruction is defined as the amount of instruction time, τ, that is given on each block when an equal amount of time is spent per block. *Steady state learner achievement* with pacing τ is defined as the final achievement level on each block in the limit of an infinite sequence of blocks.

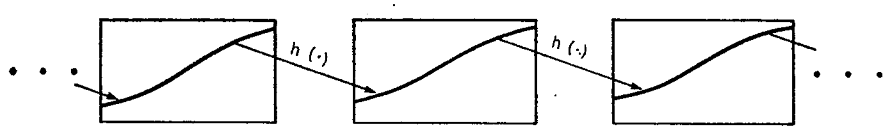

FIG. 7. Infinite sequence of similar blocks.

The concept of an infinite sequence of similar blocks, although not realizable in practical situations, is a useful abstraction for analytic purposes. This concept allows large scale, or macro, effects to be studied by means of building on small scale, or micro, learning theory models.

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 approaches infinity. In the limit, the initial and final achievement levels on the same block must be related by the block interaction function. In other words, the initial achievement levels on all blocks in the steady state are the same, as are the final achievement levels, and the instruction time spent on each block must be such that the learner progresses from this initial level to this final level.

As described by the basic model (38), the achievement profile on a block is an S-shaped learning curve. For a sequence of similar blocks, a learner has the same learning curve, p, and interaction function, h, for each block. We assume that the block interaction function satisfies

$$h(p) \leqslant \alpha p \tag{40}$$

for some positive $\alpha < 1$ so that the initial level on a block is strictly lower than the final level on the previous block. The time axis of the learning curve can be oriented so that $p(0) = 0.5$, without loss of generality. If $p(t_a)$ and $p(t_b)$ refer to the initial and final achievement levels on a block in the steady state, then the steady-state condition is

$$p(t_a) = h(p(t_b)) \tag{41}$$

where $t_b = t_a + \tau$. Figure 8 illustrates this condition.

In view of the assumption (40) on h, it is clear that for very small values of τ, the only solution to Eq. (41) is the limiting case $p(t_a) = p(t_b) = 0$. That is, for a very fast pace such that learning on a block is negligible, the repeated application of the block interaction function forces the learner state to zero. On the other hand, for very large values of τ, the final state $p(t_b)$ approaches unity and the initial state $p(t_a)$ approaches $h(1)$. For other values of τ, there may be more than one solution of Eq. (41) and in such cases, the solution giving the maximum value of $p(t_b)$ is the desired steady-state solution.

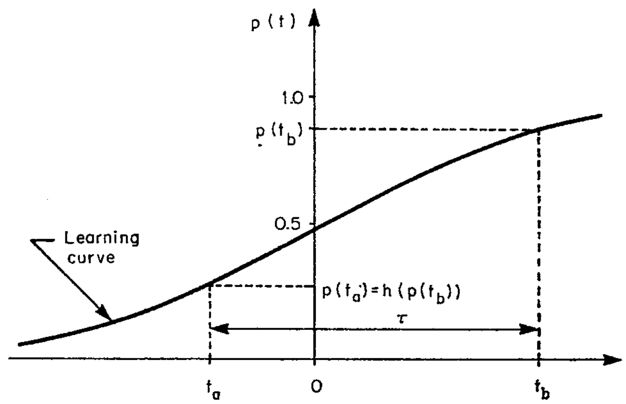

FIG. 8. Illustration of steady state condition.

For an individual learner with a particular S-shaped learning curve, the correspondence between pacing τ and steady-state learner achievement $p(t_b)$ is defined as the *steady-state response function* p_s. That is,

$$p_s(\tau) = p(t_b) \tag{42}$$

where t_b and $p(t_b)$ are given by Eq. (41). The function p_s represents the basic relationship between instruction pacing and learner achievement.

From the previous discussion, it can be seen that $p_s(\tau)$ is zero for small τ and, for large τ, is approximated by the learning curve with origin shifted so that $p(0) = h(1)$. The value τ_c, of τ, such that $p_s(\tau) = 0$ for $\tau < \tau_c$ is defined as the *critical pace.* Under suitable assumptions, it can be shown [Chant (1973]) that p_s is concave and increasing for $\tau > \tau_c$ and has infinite slope at $\tau = \tau_c$. To possess these properties, it is sufficient, although not necessary, that h be concave in the region where the learning curve is concave and that the point of inflection of the learning curve be contained in the interval (t_a, t_b). The former condition is not necessary as shown by the example in the following section. The latter condition implies that, in the steady state, each block is learned to a level above the point of the learner's maximum learning rate and that the starting point on each block is at a level below this maximum point.

The trade-off between achievement and cost is measured by defining a performance index as the ratio of steady-state achievement on a block to the time spent on the

For a fixed value of τ, gain varies with learner aptitude and is proportional to the steady state level $p_s(\tau)$. It is clear from Eq. (51) that p_s is a function only of the product $a\tau$ and not a and τ separately. Hence, if the abscissa of Fig. 10 is interpreted in aptitude units, then the figure also shows the relationship between gain and aptitude for a fixed value of τ. It is this relationship that is important for analyzing a nonhomogeneous group of learners.

Pacing for a Group of Learners

For a learner in a learning situation described by a particular characteristic learning function and block interaction function, the steady-state achievement depends on the learner's aptitude and on the pacing of instruction. The ***steady-state response reference function***, p_r, is defined such that the steady-state learner achievement for a learner with unity aptitude receiving instruction with pace τ is $p_r(\tau)$. In other words, the function p_r is that particular steady-state response function, p_s, that applies to an individual learner with unity aptitude. The function p_r (as does p_s) depends on the characteristic learning function and the block interaction function. For the example in the preceding section, p_r is given by Eq. (51) with $a = 1$ and is illustrated in Fig. 10. In view of the definition of aptitude as the reciprocal of learning time, the steady state achievement for a learner with aptitude a and pacing τ is simply $p_r(a\tau)$. The gain for this learner depends on a and τ as given by Eq. (43). The dependence on both a and τ is emphasized now by writing γ with two arguments. That is,

$$\gamma(a, \tau) = [p_r(a\tau)]/\tau. \tag{52}$$

The concept of a group of learners and a measure of performance for a group is now introduced. We assume that, for the learning situation to be analyzed, the group is composed of individual learners with identical characteristic learning functions and identical block interaction functions. The learners of the group are differentiated only by different aptitudes. As a measure of performance for the group, we define the ***group gain***, denoted Γ, to be the sum of the achievement levels of the individual learners in the group divided by the group instruction time. If there are N learners in the group, and a^i is the aptitude of learner i, then the group gain, a function of τ, is

$$\Gamma(\tau) = (1/\tau) \sum_{i=1}^{N} p_r(a^i\tau). \tag{53}$$

It is convenient, for the purpose of employing a differential analysis, to approximate the group gain (53) by an integral. The aptitude characteristics of the group are described by a ***group aptitude density function***, ϕ. If $a_{\min}$ and $a_{\max}$ denote the minimum and maximum aptitudes, respectively, of the group, then ϕ is defined on the interval

$[a_{\min}, a_{\max}]$ such that the number of learners with aptitude less than a (for $a \leqslant a_{\max}$) is given by

$$\int_{a_{\min}}^{a} \phi(\zeta)\, d\zeta. \tag{54}$$

In terms of this density function, the summation in Eq. (53) can be approximated by the integral

$$\Gamma(\tau) \simeq (1/\tau) \int_{a_{\min}}^{a_{\max}} \phi(\zeta)\, p_r(\zeta)\, d\zeta. \tag{55}$$

Equation (55) shows the dependence of group gain on instruction pacing for a group of learners with aptitude distribution ϕ and with learning characteristics represented by the reference function p_r. The optimal instruction pace for the group is the pace that maximizes this group gain.

To proceed with this maximization, it is necessary first to note that the function p_r is zero over the lower part of its domain. When the product $a\tau$ is less than some critical value, denoted here by x_c, $p_r(a\tau)$ is zero. In view of this fact, Eq. (55) for the group gain can be written

$$\Gamma(\tau) = \begin{cases} \dfrac{1}{\tau} \displaystyle\int_{x_c/\tau}^{a_{\max}} \phi(\zeta)\, p_r(\zeta\tau)\, d\zeta, & \tau < (x_c/a_{\min}); \\[2ex] \dfrac{1}{\tau} \displaystyle\int_{a_{\min}}^{a_{\max}} \phi(\zeta)\, p_r(\zeta\tau)\, d\zeta, & \tau \geqslant (x_c/a_{\min}). \end{cases} \tag{56}$$

The optimal group pace is characterized by the derivative of Eq. (56). Assuming that p_r is differentiable, the derivative of group gain, with respect to τ, is given by

$$\Gamma'(\tau) = \begin{cases} \dfrac{1}{\tau} \displaystyle\int_{x_c/\tau}^{a_{\max}} \phi(\zeta)\, \zeta p_r'(\zeta\tau)\, d\zeta + \dfrac{x_c}{\tau^3}\, \phi\left(\dfrac{x_c}{\tau}\right) p_r(x_c) \\[2ex] \qquad\qquad - \dfrac{1}{\tau^2} \displaystyle\int_{x_c/\tau}^{a_{\max}} \phi(\zeta)\, p_r(\zeta\tau)\, d\zeta, \quad \tau < \dfrac{x_c}{a_{\min}}; \\[2ex] \dfrac{1}{\tau} \displaystyle\int_{a_{\min}}^{a_{\max}} \phi(\zeta)\, \zeta p_r'(\zeta\tau)\, d\zeta - \dfrac{1}{\tau^2} \displaystyle\int_{a_{\min}}^{a_{\max}} \phi(\zeta)\, p_r(\zeta\tau)\, d\zeta, \quad \tau > \dfrac{x_c}{a_{\min}}; \end{cases} \tag{57}$$

and is not defined for $\tau = x_c/a_{\min}$.

In the neighborhood of $\tau = x_c/a_{\min}$, the derivative of Γ for $\tau < x_c/a_{\min}$ is greater than that for $\tau > x_c/a_{\min}$ due to the extra positive term in the upper line of Eq. (57) (provided that $\phi[(x_c)/\tau]$ or $p_r(x_c)$ is not zero). But the function Γ is continuous at $\tau = x_c/a_{\min}$, and hence, if $\Gamma'(\tau)$ is positive from the left and negative from the right of this value, then Γ is maximized at $\tau = x_c/a_{\min}$. This value of τ is the critical pace for the minimum aptitude learner of the group.

If the optimal pace τ_0 is such that $\tau_0 < x_c/a_{\min}$, then the upper expression of Eq. (57) must be zero. That is,

$$\int_{x_c/\tau_0}^{a_{\max}} \phi(\zeta)\, p_r(\zeta\tau_0)\, d\zeta = \int_{x_c/\tau_0}^{a_{\max}} \phi(\zeta)\, \zeta\tau_0 p_r'(\zeta\tau_0)\, d\zeta + (x_c/\tau_0)\, \phi[(x_c)/\tau_0]\, p_r(x_c). \quad (58)$$

This situation is illustrated in Fig. 11 for an arbitrary aptitude distribution ϕ. It should be noted that, with reference to this figure, if τ increases (slower pace) the curve labelled $p_r(\zeta\tau_0)$ shifts to the left and is compressed in scale.

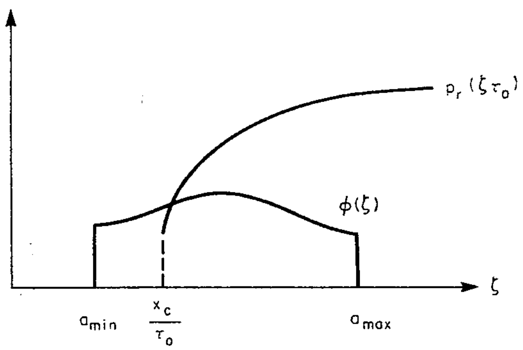

FIG. 11. Illustration of optimal pacing condition for widely spread group.

One characteristic of optimal pacing for a nonhomogeneous group is evident from Eq. (58). If for a particular distribution ϕ, τ_0 satisfies Eq. (58) then τ_0 also satisfies (58) for any other distribution that is the same as ϕ for $a > x_c/\tau_0$ but different for $a < x_c/\tau_0$. Clearly, no part of Eq. (58) depends on $\phi(a)$ for $a < x_c/\tau_0$. Thus, a large family of widely spread distribution functions with the same shape of upper portion result in the same optimal pace (locally, at least). This pace is such that only learners with aptitudes in the upper range have nonzero steady state achievement. Hence, for certain groups with widely spread aptitude distributions, the maximum gain is realized by designing the pace of instruction for only the higher aptitude portion of the group.

It can be shown, using the mean value theorem and the aforementioned properties of the reference function p_r, that, for all sufficiently narrow distributions, the optimal pace is determined by setting the lower expression of Eq. (57) for the derivative of Γ equal to zero. That is,

$$\int_{a_{\min}}^{a_{\max}} \phi(\zeta)\, p_r(\zeta\tau)\, d\zeta = \tau \int_{a_{\min}}^{a_{\max}} \phi(\zeta)\, \zeta p_r'(\zeta\tau)\, d\zeta. \quad (59)$$

In the limiting case, as the aptitude spread of the group narrows, the group approaches homogeneity where all learners have the same aptitude, and Eq. (59) implies that the optimal pace, τ_0, is given by

$$\tau_0 = x_h/a \tag{60}$$

where a is the common aptitude of the learners of the group and x_h is defined by

$$p_r(x_h) = x_h p_r'(x_h). \tag{61}$$

Figure 12 illustrates this condition. The optimal pace for a homogeneous group given by Eq. (60) is clearly the same as the optimal pace for an individual learner with the same aptitude, as can be seen from Eq. (45).

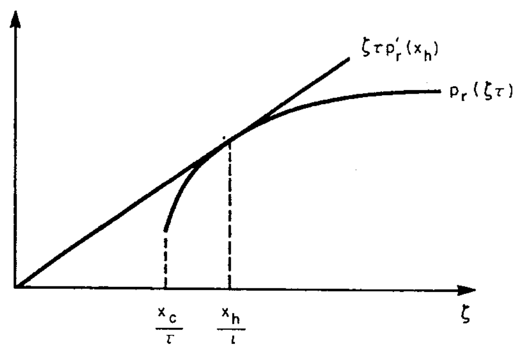

FIG. 12. Illustration of optimality condition for homogeneous group.

The optimal pace for narrowly distributed groups, defined by Eq. (59), has the property that all learners of the group have nonzero steady-state achievement (since $a_{\min}\,\tau > x_c$). This is in contrast with the result above for widely spread distributions. There is also a family of distributions, between the narrowly and widely distributed groups, for which the optimal pace is $\tau_0 = x_c/a_{\min}$. For these groups, the optimal pace is equal to the critical pace of the learner of the group with minimum aptitude.

Two other properties of the optimal group pace have been demonstrated [Chant (1973)]. The first property, which is perfectly intuitive, is that the optimal *group* pace is also the optimal *individual* pace for some learner with aptitude a_t such that

$$a_{\min} \leqslant a_t \leqslant a_{\max}\,. \tag{62}$$

This learner is defined as the *target learner*. For group pacing τ,

$$a_t = (x_h/\tau). \tag{63}$$

The second property involves the measure of the spread of the aptitude distribution. If two groups differ only in that the aptitudes of one are a positive multiple α of the other, then both optimal paces satisfy the same necessary condition [either Eq. (58) or Eq. (59)]. Furthermore, the ratio of the two paces is α.

Pacing for a Group of Learners: Example

In this section, we examine optimal group pacing for the case of a uniform distribution of aptitude. Then, the specific solution is given for the example system introduced earlier in the paper.

The uniform distribution is defined by

$$\phi(a) = \begin{cases} v, & a_{\min} \leqslant a \leqslant a_{\max}\,; \\ 0, & \text{otherwise.} \end{cases} \tag{64}$$

where v is the density of learners per aptitude unit. For a group of N learners, $v = N/(a_{\max} - a_{\min})$.

If the optimal τ_0 is such that $\tau_0 < x_c/a_{\min}$, then (58) applies and

$$\int_{x_c/\tau_0}^{a_{\max}} vp_r(\zeta\tau_0)\, d\zeta = \int_{x_c/\tau_0}^{a_{\max}} v\zeta\tau_0 p_r'(\zeta\tau_0)\, d\zeta + (x_c/\tau_0)\, vp_r(x_c). \tag{65}$$

Cancelling the density factor v and integrating the integral on the right hand side by parts yields

$$\int_{x_c/\tau_0}^{a_{\max}} p_r(\zeta\tau_0)\, d\zeta = a_{\max} p_r(a_{\max}\tau_0) - \int_{x_c/\tau_0}^{a_{\max}} p_r(\zeta\tau_0)\, d\zeta. \tag{66}$$

A change in the variable of integration yields the simpler expression

$$2\int_{x_c}^{x_{\max}} p_r(x)\, dx = x_{\max} p_r(x_{\max}) \tag{67}$$

where $x_{\max}$ denotes the product $a_{\max}\,\tau_0$.

Equation 67 may be interpreted as an equality of areas as illustrated in Fig. 13(a). Denoting the shaded areas A and B as shown in the figure, Eq. (67) implies that

$$2B = A - B \tag{68}$$

or, simply, that

$$B = A. \tag{69}$$

As shown in the figure, the point on the x-axis such that area A equals area B is denoted by x_b. Then, if the optimal group pace is such that $\tau_0 < x_c/a_{\min}$, Eq. (67) implies that $\tau_0 = x_b/a_{\max}$.

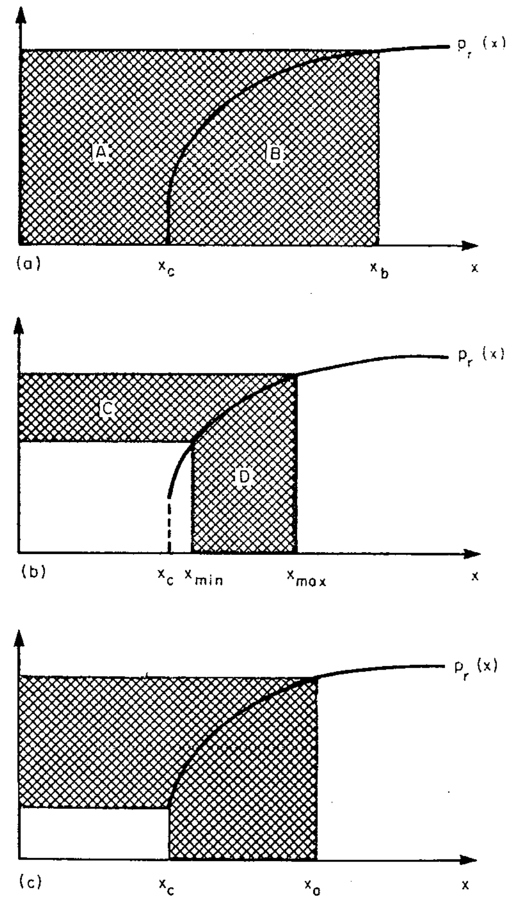

FIG. 13. Area interpretation of optimality condition for group pacing with uniform distribution.

If the optimal τ_0 is such that $\tau_0 > x_c/a_{\min}$, then Eq. (59) applies and

$$\int_{a_{\min}}^{a_{\max}} vp_r(\zeta\tau_0)\, d\zeta = \tau_0 \int_{a_{\min}}^{a_{\max}} v\zeta p_r'(\zeta\tau_0)\, d\zeta. \tag{70}$$

Simplifying, as above, yields

$$2 \int_{x_{\min}}^{x_{\max}} p_r(x)\, dx = x_{\max} p_r(x_{\max}) - x_{\min} p_r(x_{\min}) \tag{71}$$

where $x_{\min}$ denotes the product $a_{\min}\, \tau_0$. Figure 13(b) illustrates Eq. (71). Denoting by C and D the shaded areas in the figure, Eq. (70) implies that

$$C = D. \tag{72}$$

If neither Eq. (58) nor Eq. (59) applies, that is, if the derivative of group gain Γ as given Eq. (57) is positive for $\tau < x_c/a_{\min}$ and negative for $\tau > x_c/a_{\min}$, then the maximum of Γ occurs at $\tau_0 = x_c/a_{\min}$. It can be shown that, with a uniform distribution of aptitude and with the previously mentioned properties of the function p_r, the τ that satisfies the above necessary conditions is the global maximum.

We now examine the relationship between optimal group pacing and spread in group aptitude. We consider the full range of uniform distributions with $a_{\min}$ fixed and with $a_{\max}$ decreasing continuously from a very large value until it reaches $a_{\min}$, corresponding to the homogeneous group case. With $a_{\max} \gg a_{\min}$, the optimal pace is determined by Eq. (67) to be

$$\tau_0 = (x_b/a_{\max}). \tag{73}$$

For very large $a_{\max}$, this value of τ is comparatively small, which implies a fast pace. It is noted that in this case the optimal pace is such that the learners with small values of aptitude are ignored, since τ is less than the critical value for them.

As $a_{\max}$ decreases, the optimal pace continues to be given by Eq. (73), until τ_0 reaches the value $x_c/a_{\min}$. As $a_{\max}$ decreases further, the optimal pace remains at $x_c/a_{\min}$ until Eq. (71) is satisfied. During this transition period with τ_0 fixed and $a_{\max}$ decreasing, the area B of Fig. 13(a) decreases until it becomes equal to area A minus $x_c p_r(x_c)$. The value of x for this equality is denoted x_a as shown in Fig. 13(c). During this transition period the optimal pace is determined by the minimum aptitude in the group ($\tau_0 = x_c/a_{\min}$) and is at the critical value such that the minimum aptitude learner just continues to learn. As $a_{\max}$ decreases further, τ_0 is given by Eq. (71) and, in the limit when $a_{\max} = a_{\min}$, the pacing for the homogeneous group is the same as that for an individual learner with the corresponding aptitude.

In summary, the optimal pace for a group with a uniform distribution of aptitude from $a_{\min}$ to $a_{\max}$ is given by

$$\tau_0 = x_b/a_{\max} \qquad \text{for} \quad (x_b/x_c) \leqslant a_{\max}/a_{\min} \tag{74}$$

$$\tau_0 = x_c/a_{\min} \qquad \text{for} \quad x_a/x_c \leqslant a_{\max}/a_{\min} \leqslant x_b/x_c \tag{75}$$

$$x_c/a_{\min} \leqslant \tau \leqslant x_h/a_{\min} \qquad \text{for} \quad 1 \leqslant a_{\max}/a_{\min} \leqslant x_a/x_c \tag{76}$$

$$\tau_0 = x_h/a_{\min} \qquad \text{for} \quad a_{\max} = a_{\min} \tag{77}$$

where the actual value of τ_0 in Eq. (76) is determined by Eq. (71).

The example system introduced earlier for an individual learner has the steady-state response reference function p_r shown in Fig. 10. For this particular response function, the values of the parameters x_c, x_h, x_a, and x_b are 1.921, 2.151, 2.780 and 4.155, respectively. The optimal group pacing as a function of the maximum group aptitude $a_{\max}$ is shown in Fig. 14 for several values of the minimum group aptitude $a_{\min}$. When $a_{\max}$ is much larger than $a_{\min}$, the optimal pacing is determined by $a_{\max}$ only.

This is evident in Fig. 14 since all curves for various values of $a_{\min}$ are coincident on the hyperbola $a_{\max}\tau = x_b$ for large values of $a_{\max}$. As $a_{\max}$ decreases, there is a range where the optimal pacing does not change, being determined by $a_{\min}\tau = x_c$. For values of $a_{\max}$ close to $a_{\min}$, the optimal pacing is determined by Eq. (71).

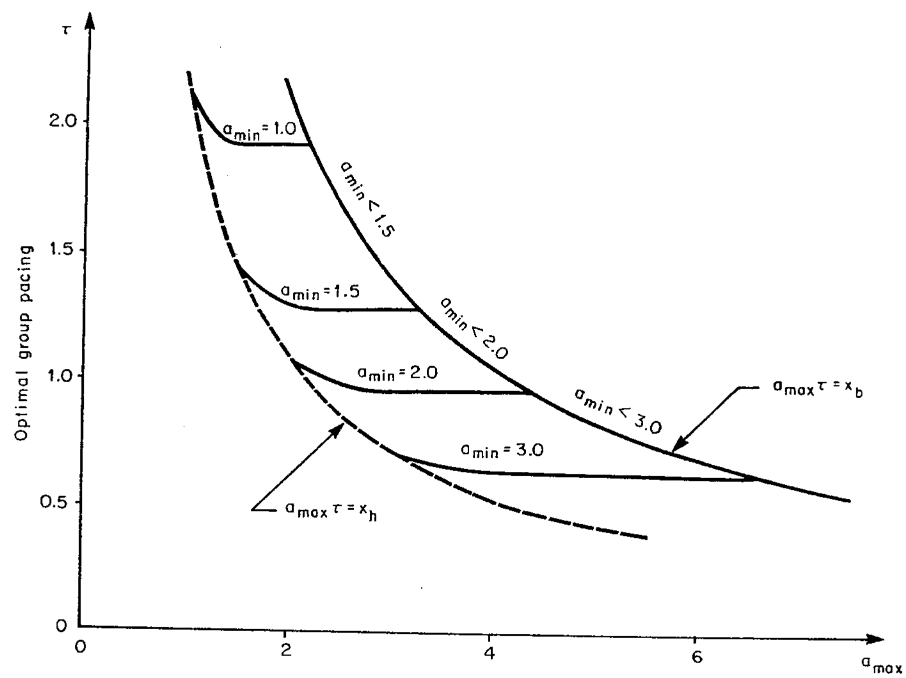

FIG. 14. Optimal group pacing versus $a_{\max}$ for several values of $a_{\min}$.

Concluding Remarks

The objective of this work was the development of a meaningful, useful theory of instruction based on psychological learning theory but with application in the educational environment. To this end, the basic model of the instructor/learner interaction was developed from learning theory concepts. This basic model was extended to account for instruction pacing, an important concept of group instruction. The essential trade-off, in the educational setting, of achievement and cost was formulated within the framework of optimization and control theory in order to examine optimal instructional inputs.

For the individual learner situation, the theory shows that intensity of instruction should remain constant throughout the learning encounter. This constant instructional input is optimal under quite general conditions including, for example, the case where learning rate changes as more material is learned. For the group learner situation, the theory shows how instruction should be paced depending on the diversity of learning rates of the members of the group. For example, for certain diverse groups, it may be optimal to design the pace of instruction so fast that the slower learners cannot stay with the group. The theory also has implications for empirical research.

For education researches, the theory points out one way in which learning theory can be used to structure some of the relationships among the variables in the educational environment. This structure then suggests which variables of the system are important and require careful study. For example, the relationship between instruction cost and its effect on learning rate, and the relative value of instruction time and learner time should be empirically determined.

For psychology researchers, the theory identifies several hypotheses that could be subjected to detailed experimentation. The concepts of instruction pacing and critical pace could be investigated by studying the relationship between learner steady-state response and instruction time per block of material. The definition of aptitude and its usefulness in specifying the nonhomogeneity of groups of learners requires further study. The equivalence, as implied by the assumptions of the theory, between aptitude and learning time over a range of related learning material should be straightforward to examine.

The theory can be extended to include other factors of the learning phenomenon so as to apply to other problems in the educational environment. An important question is the trade-off between learner achievement and instruction cost when there is the possibility of having a mix of group and individualized instruction. The treatment of this question would have to examine the effect of the size of the group of learners on the individual learner's performance. It is possible that with individualized instruction available, the pace of instruction for the group or the size of the group could be increased.

References

Carroll, J. B. School learning over the long haul. In Krumboltz, J. D. (Ed.), *Learning and the educational process*. Chicago: Rand McNally, 1965.

Carroll, J. B. A model of school learning. *Teachers College Record*, 1963, **64**, 723–733.

Chant, V. G. A mathematical theory of instruction with application to instruction pacing and class size. Ph.D. Dissertation, Department of Engineering–Economic Systems, Stanford University, Stanford, California, December 1973.

Luenberger, D. G. *Optimization by vector space methods*. New York: John Wiley and Sons, 1969.

Thurstone, L. L. The learning function. *Journal of General Psychology*, 1930, **3**, 469–493.

Received: January 7, 1974

Printed by the St Catherine Press Ltd., Tempelhof 37, Bruges, Belgium.