DOCUMENT RESUME

ED 206 681 TM 810 578

AUTHOR	Page, Ellis B.
TITLE	Operations Research as a Metaphor for Evaluation. Research on Evaluation Program Paper and Report Series.
INSTITUTION	Northwest Regional Educational Lab., Portland, Oreg.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	NWREL-REP-15
PUB DATE	Jan 79
NOTE	44p.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Algorithms; *Educational Assessment; Elementary Secondary Education; *Evaluation Methods; *Models; *Operations Research
IDENTIFIERS	*Decision Analysis; *Decision Trees

ABSTRACT

One of a series of research reports examining objective principles successfully used in other fields which can lend integrity and legitimacy to evaluation, this report presents an overview of operations research (OR) as a potential source of evaluation methodology. The nature of the methods common to this discipline are summarized and the possibility of employing these methods in educational evaluation (EE) are investigated. The author points to the idea that OR is the science of evaluation since it directly addresses decision making. In demonstrating that EE is targeted on decision making, feasable alternatives, outcome predictions and probabilities for each alternative, and estimation of costs and benefits are identified as prerequisites of EE. Decision Analysis, a subfield of OR, is then provided as a model wherein all the concerns of EE are made explicit, and an optimal solution located by organizing them into a single algorithmic structure known as a decision tree. Other transportation, network and simulation models are also applied to hypothetical EE problems. After reviewing some of the literature and the current situation in the fields of OR and EE, the author concludes that EE is one of the few fields not to have taken advantage of OR's potential. (AEF)

INTERIM DRAFT

Do not cite or quote without author's permission.

Author welcomes reactions and suggestions.

No. 15 OPERATIONS RESEARCH AS A METAPHOR FOR EVALUATION

ELLIS B. PAGE

University of Connecticut

January 1979

Nick L. Smith, Director
Research on Evaluation Program
Northwest Regional Educational Laboratory
710 S. W. Second Avenue, Portland, Oregon 97204

Published by the Northwest Regional Educational Laboratory, a private, nonprofit corporation. The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

PREFACE

The Research on Evaluation Program is a Northwest Regional Educational Laboratory project of research, development, testing, and training designed to create new evaluation methodologies for use in education. This document is one of a series of papers and reports produced by program staff, visiting scholars, adjunct scholars, and project collaborators--all members of a cooperative network of colleagues working on the development of new methodologies.

A collection of special reports has been prepared, each of which overviews a separate discipline as an alternative source of methodology for evaluation. Each report summarizes the nature of the methods common to that discipline and examines the possibilities of employing such methods in evaluation. This report, one of that collection, presents Operations Research as a potential source of evaluation methodology.

Nick L. Smith, Editor
Paper and Report Series

CONTENTS

	Page
The Corollaries of Decision Making	3
An OR Model for Decision Analysis	6
Role of Subjectivity in Decision Analysis	8
Estimating Values for Decisions	9
Other Subfields of Operations Research	12
Transportation Models	13
Network Models	16
Simulation Models	17
The OR Models as Metaphors	19
Operations Research and Education	22
References	28

OPERATIONS RESEARCH AS A METAPHOR FOR EVALUATION

A favorite colleague, given to mischief-making, has described evaluation as "50 percent measurement and 50 percent mush." In that part which is not measurement, he claims, there is little intellectual coherence, and no scientific respectability. Regarding evaluation in its present state, one finds it hard to disagree (Page, 1975, 1978). However, the argument of this chapter is that the "mush" may have made us ailing, but not terminally ill. Indeed, there are objective principles, of the sort which have led to the success of measurement and of most scientific activity, which can be applied through the field of evaluation, and can give it an integrity and legitimacy which are only fantasies today.

These saving principles are those of Operations Research (OR), a brilliantly unfolding discipline less than 40 years old. Probably not one percent of today's evaluators, even those with extensive statistical knowledge, know OR with any depth. Yet this discipline addresses exactly those problems closest to our hearts, those of decision-making. Indeed, operations research is the science of evaluation.

This article can only be a brief overview of OR, and how it relates to evaluation. First, we shall demonstrate that educational evaluation (EE) is indeed targeted on decision-making. Second, we shall consider how one subfield of OR, decision analysis, makes explicit and tractable the elements of our concerns. Third, we shall illustrate some of the other subfields of OR and possible applications in EE. With this preparation, we shall see how OR provides a metaphor for EE, in the most useful and powerful sense of metaphor. And finally, we shall consider ways in which workers may draw the two fields of activity closer to each other.

The Corollaries of Decision-Making

Clearly, evaluation's claim to uniqueness, to being usefully beyond measurement, depends on its being an aid to decision making. While not stated explicitly in much of the evaluations literature, this is at least implicit in much of the writing where EE spokemen attempt to differentiate their field from the older one of measurement (for a review of six sample texts, see Page, 1975). Evaluation is not the same as research, we are told, because EE is "concerned with the individual case," aims to "aid the planner, administrator, and practitioner," and indeed "aid in making decisions." For such purposes we "evaluate" programs, students, proposals, products, , and personnel -- all with the aim of making, or at least influencing, decisions about these subjects of our inquiry.

Having acknowledged this relation, let us see whether we can deduct some corollaries of it.

1) Any decision must necessarily concern itself with choice among alternatives. No proper "evaluation" can be of just one object. This point is curiously fuzzy in most of the evaluations literature, but it does occasionally appear. For example Scriven (ch. 1 in Popham, 1974):

> Few if any useful evaluations avoid the necessity to present data on the comparative performance of critically competitive products. All too often the data refers to some pre-established standards of merit, and the reader has no idea whether one can do better for less, or twice as well for 5 percent more, which is

the kind of information a consumer wants. . . .

Let us continue . . . this passage, for it will illuminate some other features of our decision-making:

. . . . It is not too thrilling to discover that an injection of $100,000 worth of computer-assisted instruction (CAI) can improve the math performance of a school by 15 percent if there is a possibility that $1,500 worth of programmed texts would do as well or better. There are few points where good evaluators distinguish themselves more clearly than in their choice of critical competitors. [p. 15]

One corollary of EE as decision-making, then, is the need for such a list of <u>feasible alternatives</u>.

2) A second feature of any decision making is <u>some value which may serve as a goal</u>. In the above example it is to "improve the math performance of a school." It is not clear how any decisions may be intelligently made if the process is truly "goal-free." Indeed, it is often required that such objectives be measured, as they are above, in a <u>ratio scale</u>. How else is one to interpret an improvement of "15 percent"? Clearly, then, to decide among alternatives, we need such a prediction of outcome <u>for each alternative</u>.

3) A third feature of decision making is an estimate of <u>costs</u>, accompanying each alternative. In the illustration above, costs are in terms of dollars ("$100,000 worth of computer-assisted instruction" vs. "$15,000 worth of programmed texts"). If costs are the same, then we choose in terms of value; but if values are the same, then costs become essential. But in the "real world" so prized by evaluators, some information about

both costs and benefits is essential to making defensible decisions.

What is still less acknowledged in EE, but is implicit in the above example, is the need to measure costs and benefits on the same scale. Suppose we can improve a school's math performance "10 percent" by spending an additional "$20,000" -- should we do so? Clearly, we need, for many non-trivial questions, some way of plotting a function of costs against benefits. (This is a balancing we perform all the time in real life: Else how would we decide to pay $3 for a movie?)

4) And still a fourth corollary of EE as decision making is this: We must have some knowledge of the probabilities of various outcomes for each alternative. If we choose A or choose B, that is, we need to anticipate the "chances" of various results, with their associated costs and benefits. If we know these certainly, it greatly simplifies our choice. But in the much vaunted "real world" of EE, knowledge about such probabilities must be wrung from Nature's close reserve, must be extracted like juice from any high-quality evidence we have.

And here is a rather startling, though unavoidable feature of these corollaries: These outcomes, both costs and benefits, must be related causally to the decisions. Here correlational data, collected from artifactual records either in this school or elsewhere, will not suffice. For no decision making is rational unless it bears on producing the more desirable outcome. So saying, we have prepared the way for models of Operations Research. Even more: We have made EE heavily dependent on traditional, knowledge-aimed research, for it is from true research that we can best hope to understand the required causal nexus.

An OR Model for Decision Analysis

Having framed our requirements for decision making in evaluation, we now turn to a model from O.R. to show how all these requirements are confronted and organized into a single structure. Decision analysis (DA) is only one of the models in OR. But it seems the easiest to grasp for intelligent educators unused to such thinking; and it serves as a kind of prototype of the more specialized algorithms described later in this article. It is the only model shown here in sufficient detail to be calculated. Its main features are illustrated in Figure 1.

Figure 1.

Figure 1 represents a simple decision tree containing the following structure: There is a set of nodes and branches, according to how the decision problem is envisioned by the designer. There are just two kinds of nodes: rectangular nodes are decisions, and the branches from a decision node are the alternatives among which we must choose. The circle nodes represent probabilities, or uncertainties as to outcomes. The branches from a circle are the possible events of interest. Each branch from a circle node must be flagged by some estimated probability, and these probabilities must add up to one for any circle node. At the end of each terminal branch (whether from rectangle or circle) must be some (black dot) value attached to that outcome, or plan. To complete the notation for this simple tree, we have the little toll-gate, near the letter B, where we show the estimated <u>cost</u> of that branch. Costs and values, as we have noted, must be transformed into the same units of measure, if both are necessary to include in the problem.

Let us take such a tree as representing a true picture of our decision problem, and the probabilities, costs, and values as being a reasonable approximation to their true measures. Then -- to the astonishment of newcomers to DA -- such a tree is solved entirely automatically, algorithmically, reaching decisions which are demonstrably optimal. We begin at the bottom of the tree with the terminal values. And we move up the tree by purely automatic procedures. If the node above is a probability node, then we multiply each value by its probability, thus calculating a mean value which becomes the value now attached to the node itself. For the probability (circle) nodes, these calculations are shown in Figure 1.

On the other hand, where the node is a decision (square) node, we "fold back" the branches of less value, indicating this by the double barrier as seen near the letter A. And we attach the highest branch value to the node itself. We work up the tree this way, from bottom to top, with these two simple operations of averaging and folding back (and subtracting costs). And in this simple, recursive way, we can solve a tree of any arbitrary complexity. In this illustration of Figure 3, if we were considering Plans A and B, we would automatically choose B, since 70 is greater than 60.

The curious thing about this model -- about our neglect of it in education -- is that there is <u>no alternative theory of decision making</u>.

That is, there are other OR models, but there is no rational way to decide any matter but by consideration of the variables in this tree: probabilities, costs, and benefits. And given these elements of the problem, then such a mathematical solution, as we have said, is demonstrably equal to or better than any subjective solution. (An important but neglected area of research, by the way, is the degree to which we, as human deciders, can optimize when we subjectively solve trees of this kind. But see Tillett, 1975.)

Role of Subjectivity in Decision Analysis

We must note, in Figure 1, that there are many aspects of the tree where we commonly lack firm data: often those very probabilities, costs, and outcome values. Then we must depend on judgments. These judgments, considered individually, are inevitably going to include large subjective components. But such inclusion does not render the approach unscientific. To the contrary, all science includes intersubjectivity as a cardinal test of its status (Feigl & Brodbeck, 1953, ch. 1). As Feigl points out, intersubjectivity is in fact another term for objectivity: the agreement at the most elementary level about whether a simple event did or did not occur. And one of the commonest concerns in social science is the "reliability" of measures, which in the case of more complex judgments, such as those here considered, refers to the concordance among judges acting independently. The difference between ordinary subjectivity and what I call judgment is the difference between unchecked, unverifiable individual experience -- a kind of solipcism --

and the central, desirable intersubjectivity of science. A virtue of decision science is that it tames subjectivity, giving it whatever role is necessary, but estimating its quality, making sure it is representative, and using it carefully in algorithms which are themselves objective. In a decision tree such as Figure 1, where does such judgment enter? into: 1) the design of the tree, since there may be more than one set of alternatives to consider; 2) the estimation of costs, using whatever data are available; 3) estimation of probabilities (though regression techniques may provide helpful estimates); and centrally, inescapably, 4) the estimation of values.

Estimating Values for Decisions

It is a curious characteristic of the "evaluations" literature that it has paid so little systematic attention to values. A great virtue of the OR approach is that it forces us to remedy this lack, and to make explicit some scale of worth. For a prime requirement of OR is some agreed-upon objective function, "measurable values . . . that unequivocally reflect the future well-being of the organization" (Wagner, 1969, p. 5). At first, we seem to lack any such "measurable values," once we leave the (deceptively) simple measures such as dollars of profit, or miles of travel saved. Yet we routinely analyze ratings, rankings, grades, and other judgments in our literature, and there is no needed epistemology which is not already a well-worn tool of behavioral science.

Also, contrary to what many think, there is not an unbridgeable gulf between the highest philosophical values and the most discrete test items or behavioral objectives. At Connecticut, intrigued by this problem, some of us have done work on what we call the bentee for "benefit T-score."

We have explored what we may term a "top-down tree of value" (Page, 1972, 1974; Page & Breen, 1974a, 1974b). A sample tree is shown in Figure 2.

Figure 2

With Figure 2, we can grasp the remarkable, recursive property of a value tree, and explore some of its mathematical properties. We define the bentee, at the topmost level, as including 100% of the value to be gained from an educational experience (in this illustration, the full elementary and secondary experience, the bentee being for the senior graduating from high school). We have divided this bentee into 7 areas of educational improvement (Verbal, Quantitative, etc., as shown), and we have apportioned the values according to the judgments of 101 randomly selected judges, half laymen and half professional educators. The task appears very easy for the judge to do: We simply present him/her with a list of these 7 traits, together with a paragraph description of each. These are presented in a new random order for each judge, with a sheet of paper with the 7 areas marked, and a stack of 100 chips. The judge is asked to "spend" the chips.

The method is easy and quick, and though individual judges may be wildly deviant, the group estimates rapidly reach stability as sample size increases. When we have such estimates, we can then make calculations which were formerly impossible. For example, if we have a group of high-school seniors, and standardized measures or estimates for them on each of the 7 traits, we may calculate for a student an overall bentee, by weighting each of his 7 estimates appropriately, by summing these weighted estimates for him, and by transforming this sum to a T-score (with a mean of 50 and standard deviation of 10). Such bentee scores would now permit us to rank order the students for a variety of purposes of selection, guidance, and award.

The most remarkable feature of the value tree is its generality: We can use the same techniques at any node in the tree. We can define a new bentee for, say, American literature, collecting expert opinions about the apportionment of value, building curricula and tests in accordance with these values, and assigning overall scores as the weighted sum of sub-tests. Note, this is quite similar to what is done in curriculum and test construction today, but the value-tree approach establishes a new legitimacy to the outcome, since values are <u>defined</u> in a defensible way.[1]

The tree has additional mathematical properties which permit us, in principle, to weight any node within the overall scheme of values by rules adapted from a tree of independent probabilities. (Cf. Page, 1974; Page & Breen, 1974a.) And it may be used for virtually any purpose or sub-group: third-grade pupils, learning-disabled teen-agers, pre-med or engineering majors. The tree is particularly useful when facing the problem of tradeoff: for example, where one gains in math at the expense of time taken from social studies, since it permits evaluation of people, groups, institutions, or programs in terms of <u>overall</u> educational effects.[2]

Entirely different problems of value, of course, are raised when we attempt to trade off values against costs, when these are measured in different units: for example, bentee points vs. hours of teacher time; or anytime we are faced with more than one kind of scale.

But such questions, too, have come under expert scrutiny by some workers in OR. Some issues were attacked at least two decades ago (e.g., Churchman, Ackoff, & Arnoff, 1957, ch. 4), and have been analyzed on a deep philosophical level by Churchman (1961). Some important questions of value function were considered by Raiffa (1968, pp. 51-101) in his monograph on DA. And the most thorough mathematical treatment given value tradeoffs is probably the massive work by Keeney and Raiffa (1976), which should provide much stimulation and insight for behavioral researchers. For our purposes here, we can say that such problems are not intractable; scales can be drawn together by interview and experimental techniques, to establish the tradeoffs required, in order that decision trees will, indeed, provide a framework helpful to the decision maker.

And the most important realization of this section, for those interested in EE, is that these complexities of value were not invented by OR. Rather, they reside in the real problems facing educators; and OR appears the only discipline prepared to make them explicit and to balance them off in an intelligible way.

Other Subfields of Operations Research

Suppose, now, that we are advising administrators on a different problem, the assignment of teachers to courses. We can assign each teacher only to a certain number of courses, and each required course in the curriculum must be taught. We have a specified goal: We wish our assignment to optimize the sum of ratings given by department heads to each teacher/course combination.[3] How should we design our OR model? We could set out to design a decision tree, with every possible teacher/course combination complex having a branch of its own. But the number of such branches would become extremely large. For example, if there were just 5 teachers and 20 courses for them to cover (each teaching four courses), then there are 300 billion possible assignments![4] DA is obviously a poor model, then, for such an assignment problem, and we turn to the broader field of OR for a more suitable choice.

Transportation Models

For such problems, a suitable framework is an overarching one called a transportation model. Such a model provides one row for each "supply point" (in this case, a teacher) and one column for each "destination" (in this case, a course). Instead of generating billions of possible assignment tables, then, we are concerned with just a 5 x 20 table (in this sample problem), and a limited number of iterations over this table. Or if there are, say, only 7 unique courses, then the table is only 5 x 7 in this small example. Such a reduced table is illustrated in Figure 3.

Figure 3

As we note, for each teacher/course combination, there is a benefit r_{ij}, here a rating of quality.[5] For each cell in the Figure, there will also be an x_{ij} with values ranging from zero (when the teacher teaches no sections of that course) to 4 (if, say, one teacher teaches all four hours available in Course #5). The aim of the assignment algorithm, then, will be to maximize

$$\sum_{i=1}^{5}\sum_{j=1}^{7} x_{ij}r_{ij} .$$

This maximization is carried out under the constraints that each teacher instruct the number of agreed hours, and that each course section be covered by an instructor.

An expert example of such analysis is provided by Tillett (1975). In a real-world, serious study of such assignments, there are of course many considerations beyond those above, and Tillett adjusted his model for these: For one thing, neither ratings nor preferences are alone enough to make out a good assignment. Both must be considered and balanced appropriately, sometimes in a way different for each teacher (depending, for example, on seniority). This can be neatly done within the same general algorithm (p. 102). For another, teacher preferences are complicated by the number of sections assigned of a course (I may like to teach 9th grade algebra, but four sections of it.... ?). Tillett therefore extended his model to a third dimension (the number of sections of each ij combination). And in order to make sure he obtained feasible solutions, he abandoned the fast algorithms for one called zero-one integer programming. Obtaining data from 7 Connecticut high-school Math departments, he compared the optimal solutions for each, against the current schedules actually being taught. His results are shown in Tables 1 and 2.

Table 1

Table 2

Naturally, since the algorithm does optimize the objective function, there is never a case where the solution does not meet or surpass the current assignment, when only one dimension of benefit is considered. Thus; when preferences are maximized in Table 1, the preference gain is very clear. When we compare these preference solutions with the effectiveness ratings; however, we find there has been some tradeoff, and the existing schedules are at times superior. And we see the same result, in reverse, when we maximize only the effectiveness ratings. (Tillett did not apply the available techniques for combining the two benefits in appropriate ways.) The clearest message from such work, then, is that our choice of value dimensions is of fundamental importance to our decision-making.

Now suppose an evaluator is called on to help make assignments of this general sort. Clearly, the OR framework can enormously aid in knowing what information to collect, what opinions and preferences, and how to combine them into a feasible and defensible solution. There is also a side benefit from nearly all models of operations research: After finding one or more optimal solutions, one can do sensitivity analysis to find how stable it is under shifting estimates of benefit. And one can fairly easily calculate alternative solutions, if the automatic one reaches produces some infeasible feature not anticipated in the design. It helps us ask questions, as well as make use of the answers.

Network Models

Another big class of OR models can be adapted for large, unknown number of educational problems. These are the network models, which are probably most familiar to educators as PERT diagrams (e.g., Cook, 1966). But there are countless other potential applications. For example, let us assume a question of curriculum design. We have a certain amount of time, a certain amount of material to cover (which we cannot do justice to), and estimates of cost and value for different levels of effort for these materials. How can OR help us sort out all this information, and come up with the best possible design?

For simplicity, let us assume we have just 10 hours of study time, to get through four chapters of a textbook. For each chapter we have a set of alternative levels of effort, with a cost and value estimated for each level. If we think of beginning at one end of a network, going through the four chapters, and ending at the other end of the network, we might draw the problem as in Figure 4.

Figure 4

Now we may think of our problem as picking our way through a maze of paths. Here we are not just picking the shortest path, as we might in planning a trip; and we are not just picking the longest path, as we might when anticipating difficulties in PERT. Rather, we are picking that path which produces the most <u>efficient</u>

trip, gaining us the best ratio of benefit-to-cost, possible within our set limitation.

And we notice a peculiarity of the network of Figure 4. There are five successive stages of the trip, and there is no line, for example, directly connecting Chapter 1 with Chapter 3. In fact, this sort of problem is therefore nicknamed the "stagecoach" problem (Wagner, 1969, p. 256), and it permits much easier solutions than if all nodes were connected with all others. Because the model is solved by moving from one stage to the next, solving each in turn, it is classified as a "dynamic programming" algorithm.

Network problems, like others in OR, allow us to test the sensitivity of the solution to errors in estimate, and to try out various alternative values in the input. Here again, we have a framework which helps us ask the right questions, and fit the answers into a rational set of suggestions, together with some estimate of possible errors in those suggestions.

Simulation Models

In seeking such legitimate approaches to evaluation, we have briefly considered three major kinds of OR models: decision analysis, transportion, and network designs. These and others have in common the enormous benefit of providing "optimal" solutions, either maximizing or minimizing some function of agreed worth. But there is a large class of OR models under the general head of "simulations", which permit us to search out a large number of possibilities, but which do not automatically home in on the best alternatives.

Simulation models, in fact, are so various as to be almost indescribable, since they include the vast quantity of OR problems which cannot be analyzed by one of the standard optimization models. Simulation is very valuable: We often wish to test a plan without the enormous dangers and costs of a real-world trial, so we set out the way the plan would be expected to work, often in the form of a flow-chart. We also set out some statistical features of the events we expect to encounter, often using random-number generators to simulate their occurrence. Now we write a computer program (probably using some of the simulation packages available), and run a large number of cases, tabulating the results in summary ways. We gain whatever wisdom we may gather from, say, "experiencing" 10,000 such cases.

Let us consider a more concrete example. We wish classes to make full use of the new school library being planned, but we are concerned about overwhelming the librarian with check-outs, possibly leading to tardiness, confusion, poor discipline. Such overloads are random events, but we have some knowledge about the mean and variance of check-out demand for classes. Now there is a particular large branch of simulation, so well developed and that it is often treated separately: queueing theory. By the use of certain formulas in this theory, we are able to "try out" several plans for the library, finding for each how often we get into difficulty, and how bad the difficulty might be.

Like the other OR models, simulation encourages us to think hard about our problems, identifying key features, guiding our collection of information, setting up our algorithms. When we run the problem, we get cheaply what might otherwise have cost very dearly. But whatever the model, the solutions are only as good as the match of the model with the world.

The OR Models As Metaphors

The argument of this article is that decision-making is the most important distinction of educational evaluation, and that Operations Research is the science of decision-making. Thus, OR models would seem to be our preferred metaphor for evaluation. But the introduction of this term should be justified. One scholar recently wrote that "no systematic way for attacking metaphors exists at this time" (Guba, 1979). I believe, to the contrary, that some ways of attacking metaphor (or at least its first cousin, analogy) have been brilliantly synthesized in recent years. My preferred examples are all from the new science of artificial intelligence (e.g., Slagle, 1971). For example, Thomas Evans (1968) has created an automatic, rule-driven analyzer of geometric analogy problems of the sort seen in Figure 5.

Figure 5

His program ANALOGY, even a decade ago, passed tests with the metaphoric insight of an American high-school student.

In the more directly verbal realm, Reitman (1965) programmed a solution algorithm to the classic sort of word analogy:

HAND : GLOVE :: FOOT : (TOE, FINGER, SHOE, LEG).

In this problem, most would agree that the third choice, SHOE, is the correct response. But why? Apparently because both sides of the equation will then satisfy a relation of the form:

CLOTHING(HAND, GLOVE); and

CLOTHING(FOOT, SHOE); or more broadly,

<u>Rxy</u>,

where <u>R</u> is a specified relation, and <u>x</u> and <u>y</u> are ordered arguments which satisfy that relation.

Such relations are at the heart of metaphor analysis. Indeed, even geometric analogies (as in Figure 5) are translated into complex list-processing strings (using LISP) which express many properties of forms, and relations among them, and permit pattern recognition by comparisons among such strings. Even visual metaphors, then, are converted and solved at the level of symbolic logic. [6]

In evaluation, we are seeking metaphors which will provide the best springboard to our successful performance. We seek comparisons of a real-world problem (let us say, <u>Rxy</u>) with some other problem (say, <u>R'ab</u>) which will serve as a useful model. We are applying to our problem an implicit equation of the form:

$$Rxy \stackrel{?}{=} R'ab \ .$$

There are two sources of difficulty: 1) whether we have a clear idea of the way the model itself works: <u>R'ab</u>; and 2) whether there is indeed a useful match of the two relations, that is, whether the model is usefully isomorphic with the real-world problem.

On the first point, OR seems brilliantly ahead of any other metaphor of decision-making. OR has the enormous advantage of all mathematics, of bringing to bear a system of demonstrable theorems, of certain truth. Virtually no other, non-mathematical metaphor can hope to provide the intellectual anchor, then, that OR provides. Compare its working model of decision analysis, for example, with such intrinsically ambiguous metaphors as history or journalism.

On the second point -- the fit of the model to the real-world problem -- OR shares some of the weaknesses of other metaphors. That is, one must labor to establish that R really is usefully equivalent to R', and that x and y are usefully analogous to a and b of our OR model. As Norman Campbell put it,

> It is never difficult to find a theory which will explain the laws logically; what is difficult is to find one which will explain them logically and at the same time display the requisite analogy. . . If it were found that the analogy was false it would at once lose its value. (Feigl & Brodbeck, 1953, p. 289)

All metaphors, then, share the burden of proving their own accuracy. But even here, OR metaphors seem to enjoy unusual status, since they routinely provide algorithms for describing the real-world problem (for example, by requiring lists of alternatives, and the values of probabilities, costs, and benefits) in ways which will at least probe for any falsity in the analogy.

When these two advantages are put together -- the power of the mathematical model in itself, and its heuristic power in exploring the dimensions of fit to the real world,- then OR seems uniquely suited to provide the most useful metaphors for educational evaluation, with its attendant decision-making.

Operations Research and Education

This final section is intended as a guide to the literature of OR, especially of those topics most salient for educational evaluation. A masterful, prizewinning summary of the field is the 1000-page work by Wagner (1969). This is at once a huge textbook and a handbook for professional workers in OR. There are very numerous exercises and word problems, quite a few with a behavioral or educational flavor. Other comprehensive texts in OR are by Hillier and Lieberman (1974), and by Gaver and Thompson (1973). All of these have many problems, and would be suitable for courses of two semesters or more.

For those seeking a particular emphasis on decision analysis, there is an OR text by Trueman (1974), which devotes six chapters to such decision trees. And if one seeks a combination of DA with our more classic statistical approach, there are statistics textbooks (without other OR techniques) by Hamburg (1970) and by Winkler and William L. Hays (1975), the well-known psychological statistician. It seems certain that more statistics texts will move in the direction of DA, as Bayesian approaches become more popular (e.g., Novick & Jackson, 1974), since there is some affinity between the two areas. And for a specialized introduction to DA, there is still no better work than that of Raiffa (1968).

As we have noted, psychologists and educators have been slow to discover OR, but there are some green shoots of interest. Some basic OR concepts have filtered into education, especially in flow-charting (e.g., Kaufman, 1972), PERTing (Cook, 1966), or PPBS (Hartley, 1968). These have typically used graphic aids from OR, but seldom the advanced mathematical techniques, let alone optimization, which are the core of OR as a professional discipline. In fact, ~~there are~~ at least two introductory books for educators (Banghart, 1969; Van Dusseldorp, Richardson & Foley, 1971), but these are clearly aimed at familiarization with some of the concepts, and not at a level of useful expertise. A specific OR textbook for educators and other behavioral scientists, then, seems a gap waiting to be filled.

Educators may not realize just how advanced and commonplace OR is, in other fields than ours. Courses exist in various university departments, including statistics, engineering, industrial administration, and sometimes departments of their own. Ph.D.'s are given in many universities. In the U.S., large and active professional groups are focused in major organizations such as the Operations Research Society of American (ORSA) and the Institute of Management Sciences (TIMS). These put out major technical journals, Operations Research and Management Science, and team up for the more applied TIMS/ORSA Interfaces, and for large national and international congresses, with programs running over 300 close pages. OR is at the heart of much serious thinking in government, industry, and the military. And in Schools of Business Administration, the typical MBA may routinely take two courses focused on OR approaches. Indeed, it seems to be that Education is the discipline which is out of step in OR (or which has hardly begun to march!).

Yet there have been a number of efforts to apply OR approaches to the understanding of educational problems, and reviews of some of these have been written by McNamara (1971) and by Johnstone (1974). Much of the work has been done in higher education, since OR specialists, with the necessary abilities and skills, have most frequently turned to nearby problems in their own universities.

It is clear that OR techniques have flourished most in fields where there are easily available objective functions, such as time saved, or dollars profit, or hits on target. Our educational objectives, however, are intrinsically less tractable; it is a remarkable achievement of the past half century, that we have (in Western nations) reached some fragile consensus about the importance of scoring well on objective tests. And such measures are an indispensable basis of good educational decision-making (for a technical, early approach, see Cronbach & Gleser, 1965).

Despite the remarkable, incremental progress of educational measurement, there are still problems in using test scores as objective functions for OR. One of these is that a test score, at best, is of interval-scale measurement, whereas for many OR applications we need the strength of ratio scale, that is, using a known zero point. One solution is to concentrate on the change of test scores, since a zero point of a sort is established for no-change. Yet there are psychometric problems, even beyond the relative instability of measurements of change. Here again, is a promising field for serious research.

Another problem of test scores, as we have seen, is putting them in proper perspective as objectives of the educational enterprise. Here we have noted some strong basic work in multiple values (Raiffa, 1966; Churchman, Ackoff & Arnoff, 1956, ch. 6; Churchman, 1961; and especially Keeney & Raiffa, 1976). More applied work on this general problem has been cited (Page, 1972, 1973, 1974; Page & Breen, 1974a, 1974b). The bentee methods have also been applied by Wayne Martin (of the National Assessment of Educational Progress) and by Dr. William Streich (of the Farmington, Conn., Public Schools). Other major work on the values of objectives has involved the "Delphi" technique (Dalkey, 1969), which has aimed more at the creation of consensus among a group than at the discovery of independent opinion. None of these have explicitly used values to combine test scores into single measures (such as described in the bentee literature), but the potential is clearly present.

Apart from test scores, of course, is the general use of independent ratings of value, on a ratio scale. Once these are done, then OR techniques may be applied to optimize such values. And this has been done in curriculum design, as we have seen (Page, Jarjoura & Konopka, 1976; Page, 1976; Konopka, 1977), in teacher assignment (Tillett, 1975), and most recently in the curriculum planning and scheduling for high school students (Grandon, in press). In all of these, however, the "applications" have consisted of the transfer of the general theory to a specific area, and the discovery of opinions, and the design and running of appropriate programs to optimize the objectives, given the data. The applications have not involved the carrying out of plans generated by OR. Such complete applications remain for the future.

Will educational evaluation turn toward Operations Research, as evaluators become more institutionalized, and the evaluative discipline takes on a more structured form? There is surely a movement in this direction, among at least a few of the more quantitatively sophisticated in the evaluations movement (e.g., Edwards, Guttentag & Snapper, 1975; Levin, 1975; Carnoy, 1976 -- and Gene Glass in his introduction of Carnoy). But the movement is still very slight. And there is a strong counter-movement, subjectivist and anti-technological, also among major opinion leaders in evaluation (e.g., Stake, 1978). And we can point to other applied fields in education (such as guidance) where there has been steady slide <u>away</u> from the technical use of evidence, more and more toward a "soft" epistemology, doctrine accepted on faith, and an ever heavier reliance on "authorities" in the field.

In educational evaluation, we may have a rather similar situation. We have a strongly entrenched administrative leadership, all having risen to their positions without any acquaintance with OR. OR is little taught in Departments of Educational Administration. When we turn to research specialists, we have those who are, in Education, best prepared to understand the material. Yet there is still a major obstacle for them to overcome, as well; for hardly any part of their training, neither analysis of variance nor multivariate analysis, will have prepared them to read the material easily. Which brings us to the single major problem with OR and EE: It is hard. Even for measurement specialists, it requires months of new effort to grasp the fundamentals. For the vast number of EE workers, who emphasize verbal, personal, and political skills far more than technical ones, OR, as an advanced practicing specialty, may be hopelessly out of reach.

Ernest Anderson (1970), recognizing the problem of OR's interface with education, asked about "The Little Man Who Isn't There":

> * the Operations Researcher trained in and dedicated to education?
> * the applied Educational Researcher who can and will ferret out needed data and relationships even when these hold no promise of a good Journal article or convention report?
> * the Cliant who can understand and use what The Operations and Educational Researcher have to offer him? (p. 3)

We can repeat the question. Yet as C. P. Snow wrote, "The scientist has the future in his bones." Operations Research has the most to offer of any metaphor for the evaluative process. It provides a framework to tie together the best information and insight into a system to help us make important decisions. In the long run, if not immediately, it must prevail as the principal technology for evaluation.

References

Anderson, G. E., Jr. Operations research: A missing link. Educational Researcher, March 1970, 21, 1-3.

Banghart, F. Educational systems analysis. Toronto: Collier-Macmillan, 1969.

Bloom, B. S., Engelhart, M. D., Furst, G. J., Hill, W. H., & Krathwohl, D. R. Taxonomy of educational objectives: Handbook I, the cognitive domain. New York: David McKay, 1956.

Bobrow, D. G. Natural language input for a computer problem-solving system. In Minsky, M. L. (Ed.), Semantic information processing. Cambridge: M. I. T. Press, 1968. Pp. Pp. 135-215.

Carnoy, M. The economic costs and returns to educational television. In G. V. Glass (Ed.), Evaluation Studies Review Annual, vol. 1. Beverly Hills, Calif.: SAGE, 1976. Ch. 9, pp. 219-261.

Churchman, C. W. Prediction and optimal decisions: Philosophical issues of a science of values. Englewood Cliffs, N.J.: Prentice-Hall, 1961.

Churchman, C. W., Ackoff, R. L., & Arnoff, S. L. Introduction to operations research. New York: Wiley, 1957.

Cook, D. Program evaluation and review technique: Applications in education. (Cooperative Research Monograph No. 17.) Washington, D.C.: U. S. Department of Health, Education, and Welfare, 1966.

Cronbach, L. J., & Gleser, G. Psychological tests and personnel decisions, 2nd ed. Urbana: Univ. of Illinois Press, 1965.

Dalkey, N. Analyses from a group opinion study. Futures, December 1969, 1, 541-551.

Edwards, W., Guttentag, M., & Snapper, K. A decision-theoretic approach to evaluation research. In Struening, E. L., & Guttentag, M. (Eds.) Handbook of evaluation research, vol. 1. Beverly Hills, Calif.: SAGE, 1975. Ch. 8, pp. 139-182.

Evans, T. G. A program for the solution of a class of geometric-analogy intelligence-test questions. Ch. 5 in Minsky, M. (Ed.), Semantic information processing. Cambridge: M. I. T. Press, 1968. Pp. 271-353.

Feigl, H., & Brodbeck, M. (Eds.) Readings in the philosophy of science. New York: Appleton-Century-Crofts, 1953.

References

Anderson, G. E., Jr. Operations research: A missing link. Educational Researcher, March 1970, 21, 1-3.

Banghart, F. Educational systems analysis. Toronto: Collier-Macmillan, 1969.

Bloom, B. S., Engelhart, M. D., Furst, G. J., Hill, W. H., & Krathwohl, D. R. Taxonomy of educational objectives: Handbook I, the cognitive domain. New York: David McKay, 1956.

Bobrow, D. G. Natural language input for a computer problem-solving system. In Minsky, M. L. (Ed.), Semantic information processing. Cambridge: M. I. T. Press, 1968. Pp. Pp. 135-215.

Carnoy, M. The economic costs and returns to educational television. In G. V. Glass (Ed.), Evaluation Studies Review Annual, vol. 1. Beverly Hills, Calif.: SAGE, 1976. Ch. 9, pp. 219-261.

Churchman, C. W. Prediction and optimal decisions: Philosophical issues of a science of values. Englewood Cliffs, N.J.: Prentice-Hall, 1961.

Churchman, C. W., Ackoff, R. L., & Arnoff, S. L. Introduction to operations research. New York: Wiley, 1957.

Cook, D. Program evaluation and review technique: Applications in education. (Cooperative Research Monograph No. 17.) Washington, D.C.: U. S. Department of Health, Education, and Welfare, 1966.

Cronbach, L. J., & Gleser, G. Psychological tests and personnel decisions, 2nd ed. Urbana: Univ. of Illinois Press, 1965.

Dalkey, N. Analyses from a group opinion study. Futures, December 1969, 1, 541-551.

Edwards, W., Guttentag, M., & Snapper, K. A decision-theoretic approach to evaluation research. In Struening, E. L., & Guttentag, M. (Eds.) Handbook of evaluation research, vol. 1. Beverly Hills, Calif.: SAGE, 1975. Ch. 8, pp. 139-182.

Evans, T. G. A program for the solution of a class of geometric-analogy intelligence-test questions. Ch. 5 in Minsky, M. (Ed.), Semantic information processing. Cambridge: M. I. T. Press, 1968. Pp. 271-353.

Feigl, H., & Brodbeck, M. (Eds.) Readings in the philosophy of science. New York: Appleton-Century-Crofts, 1953.

Gaver, D. P., & Thompson, G. L. Programming and probability models in operations research. Monterey, Calif.: Brooks/Cole, 1973.

Grandon, G. B. The optimization of high school curriculum assignments. Socio-Economic Planning Sciences (Great Britain), in press.

Guba, E.G. A proposed introductory section for metaphor adaptation reports (tentative title). Alternative sources of evaluation methodology. ?

Hamburg, M. Statistical analysis for decision making. New York: Harcourt, Brace & World, 1970.

Hartley, H. J. Educational planning, programming, budgeting: A systems approach. Englewood Cliffs, N.J.: Prentice-Hall, 1968.

Hillier, F. S., & Lieberman, G. J. Introduction to operations research, 2nd ed. San Francisco: Holden-Day, 1974.

Johnstone, J. N. Mathematical models developed for use in educational planning: A review. Review of Educational Research, 1974, 44(2), 177-201.

Kaufman, R. Educational system planning. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Keeney, R. L., & Raiffa, H. Decisions with multiple objectives: Preferences and value tradeoffs. New York: Wiley, 1976.

Konopka, C. D. Applying dynamic programming to curriculum design of public school courses. Unpublished Ph.D. dissertation, University of Connecticut, Storrs, 1977.

Levin, H. M. Cost-effectiveness analysis in evaluation research. In Guttentag, M., & Struening, E. L. (Eds.) Handbook of evaluation research, vol. 2. Beverly Hills, Calif.: SAGE, 1975. Ch. 5, pp. 89-122.

McNamara, J. F. Mathematical programming models in educational planning. Review of Educational Research, 1971, 41(5), 419-446.

Novick, M. R., & Jackson, P. H. Statistical methods for educational and psychological research. New York: McGraw-Hill, 1974.

Page, E. B. Seeking a measure of general educational advancement: The bentee. Journal of Educational Measurement, 1972, 9(1), 33-43.

Page, E. B. Effects of higher education: Outcomes, values, or benefits. In Solmon, L. C., & Taubman, P, (Eds.) Does college matter? Some evidence on the impacts of higher education. New York: Academic, 1973. Pp. 159-172.

Page, E. B. 'Top-down' trees of educational values. Educational and Psychological Measurement, 1974, 34(3), 573-584.

Page, E. B. Evaluating 'Evaluation.' Review of Education, 1975, 1(1), 133-140.

Page, E. B. The optimization of educational values in Navy curriculum design. Proceedings of the American Statistical Association: Social Statistics, Part II: 1976, pp. 655-659.

Page, E. B. Should educational evaluation be more objective or more subjective? More objective! Educational Evaluation and Policy Analysis, 1978, in press.

Page, E. B., & Breen, T. F. III. Educational values for measurement technology: Some theory and data. In Coffman, W. E. (Ed.), Frontiers in educational measurement and information processing. Boston: Houghton-Mifflin, 1974. (a) Ch. 3, pp. 13-30.

Page, E. B., & Breen, T. F. III, Factor analysis of educational values across two methods of judgment. Proceedings of the 15th Interamerican Congress of Psychology (Bogotá, Colombia), 1974, pp. 106-107. (b)

Page, E. B., & Canfield, J. Design of Navy course structure through a dynamic programming algorithm. Report for the U. S. Navy Personnel R. & D. Center, San Diego, Calif., June, 1975.

Page, E. B., Jarjoura, D., & Konopka, C. Curriculum design through operations research. American Educational Research Journal, 1976, 13(1), 31-49.

Popham, W. J. (Ed.) Evaluation in education: Current applications. Berkeley: McCutchan, 1974.

Raiffa, H. Decision analysis: Introductory lectures on choice under uncertainty. Boston: Addison-Wesley, 1968.

Reitman, W. R. Cognition and thought: An information-processing approach. New York: Wiley, 1965.

Slagle, J. R. Artificial intelligence: The heuristic programming approach. New York: McGraw-Hill, 1971.

Stake, R. The case study method in social inquiry. Educational Researcher, 1978, 7(2), 5-8.

Tillett, P. I. Optimization of secondary teacher assignments using operations research. Socio-Economic Planning Sciences (London), 1975, 9, 101-104.

Trueman, R. E. An introduction to quantitative methods for decision making. New York: Holt, Rinehart & Winston, 1974.

VanDusseldorp, R. A., Richardson, D. W., & Foley, W. J. Educational decision-making through operations research. Boston: Allyn & Bacon, 1971.

Wagner, H. M. Principles of operations research: With applications to managerial decisions. Englewood Cliffs, N. J.: Prentice-Hall, 1969.

Winkler, R. L., & Hays, W. L. Statistics: Probability, inference, and decision, 2nd ed. New York: Holt, Rinehart & Winston, 1975.

Footnotes

1. Some work has been done in pushing the bentee at least a large step closer to application. Wayne Martin (of the National Assessment of Educational Progress) has explored values in Social Studies. Some of us have explored the redesign of a large and important course in electronics with the U. S. Navy (Page & Canfield, 1975; Page, 1976). And Dr. William Streich (of the Farmington, Conn., Schools) has studied course structure in high school departments of English and History.

2. ~~Note 2.~~ The well-known Bloom (et al, 1956) Taxonomy of Educational Objectives did not address itself to "values" of the various goals in the way we are describing, and thus is not directly useful for choice among programs. The various "levels" within the Taxonomy, on the other hand, can be weighted and incorporated within the bentee strategy.

3. We could just as well optimize the sum of preferences of those teachers, or the sum of ratings by students, or some combination of any of these.

4. This calculation treats the courses as unique, and ignores [illegible] schedule constraints. Then the possible permutations are 20!/(4!4!4!4!4!).

5. In the basic "transportation" model, each cell might rather be flagged by a cost c_{ij}, and the goal would be to minimize the total cost of the final solution. Mathematically, the adjustment is trivial.

6. The STUDENT program of Daniel Bobrow (1968) is perhaps still closer to the sort of metaphor we seek, since it converts natural language (English) problems into algebra, solves them, and translates the results back into English.

(Tables 1 and 2)

Table 1. Results of scheduling to maximise preference ratings in seven Connecticut High School Departments

Departments	1	2	3	4	5	6	7
Mean preference per course-section							
Existing schedule	7·5	7·6	8·1	7·8	8·1	6·1	7·6
Optimal schedule	8·1	8·1	8·4	8·0	8·1	7·0	7·8
Mean effectiveness per course-section							
Existing schedule	8·0	7·6	7·2	8·2	8·8	8·3	7·2
Optimal schedule	8·2	7·3	7·7	8·2	8·7	8·5	7·1

Table 2. Results of scheduling to maximize effectiveness ratings in seven Connecticut High School Departments

Departments	1	2	3	4	5	6	7
Mean preference per course-section							
Existing schedule	7·5	7·6	8·1	7·8	8·1	6·1	7·6
Optimal schedule	8·1	6·7	5·8	4·3	8·0	5·8	6·7
Mean effectiveness per course-section							
Existing schedule	8·0	7·6	8·1	8·2	8·8	8·3	7·2
Optimal schedule	8·2	7·8	8·6	8·4	8·8	8·8	7·8

Reprinted from Tillett (1975, p. 103, copyright Socio-Economic Planning Sciences, Great Britain), with the kind permission of author and publisher.

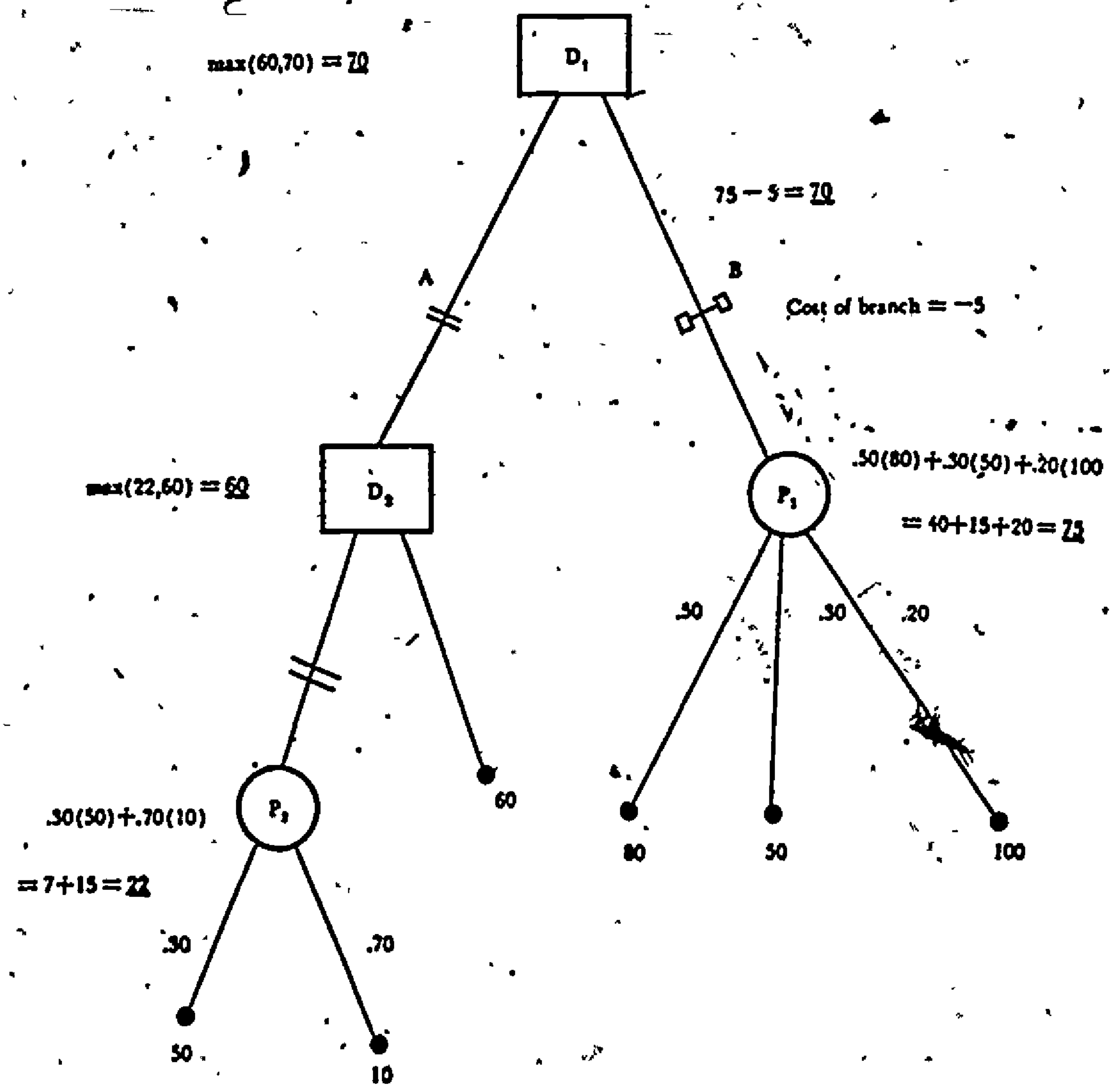

Figure 1. A decision tree.

A decision is reached by tracing out the branches as far as possible, assigning values to each terminal node, and probabilities to each branch from a P node. P nodes are then solved (working from the bottom up), by averaging out the branches. And D nodes are solved by folding back all but the most valuable branch as evaluated below each D. For vocations, the probability values are determined by knowledge of both the world and self, as are also the terminal values. Technical procedures can be applied to aid all such determinations.

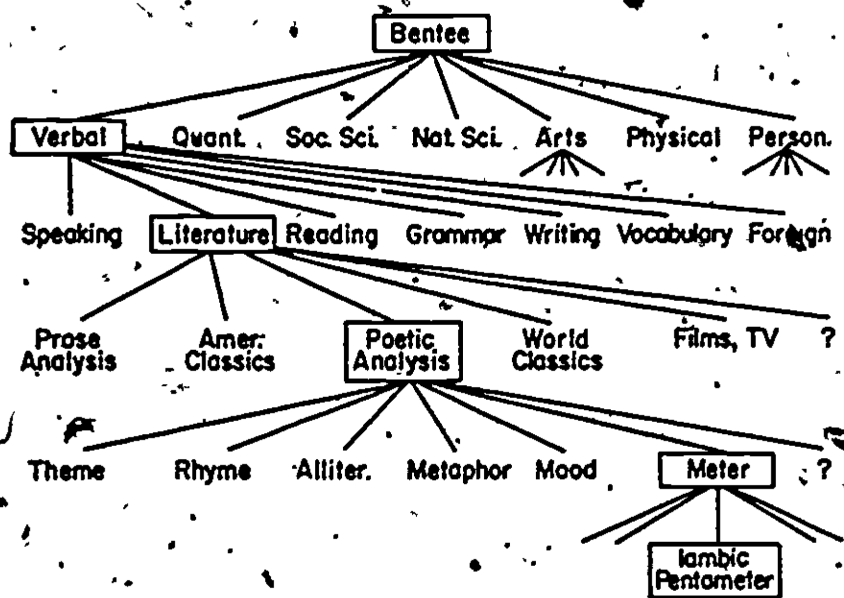

Figure 2. A bentee tree of educational value. If the bentee is defined as the "overall educational benefit" for a 12th-grade student, then the values of each part of interest in the tree may be defined by the token or other method. As noted by this tracing out of poetic interests, it is possible to move from the highest philosophical value to the most discrete item or instructional objective, within only a few subdivisions or "generations."

Source: E. B. Page, "'Top-down' Trees of Educational Values," Educational and Psychological Measurement, 1974, 34, 573-584.

Courses Offered

Teachers	1	2	3	4	5	6	7	Supply
1	r_{11}	r_{12}	.	.	.	.	r_{17}	4
2	r_{21}							4
3	r_{31}							4
4	r_{41}							4
5	r_{51}						r_{57}	4
Demand (sections)	3	2	1	2	8	3	1	(20)

Figure 3. Assignment of 5 teachers to 7 courses, 20 sections. In this "transportation" problem, each teacher has a row, and each course-type a column. The number of sections of each course required is the "demand"; and the number of sections each teacher is available is the "supply." For each cell, r_{ij} is the rating (or preference) of that teacher for that course-type. The aim is to assign teachers to courses in order to maximize the sum of r_{ij} in the assignment.

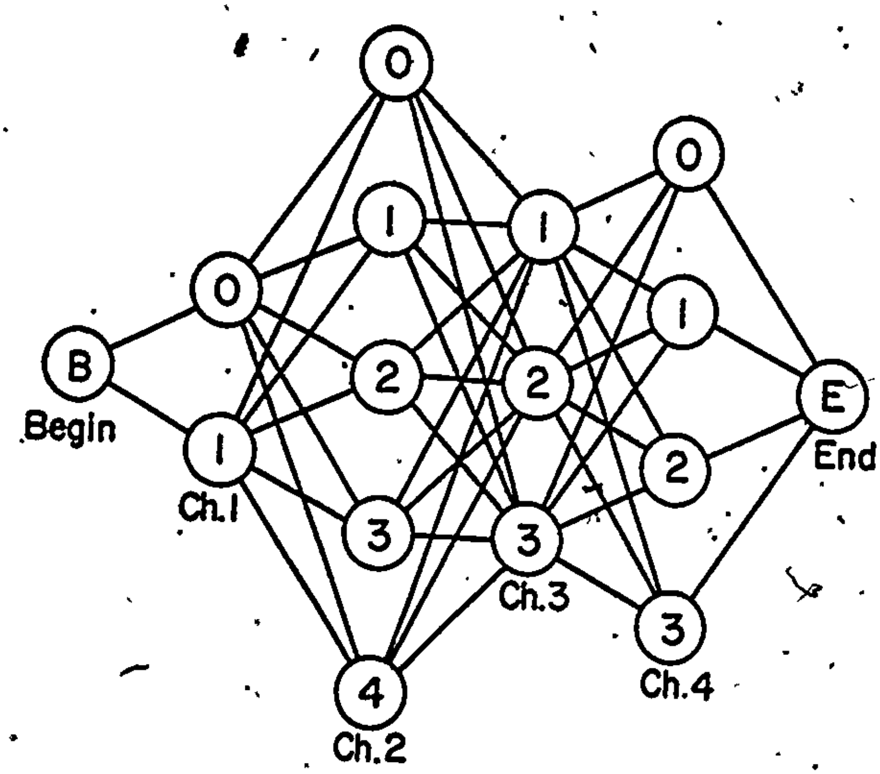

Figure 4. A textbook problem seen as a network. The student begins at the left and ends at the right, having studied each chapter at some level. Each node has a cost, and a benefit. The best path from B to E maximizes the benefit, while keeping costs with some permissible amount.

From Page, Jarjoura, and Konopka (1976, p. 41).

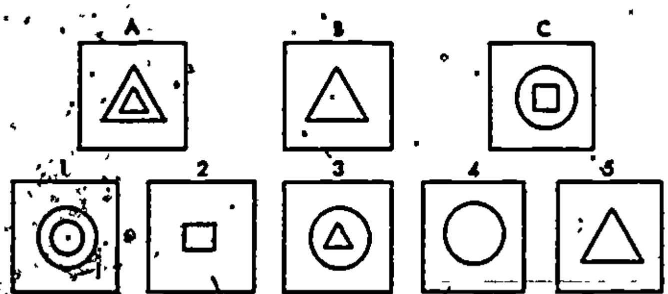

Figure 5. A "geometric analogy" intelligence-test item, of the sort solved by the computer program ANALOGY (Evans, 1968, p. 271). The aim is to complete the relation "A is to B as C is to what?" by choosing correctly among the five numbered alternatives. The accepted response is #4. Copyright 1968 by the M.I.T. Press, and reprinted with permission of the author and publisher.