THE OPTIMIZATION OF EDUCATIONAL VALUES IN NAVY CURRICULUM DESIGN

Ellis B. Page, University of Connecticut, Storrs

Summary

Curriculum theory has been recently shown to be analyzable with the quantitative tools of dynamic programming (DP), yet such applications are practically non-existent. What is required is (a) a list of candidate units or topics possible for inclusion; (b) for each candidate unit, at each considered level of instruction, a cost, usually expressed in time; (c) for each such unit, a value; and (d) some overall maximum acceptable cost. The most difficult problem is the fair and functional establishment of values, by instructors or other judges. Methods of apportionment and scaling were developed for a major technical Navy course, and found to have acceptable reliability. But a general phenomenon of "averaging" in judgment works against the more costly units when the DP algorithm is used, and compensations therefore need to be applied. Several statistical approaches are outlined for overcoming such a bias and making the technique usable in course design.

INTRODUCTION

We may say that a "curriculum" is a selection of materials for instruction, and an allocation to each material of some limited resource, usually time. A curriculum "problem" may be said to exist when when the candidate materials for such instruction exceed the available resource. Then we are forced to choose among the candidate materials, and to arrange the resource allocation such that we do not exceed the time available. Furthermore, if we wish to make the best possible curriculum, we will try to optimize the value of the materials included.

Recent theoretical work (Page, Jarjoura & Konopka, 1976) has treated this curriculum problem as a problem in operations research, and especially as a problem in dynamic programming (DP). We may picture it simply as if it were a college textbook, and we wished to move through the text assigning each chapter to some level of instruction (including the zero level, which would be omission of that chapter). Each level for each chapter has some estimated cost (in terms of hours) and some estimated value (in units to be discussed). And for the book as a whole, we begin with some maximum total cost (in hours assignable). The solution to the problem, then, consists in assigning each textbook chapter to some level of study, such that the total hours consumed do not exceed the maximum allowable, and such that the estimated value is as high as possible.

The theoretical analysis treated this textbook problem as if it were brand-new, and as if the "levels" for each chapter might be chosen (such as "read chapter once lightly") quite independently of the assignment of costs and values. In the applied situation herein studied, there were discovered problems which required adjustments in the approach, which considerably affect the results of the DP algorithm. Most particularly troublesome were certain behaviors in the assignments of value to the units, and in the effects of these behaviors on the computer solutions.

AN APPLIED NAVY ANALYSIS

In practical terms, the course Sonar Electronics Intermediate may be one of the most important courses in the Navy. Taught in San Diego, it had a well-experienced and able instructional staff, familiar with the course, with the fleet requirements, and with the trainees. Some of these instructors served as judges for the values of the larger units within the course, and of the topics within each unit.

The specification of costs

Navy courses are commonly taught, like certain civilian instruction, from a thoroughly worked-out curriculum guide. Such a guide will specify the content of each portion, the problems to be presented and solved, and the time to be allotted to each. In general, the materials of such a guide are explicitly allotted, then, a certain "cost" for our curriculum analysis.

After some experimentation, the most useful way of describing the costs for proposed changes, then, was in terms of the current costs. And certain modifications of that current cost were essayed, as follows: No hours assigned, 40% of current costs, 80%, 100%, 120%, and 240%. The guiding principle was that there would be five possible levels for a unit other than its present level. Two of these would be quite conservative in their change (20% to either side of the present level), and two of these would be more radical (halving or doubling these more conservative changes), and one would represent the possibility of excluding that unit from the Navy course.

The stimulus that resulted from these calculations, then, may be seen in Figure 1.

PROPORTION	0	40	80	100	120	240
HOUR COST	0	35*	70*	87	100*	200*
ESTIMATED VALUE	0			100		

Figure 1. The stimulus presented for scaling value in a Navy course for one unit.

In Figure 1, it is seen that the proportions of cost are scaled from 100%, which is the current program, and which uses 87 hours of instructional time. The hour costs marked by asterisk (*) are scaled to match the percent of 87 hours, as indicated in the top row. (Thus 35* hours is rounded from .40 x 87 = 34.80.) These cells, therefore, become hypothetical costs, amounts of time which might be considered by the judges of the course.

The specification of values

As noted above, the current costs for a Navy course are as prescribed by the established outlines. And the hypothetical costs are arbitrary, scaled for a spread of radical and conservative deviations from the current costs. On the other hand, the specification of values is performed by the expert judges themselves.

There are two acts of judgment performed by the experts in a curriculum. These may be done by the same judges, or by different judges; and they may be done at one sitting, or at two. The acts are:

a) Assigning values to units of the current curriculum. Suppose there are five units under consideration. Then one is given 100 tokens (poker chips, all of same value), and asked to "spend" these tokens in accordance with his beliefs about the values of the units as currently taught. He will spend all 100, so that his registered evaluation will be expressed as v_{ij}, where he is the ith judge, appraising the jth unit, and where

$$\sum_{j=1}^{m} v_{ij} = 100.$$

b) Assigning proportional values to hypothetical levels of cost. In a separate operation, a judge is asked to fill in the four blank cells in Figure 1. He is asked to assign a percent of value to a reduced, or an increased, effort dedicated to the unit in question. The first task, above, compared each unit with each other unit in assigning value. The present task scales the hypothetical value of certain fixed amounts of hypothetical effort. The judge's scaling, then, will be expressed as R_{ijk}, where he is the ith judge, appraising the kth level of the jth unit. These values must be non-decreasing, as k rises from 0 to 6.

RELIABILITIES

Reliabilities of the token strategy

For this pilot research, there were five Navy instructors who served as judges of the values. All were widely experienced in Sonar Electronics, both as instructors of the course and as technicians in the Fleet. Independently of each other, they all assigned token weights to the five units of the course, and to the topics of three of these units. A measure of concordance could then be taken by the correlation ratio,

$$\eta = \sqrt{\frac{SS_{bet.units}}{SS_{total}}}.$$

For the course as a whole, such analysis yielded a correlation ratio of .66. And for Units 1, 4, and 5, the corresponding statistics were .74, .76, and .76. Despite the small number of judges, these agreements were all statistically significant.

Therefore, it may be inferred that, among these instructors, there was some concordance of viewpoint, which renders their mean judgments of some importance.

Reliabilities of the level scaling technique

As noted above, each judge would scale the hypothetical value of each unit for each level (these levels being 40%, 80%, 120%, or 240% of the current expenditure). In order to appraise the concordance of judgment, and to avoid the problems of monotonicity, each level would be considered independently. For the overall level scaling of the course, across five units and five judges, at the 40% level of effort, the data are shown in Table 1.

THE 40% LEVEL FOR ALL UNITS

Judges	Units 1	2	3	4	5
A	80	100	40	80	50
B	50	40	10	10	60
C	40	30	10	5	20
D	90	35	25	10	20
E	25	70	10	30	25

η = .56

From sixteen such tables as Table 1, it was possible to calculate the correlation ratios for all the investigated levels, for the course as a whole, and for three units of that course. The results are as seen in Table 2, and the levels of significance are as indicated.

TABLE 2

CORRELATION RATIOS FOR FOUR LEVELS OF EFFORT FOR THE COURSE AND THREE UNITS

	Levels of Effort 40%	80%	120%	240%
COURSE (5 units)	.56*	.65*	.27	.35
UNIT 1 (20 topics)	.45	.72**	.49	.63**
UNIT 4 (19 topics)	.61**	.54*	.44	.45
UNIT 5 (9 topics)	.43	.62*	.54	.43

* Sig. at .05 level
** Sig. at .01 level

From Table 2, it is apparent that the agreement concerning levels of hypothetical value is enough to work with, when taken as an average. Furthermore, it was observed that instructors had no apparent difficulty, no great hesitation, in assigning these values. Familiarity with the course had given them fairly strong opinions about the units of the course, and about the topics of the units.

OPTIMIZING THE ASSIGNMENT OF HOURS

From such measures of value of the units, and of the hypothetical levels of effort, it is possible to compute an objective function for the dynamic program algorithm. First, each $\underline{v}_{ij}$ and each $\underline{R}_{ijk}$ is averaged across judges. Then the product of these averages becomes the (usually non-linear) objective function for the algorithm.

For example, in Figure 1, scaling percentages for the four empty cells were, from left to right, 57, 93, 107, and 110. The mean value of the unit, taken as a whole, was just .12 of the course's total value. Therefore, a new line to Figure 1, of "adjusted value," may be calculated, and these values will be 0, 6.84, 11.16, 12.00, 12,84, and 13.20. The value is thus seen to be curvilinear, with much of the current value (12.00) achieved with 20% less cost (11.16), and little to be gained by increasing the hourly expenditure for the unit.

The mathematical model is that of a "distribution of effort" problem, in standard works (cf. Wagner, 1970, pp. 254-257). The assignment of each unit or topic to a specific level of effort is made, such that the total expenditure of hours does not exceed some fixed quantity, and such that the sum of values of those assignments is maximized. And Moonan (1976) recently provided a useful algorithm for the solution of such problems.

Comparing the current and optimal assignments

From such dynamic programming, what is of interest is a comparison of the solutions with the values of the current strategy. What is also interesting is the hypothetical values which could be achieved with even <u>less</u> expenditure of hours than is currently given. The result of such comparison is seen in Table 3.

Obviously, the algorithm was highly successful in finding an assignment which would increase the estimated value of the course. For the apportionment of time, at the current level, to the course as a whole, the increases over the current value amount to 7%, and for the three units examined, the gains are 29%, 15%, and 9%. In fact, even with the allotted time reduced, it is possible to increase the estimated value, as is evidenced in the two other columns of data in Table 3. With a 10% reduction in time-cost, it is possible to maintain or surpass the current value. And even with a one-quarter reduction in time-cost, the estimated values are sustained not too much below the current values.

TABLE 3

VALUE ACHIEVED FOR FOUR TEST CASES AT THREE ALLOCATION LEVELS THROUGH DYNAMIC PROGRAMMING

	Allocation level compared w. present		
Case	100%	90%	75%
ALL UNITS	107	99	92
UNIT 1	129	124	118
UNIT 4	115	108	97
UNIT 5	109	103	89

THE NATURE OF THE SOLUTIONS

When the optimal assignment tables are produced (Page and Canfield, 1975, App. F), it is possible to study the relation of the optimal assignment in comparison with the current assignment. We may consider Figure 1 to contain six levels of effort, 0, 1, 2, 3, 4, and 5, corresponding with the columns, with level 3 representing the current solution. A "conservative" solution would be one in which level 3 was frequently the optimal level. A "radical" solution, on the other hand, would be one in which level 3 seldom appeared in the new assignment table.

By such a criterion, the solutions for this Navy course were decidedly "radical." Level 3 appeared only once of 15 possible decisions for the three tables of the course as a whole. It appeared only six times of a possible 60 for Unit 1; only eight times of a possible 57 for Unit 4; and only 4 times of a possible 27 for Unit 5. Given that this is a major Navy course, developed over years of intensive effort by many able people, such radical changes must be cause for very careful study -- not wrong per se but requir- great caution in interpretation. What is the nature of the solutions, and why are they so radical?

Part of the answer is that, in the solutions, the zero level appears a total of 24 times. That is, there are 24 occasions in the solution tables where a unit or topic is reduced to nothing. Especially, it appeared that this reduction to nothing sometimes occurred in the largest, most costly units or topics. To analyze this possibility, Table 4 was generated.

Table 4 reveals an apparent psychological phenomenon of the valuing strategy: a rather weak relation between the token value of a unit, and its current time cost. The token values are generally related <u>positively</u> to the current costs,

TABLE 4

CORRELATIONS BETWEEN CURRENT TIME-COSTS AND TWO MEASURES OF VALUE

CASE		Correlation of Time-Cost With	
		Token Value	Level Assign.
ALL UNITS	(5 units)	.41	-.63
UNIT 1	(20 topics)	-.45	-.65
UNIT 4	(19 topics)	.17	-.10
UNIT 5	(9 topics)	.81	-.03

but weakly. There is, in tokening, an apparent averaging which goes on, and which means that the variance in token proportion is much lower than the variance in cost proportion.

The effect of such averaging is seen in the last column of Table 4, in a strong negative bias against the larger units or topics. Subsequent debriefing of the judges revealed that it was not their intent to eliminate these large units; rather, it reflected that tendency to average their values, to lower the between-unit variance. There was, in fact, something of the feeling, though not verbalized during the rating, that they were evaluating the worth of a unit in terms of the "value per hour." But the effect of such valuing is to lower the estimated value for the unit (or topic) as a whole.

Yet the DP algorithm spends the total available hours in the most cost-effective way possible, which means that more effort is expended on those units where the value per hour is higher. And the algorithm computes the "value per hour" by dividing the total value by the number of hours currently allotted. Let us suppose two units, one of 50 hours and one of 10. If we appraise the first to be worth 20 tokens, and the second to be worth 10, then there is a correlation between cost and value. Yet the value per hour of the first is 20/50 = .4, while the value per hour of the second is 10/10 = 1.0. Thus in any DP algorithm, the second, shorter unit will commonly be emphasized. This phenomenon, then, is a feature of judge behavior, which must somehow be adjusted to render the the DP algorithm feasible in applied settings.

DISCUSSION

In a good many applications of OR, any difficulty is caused by an insecure fit of the model to human behavior. Most especially may this be true in educational applications, which have long suffered from inadequate measures of value (Page, 1972, 1974; Page and Breen, 1973). In the present application, there is here revealed a serious ambiguity, in the behavior of the judges, about what constitutes the "value" of a unit, whether their judgment reflects the worth of the unit as a whole (as it properly should, for the DP algorithm to function), or whether their judgment also reflects a factor of the worth of the unit per hour. The problem should call forth the ingenuity of workers in psychological scaling, and in operations research.

The following suggestions, therefore, are for future research in the field. They are broken into two sorts of strategy, one aimed at behavioral adjustments to this averaging tendency, and the other at statistical adjustments:

Behavioral adjustments

1) Begin the token strategy by placing before the judge a stack already proportional to the current time costs. This might serve to remind him of the overall nature of the evaluation, and to counteract the averaging tendency.

2) Or the judge might be explicitly told that it is the "value per hour" that he is estimating. Then, to use the DP algorithm, the weight of the unit as a whole would be the judge's estimate multiplied by the number of hours.

Statistical adjustments

3) The tokening might be skipped, and judges used only for the relative scaling of value at the various levels. To some extent, important information about the unit value is contained in such scaling. The statistical adjustment, for purposes of the DP algorithm, would be simply to insert a pseudo-token value equivalent to the proportion of current time costs.

4) The token judgments, once collected, might be adjusted so that the variance of their proportions would be made equivalent to the variance of the proportions of current time costs. This technique would leave the relative values of the judges undisturbed, but would fan them out in a way to eliminate the effects of the "averaging tendency."

Research along these lines should be successful in rendering the DP algorithm useful in the modification of curriculum of many different kinds.

ACKNOWLEDGEMENTS

This research and development was performed under NPRDC Code 306, Test and Application of Training Technologies and Systems (cf. Page and Canfield, 1975). The specific application was stimulated by the need to evaluate the course, Sonar Electronics Intermediate, taught by the Fleet Anti-Submarine Warfare School, San Diego. Earlier work with the same course was done by Stern and Aiken (1969).

During this work, valuable advice and assistance were given by E. C. Aiken, E. F. Alf, A. G. Archibald, J. G. Balaban, R. Boller, J. D. Carter, Jan D. Dodson, T. M. Duffy, J. D. Fletcher, E. I.

Jones, K. O. Leland, Carolyn L. McLandrich, W. J. Moonan, F. A. Muckler, J. J. Regan, B. Rimland, D. L. VanKekerix, H. L. Williams, and J. H. Wolfe, all of the Navy Personnel R & D Center, San Diego. Particular help was also given by P. V. Asa-Dorian and the instructors of the Curriculum Development and Evaluation Center, Fleet ASW School. Computer assistance was received from G. Grandon and C. Konopka, of the University of Connecticut, Storrs. Thanks are also due to P. Hawkes and A. Lubin. The most constant and direct collaboration was given by Dr. James Canfield, without whom this report would not have been possible.

REFERENCES

Moonan, W. J. DYNA: Dynamic programming analysis and the distribution of effort. Paper presented at the Military Testing Association Conference, October 1976, Pensacola, Florida.

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

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

Page, E. B., and Breen, T. F. III. Educational values for measurement technology: Some theory and data. In W. E. Coffman (Ed.), Frontiers of educational measurement and information systems. Boston: Houghton-Mifflin, 1973.

Page, E. B., and Canfield, J. Design of Navy course structure through a dynamic programming algorithm. Consultant report at the Navy Personnel Research and Development Center, San Diego, June 1975. Pp. 69.

Stern, H. W., and Aiken, E. G. Training feedback study of the Sonar Technician Intermediate Electronics Course. Navy Personnel R & D Center, SRR 70-2, September 1969.

Wagner, H. M. Principles of management science: With applications to executive decisions. Englewood Cliffs, N.J.: Prentice-Hall, 1970.