DOCUMENT RESUME

ED 054 230 24 TM 000 826

AUTHOR Borg, Walter R.
TITLE Some Thoughts About Product Evaluation at the Far West Laboratory.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-6-2931
PUB DATE 8 Jan 71
CONTRACT OEC-4-7-062931-3064
NOTE 9p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Educational Equipment, Educational Programs, *Equipment Evaluation, *Evaluation, *Evaluation Techniques, *Marketing, Merchandise Information, Merchandising, Objectives
IDENTIFIERS *Evaluation Continuum

ABSTRACT

Questions are discussed concerning the problem of evaluating an educational product for its potential marketing success. A strategy is proposed for identifying various levels of evaluation for a product, and for selecting the evaluation level that will be accepted as yielding sufficient evidence for the desirability of marketing the product. (MS)

1/8/71

EA

SOME THOUGHTS ABOUT PRODUCT EVALUATION AT THE FAR WEST LABORATORY

Relevant Evaluation Questions

The problem of evaluating an educational product and making a decision as to whether or not the product should be marketed, is a highly complex one. The program staff of the Laboratory has identified four questions that appear to be of major importance in arriving at the aforementioned decision. Although at this point we do not suggest that these are the only questions that are relevant or even the most important questions that could be asked, we do maintain that these four questions are important and can form the basis for further thinking about the overall question of educational product evaluation.

The first question can be stated briefly as: "Will the product sell?" However, since there is some evidence in the world of commerce that virtually any product will sell if an appropriate and extensive sales effort is made, perhaps a more useful way to state the question is "What will have to be done in order to sell this product?"

A second question that must be answered: "Is the product creditable?" It is possible to build an effective product which, because of the biases or values of the potential users, will not be accepted by them as creditable. For example, it is doubtful that a nursery school program based on the principles of contingency management and behavior modification would be adopted by most nursery school teachers even if this program had proven to be highly effective. Since the majority of nursery school teachers have a heavy clinical orientation, a program based on the principles of behavior modification would not be creditable to them regardless of the merits that

it might have. There are other examples of areas in education where creditability plays a large role in the consideration of educational products for possible use. Foreign language programs built primarily around memorization and application of the rules or grammar would not be creditable to the vast majority of foreign language teachers in today's schools. Similarly, a fine arts program built around the premise that artistic talent can best be developed by having the learner copy the paintings of known masters would lack creditability to the majority of art teachers in today's schools.

The third question that we have considered is: "Will teachers use the product?" A great many educational products have appeared over recent years that are both creditable to the user and saleable but which are not used once they have been purchased. Like the previous two questions, the question of usability is a very complex one. Products are purchased and then not used for a great many reasons. For example, some products are not used simply because their use makes demands on the teacher which the teacher is not prepared to meet. New audio-visual equipment sometimes require skill and training which the teacher does not have. New curricular materials often require more preparation time than the teacher has available. New instructional packages may require the teacher to break long established habits.

The fourth question and the one with which this paper is primarily concerned is: "Does the product work?" This is, of course, the classic evaluation question, and is concerned with the degree to which the specific product objectives are achieved when the product is used in the manner pre scribed by the developer. Many developers would regard this question as

the most important of the four that we have enumerated. However, debates about the degree of importance of these four questions are somewhat pointless in view of the fact that if the product fails in any of these four areas, it has virtually no chance of making an impact upon education.

A great deal of heated argument has evolved around the means that should be used in obtaining an answer to this question. Many such arguments appear to be based on the idea that there is only one right way to evaluate an educational product. This paper, on the other hand, is based on the premise that for any given educational product there are a great many evaluation procedures that are appropriate and that provide at least some relevant evidence. In this paper we will propose a strategy for identifying various levels of evaluation for a given product and selecting the level of evaluation that will be accepted as yielding sufficient evidence for the product in question.

An Evaluation Continuum

The first step in this process is to develop a continuum of evaluation procedures for the product to be developed. This continuum should describe procedures ranging from the least rigorous evaluation to the most rigorous evaluation.[1] However, it is doubtful that a purely quantitative continuum can be developed, since at certain points along that continuum we will also find that qualitative differences exist in the kinds of evidence that are yielded.

In order to provide an example, we have taken the Minicourse Component of the Teacher Education Program and have attempted to outline a continuum

1. The most rigorous procedure would be that which produced the most reliable evidence on the degree to which the product achieved its objectives.

by specifying a number of levels of evaluation. In looking over this continuum, the reader will note that the various steps generally move from less rigorous to more rigorous evaluation of the course objectives. The reader will also note that at certain points qualitative differences exist between the type of information gathered at one level and the type of information gathered at the next level. I am sure that most readers will also observe that only a few of the possible points along this continuum have been identified and described. Actually, as is the case with any continuous distribution, it would be possible by making adjacent steps small enough to approach an infinite number of different levels of evaluation. The authors suggest that the 12 steps given are sufficient for the purposes of illustrating what an evaluation continuum is like and concedes that in all respects except illustration, the following continuum is inadequate.

Broad Objective

To improve specific aspects of pupil performance by developing teacher skills related to the pupil performance in question.

Levels of Evaluation

1. The developer will infer the effectiveness of the product by a general, subjective appraisal of the product itself.
2. An independent expert will infer the effectiveness of the product by a general, subjective appraisal of the product itself.
3. A panel of developer experts and potential users will independently infer the effectiveness of the product by a general, subjective appraisal of the product itself, and their judgments will be pooled.

4. The developer will infer the effectiveness of the product by a specific objective appraisal of the product itself.
5. An independent expert will infer the effectiveness of the product by a specific objective appraisal of the product itself.
6. A panel of developers, experts, and potential users will independently infer the effectiveness of the product by a specific objective appraisal of the product itself, and their judgements will be pooled.
7. A questionnaire will be administered to teachers taking the Minicourse asking their perception of their own and their pupils' performance.
8. A semi-structured interview will be conducted with a sample of teachers who have taken the Minicourse. Interviews will focus sharply on specific examples of relevant pupil and teacher performance.
9. Tests which indirectly measure variables that are related to the relevant pupil performance will be administered to pupils and analyzed.
10. Relevant teacher performance in the classroom will be observed at previously specified times by observers whose presence is known by teachers and pupils. Pupil performance will be inferred from analyses of teacher performance.
11. Relevant pupil performance in the classroom will be observed at previously specified times by observers whose presence is known by teachers and pupils.
12. The classrooms of large random samples of teachers taking (T) and not taking (C) the Minicourse will be recorded on videotape prior to the treatment and after the treatment over a period of one academic year following completion of the treatment. All videotape will be analyzed to determine the quality and quantity of performance of each pupil in

the T and C classrooms. Furthermore, random time samples of each pupils' out-of-school behavior will be recorded on videotape with pupils not knowing his behavior is being recorded. These videotapes will also be analyzed for specific pupil performance to determine degree of transfer to out-of-school situations. Pupil characteristics relevant to the performance in question will be measured and interactions among characteristics and performance will be determined.

Building an Evaluation Continuum

The first step in establishing such a continuum is to spell out the objectives of the educational product as clearly as possible. Since in this example, we are dealing with a group of educational products (the Minicourses) we have stated a broad objective as a first step. In dealing with a specific product, it would be preferably to spell out specific behavioral objectives for the product at this point.

The next step in developing a continuum of evaluation levels is to describe the most rigorous evaluation procedure that the investigator can devise for evaluating the product's achievement of its objectives. For most educational products, this level of evaluation could not be achieved simply because the cost of reaching this level would be so high that it could not be justified. For many educational products, it is unlikely that the most rigorous level of evaluation could be carried out even if the investigator had unlimited resources at his disposal. Thus, the development of a continuum of evaluation suggests that some sort of a balance must be struck between the rigor of the evaluation data obtained and the cost of obtaining it.

The next step in the development of the evaluation continuum requires the investigator to identify the least rigorous evaluation effort that can provide any sort of usable information about the product. At this point the investigator has identified the two extremes of the continuum. He must then move gradually along the continuum and describe enough additional points so that the overall structure of the continuum is reasonably clear. If the continuum were purely quantitative, ranging from the least rigorous to the most rigorous evaluation procedures within a given evaluation process, the next step would be to move up or down the continuum until the investigator had reached a point that he regarded to be the ideal compromise between rigorous evaluation and reasonable cost. Unfortunately, however, the investigator is likely to discover after he has filled out a number of points on the continuum that at certain points the evaluation process shifts, thus introducing qualitative differences between adjacent steps. In our example, substantial qualitative differences appear to exist at a number of points, such as between steps 3 and 4 and between steps 6 and 7. Actually, one could find slight qualitative differences in the nature of the data collected between any two steps on the continuum in addition to the quantitative differences related to the level of rigor.

Since both quantitative and qualitative differences are likely to exist, the final step in using the evaluation continuum requires identifying one or more evaluation descriptions along the continuum that together constitute a "satisfactory" evaluation of the product and collecting evidence called for in these descriptions. I would suggest that with regard to Minicourses, the satisfaction of the evaluation requirements stated at

points 6, 8 and 11 would provide a reasonably cost effective evaluation process for a Minicourse. In most educational product evaluations, what is "satisfactory" must usually be considered in terms of cost versus effectiveness.

Other steps could be taken that would add to the sophistication and the usefulness of the evaluation continuum. For example, an experienced educational developer could make a reasonably accurate estimate of the cost of carrying out the evaluation described in each of the steps specified in the continuum. Such cost estimates would surely be useful in making a more refined decision as to what points in the evaluation continuum would be applied in evaluating the given product.

The reader will note that many of the steps in the evaluation continuum refer not only to question 4 ("Does the product work?"), but also to the other evaluation questions stated at the beginning of this paper. Item 6, for example, would probably provide information on course creditability and usability as well as providing limited evidence on effectiveness.

Summary

In summary we would like to suggest several points related to the evaluation continuum and the overall process of educational product evaluation:

1. There is no single evaluation process that is universally good or bad. For any given product, a continuum of possible evaluation strategies exist and the strategy to be used should be drawn from that continuum primarily on a cost benefits basis.
2. If we constructed a dozen evaluation continuums for a dozen different educational products, it is probable that certain common features

would be present in many of these, since educational evaluation is built upon a limited number of processes such as observation, interviews, tests, questionnaires, simulations, etc. However, in spite of the fact that many evaluation continuums are likely to be similar, it is probably necessary to tailor-make each continuum to fit the product in question.

3. If any general guidelines can be laid down about the use of the evaluation continuum, one possible rule is that one should go as far towards the rigorous end of the continuum as the practical considerations of cost, time, etc., will permit.