

# Evaluating Health Education

By ANDIE L. KNUTSON, Ph. D.

Health people frequently ask: How can we tell if we are accomplishing anything? What methods should be used to evaluate our work? What are the criteria for effectiveness in health education?

This paper is a general approach to, or a philosophy of, evaluation. It outlines some basic concepts to consider when planning and conducting health education evaluation. The concepts, although presented independently, are closely interrelated and supplement each other.

Three things of utmost importance in evaluating health education are (a) the correct selection or development of methods, (b) the proper use of the methods selected, and (c) the sound interpretation of the data obtained. Given below are a few basic guidelines which should be considered in selecting, applying, and interpreting the techniques of measurement used in evaluating health education.

## Exploratory Fact-Finding

Planning evaluation studies of health education requires adequate exploratory fact-finding. Such explorations should be considered as part of planning a health education program. They should assure that the program is directed toward satisfying expressed needs or interests of the people for whom it is designed and that the methods used tie in with existing behavior patterns. Careful exploration is particularly necessary when the program aims to change behavior. If such problems are not considered

as part of planning a health education program, it may be based on false or inaccurate assumptions regarding behavior and have little possibility of success. Later efforts at evaluation would have limited or no value.

For example, before attempting to measure the effectiveness of a program to inform people about the nutritional value of proteins, carbohydrates, vitamins, minerals, and fats, one must know what meanings these terms convey. How can the concepts described by these terms be most easily incorporated into the usual pattern of living? Do people shop for proteins and carbohydrates or do they buy meat, vegetables, and baked goods? How can they improve their food habits with a minimum of change in their current patterns of buying food and planning meals?

A program designed to increase the use of milk may miss its goal if the people for whom it is planned do not consider milk an essential food. Or an information program on the early symptoms of cancer may not lead to action unless these symptoms are described in terms people understand and can recognize.

## Specific Goals

The broad purposes of an educational program must be broken down into specific, concrete, definable goals before evaluation can be undertaken. When the objectives are fully and concretely defined, one can determine in advance what kind of data will yield the best evidence of their achievement. A study design can then be developed to collect the necessary data.

On the other hand, if an over-all evaluation is attempted without breaking the broad purposes of the program into specific goals, the results are likely to be inadequate, yielding personal impressions rather than objective meas-

---

*Dr. Knutson, a social psychologist, formerly was associate director of the Office of Public Opinion Research of Princeton University. He is now chief of the experimental and evaluation services branch of the Division of Public Health Education, Public Health Service.*

---

urements. (See references 1, 2, and 3 for experimental evidence of the errors which result when general ratings are used in attempting over-all evaluation.) Such impressions will vary according to the experiences and interests of the persons doing the evaluating. As shown by studies in social psychology (4), each evaluator will tend to see in the program the things which are meaningful in terms of his own experiences. He will judge it against his own unique and undefined scale of values. One evaluator may view a part of the program in which he is interested and judge it to be good; another may rate this same part as poor; while a third may not even consider it to be significant and will base his judgment on some other aspect of the program. In the end, the evaluation is essentially a variety of unrelated personal opinions.

### General vs. Specific Evaluation

The interrelationship between the findings on the achievement of specific goals must also be considered. The whole is greater than and sometimes different from its parts. While a general evaluation may be referred to as such, in practice it means that an empirical judgment has been made, based on various specific evaluations.

In making a general health examination, the doctor measures specific things, such as height, weight, blood pressure, pulse, and vision. His general evaluation is limited by the degree to which he can break it down into measurable specifics and by his understanding of the interrelationship of these specifics. At the same time, the examination is limited by his own experience, through which he has learned to see certain things to the exclusion of others, and by his own concept of general health as it relates to a patient. Accuracy in interpretation is limited to a person's ability to perceive and to understand. Thus, a general practitioner might have difficulty in diagnosing an unusual illness because he does not know the specific disease, because he has had no experience in detecting it, or because the disease is completely unknown.

As the doctor acquires broader experience and his tools become more numerous and refined, he can evaluate specific items in the general examination with a higher degree of precision and

reliability. Accordingly, the results of his general examination gradually approach a total evaluation of his patient's health.

The measure of success of an activity is how well it achieves its intended purpose. This principle applies no matter what the purpose, whether it is to raise curiosity, to inform, or to produce action.

Health materials or programs may also provide information or influence behavior in some way other than the one intended. A person reading a health pamphlet, seeing an exhibit, or having some other educational experience may find in it something that directly ties in with his experiences and interests and thereby helps him to achieve his own goals. His knowledge, understanding, and behavior may all be influenced. He would probably consider the materials or program very successful.

From the point of view of the educator, however, such materials or programs should not be considered successful unless the intended objectives are also achieved. It matters not if the action caused is even more desirable than the action hoped for. To interpret as indications of success evidence of behavior changes other than those intended is to set up *post hoc* objectives.

Success, from the viewpoint of those who plan materials or programs, need not conflict with success from the standpoint of the persons for whom they are planned. Studies of human behavior suggest that educators will be most effective in achieving their purposes when such purposes are defined in terms of the wants and goals of members of the intended audience and are tied in with their way of life.

### Criterion of Effectiveness

Concrete evidence that an objective has been achieved is the only realistic criterion for measuring effectiveness. Such evidence should be distinguished from criteria which measure some of the conditions necessary to achieve the objective.

The number of items of material distributed, the number of persons attending a movie or exhibit, the amount of attention attracted by exhibits or other materials, the level of readability, and similar items represent conditions

necessary for effectiveness, although they are sometimes cited as measures of effectiveness. Certainly a movie, an exhibit, or a lecture must be seen, read, or heard to be effective, but that may not be sufficient. The observer may still be left uninformed, misinformed, or with negative attitudes, and may take no action.

The difference between conditions necessary for effectiveness and evidence of effectiveness itself is generally accepted in psychological research but is not yet widely recognized in health education. Attention has been called to it by Derryberry (5) as follows:

. . . in the past, counts of the numbers of spectators and average observation time have been cited as positive evidence of an exhibit's effectiveness. However, such records do not constitute positive evidence of the effectiveness of exhibits, because the records fail to indicate what, if anything, the spectator learns. Certainly there is no assurance that he is learning the message intended. So far as can be detected from observations, the visitor may be drawing wrong conclusions from the material in the display. Consequently, it is apparent that such records are valuable only in a negative sense—negative in that an exhibit cannot be effective as an educational medium if it fails to attract attention of visitors or fails to sustain their interest long enough for them to complete inspection of it.

The failure of a program to satisfy the conditions necessary for effectiveness leads to a screening out of some members of the potential audience. For example, all members of the potential or intended audience who did not receive a pamphlet, attend a movie, or see an exhibit (were not exposed to the educational effort) are eliminated because one of the conditions for effectiveness (exposure) was not met. Similarly, lack of interest and understanding will screen out other members of the intended audience. The fact that some individuals were not screened out by these limiting conditions and did participate in the program is not adequate evidence, however, that the program was successful for them (6).

### **Selection of Methods**

The method used to evaluate a program should be especially selected or constructed in terms of the particular goals to be reached. Accordingly, evaluation plans should be devel-

oped along with program plans. Developing evaluation plans at this time will help to assure that the goals are specific and that steps are taken to obtain an adequate base line from which to measure change. The evaluation approach and findings can thus be tied in more closely to program needs, goals, and methods so that the findings will be of maximum use to those carrying on the educational program.

Two serious hazards are involved in the alternative practice of developing a list of evaluation techniques and then selecting from this list certain procedures to measure a program's achievement. First, the technique most effective for evaluating one program may prove worthless in another because of differences in goals or in methods of approach. No one technique of evaluation is adequate for all situations. Secondly, there is always the danger that some technique will look particularly interesting and changes will be made in the program to facilitate its use. This aspect of technique influencing the definition of problems and goals is one of the greatest dangers in social research (7).

### **The Base Line**

The measures used to evaluate health education seldom start from absolute zero. Individuals have invariably acquired some health knowledge, attitudes, and practices before a program starts. A base line of zero cannot usually be assumed.

The evaluator who studies the effectiveness of a nutrition education program, for example, must recognize that individuals already have considerable information about food, although some of it may be incorrect. They have also developed definite food attitudes and practices before participating in the program. Some base line of their present knowledge, attitudes, and behavior, therefore, must be obtained prior to estimating changes.

Such a base line should be viewed as a relative rather than an absolute starting point. It usually describes the average situation at the beginning of the program and indicates the degree to which individuals vary about this average. Gains should be interpreted as progress from this average starting point rather than as progress from zero.

A corollary to the preceding concept is that the measurements employed in evaluating health education seldom, if ever, have equal psychological units. The units within a scale of measurement usually vary somewhat, making interpretation of data difficult.

For example, it is erroneous to infer that a person who scores zero on a test has no information, positive attitudes, or practices, or that a person who gets a perfect score is completely informed and wholly positive in attitude and action. Likewise, a person who receives a score of 80 cannot be considered to be twice as educated as one who obtains a score of 40. Even individuals who obtain identical scores cannot be assumed to have identical knowledge or understanding.

### **The Indirect Nature of Measurements**

Evaluative measurements are nearly always indirect measures of the effects of the program on behavior. The findings of evaluation must be interpreted with full recognition of the specific indices used for measurement.

The personal and social variables being measured in a health education program are more complex than the variables in physical objects. For example, units of measurement for weighing yield only indirect evidence of the weight of an object. The standardized unit for such weighing, however, has been in use for so long that weight given in these units is accepted as direct evidence of the object's weight.

No comprehensive standards have been developed to measure all the personal and social variables in human behavior. New scales of measurement must be devised and validated almost every time one attempts to determine the amount and accuracy of a change in a person's information, the direction, extent, and intensity of a shift in attitudes, or the nature and direction of an adjustment in behavior. Each scale of measurement so developed provides only sample evidence on the behavior being measured.

As Thorndike (8) has pointed out, intelligence cannot be measured directly, nor is it known for certain what intelligence represents. Intelligence tests measure the effects or products of intelligence. They have proved useful for interpreting behavior and for making pre-

dictions about future behavior. In interpreting the findings from such tests, however, it is essential to bear in mind that intelligence is inferred from indirect evidence.

Likewise, in attempting to measure attitudes or changes in behavior, the data are limited to the things that can be observed in verbal or physical behavior. Great caution should be observed in imputing meanings or motivations to the data observed.

### **Planning for Future Needs**

Each practical evaluation should be planned to satisfy long-range needs of health education while meeting the immediate project requirements. Each study will then contribute toward the testing of some long-range hypothesis or principle without much additional cost or effort. The findings of different studies will supplement each other in a more meaningful way while unrecognized duplication of effort is minimized. Otherwise, evaluations may yield isolated bits of information that contribute little or nothing toward improving educational methods. The findings should help to answer the challenges so often made: If this won't work, what will? How can this evaluation help us to do it better the next time? How can we relate these results to the findings of other studies?

Thus, evaluators are faced with two questions. The first is: How can evaluations of a specific program be made more meaningful and useful to those responsible for achieving immediate program goals? The answer to this question calls for better experimental design for each study.

The second question is: How can an evaluation of a specific program be made to furnish data for long-range needs? The answer to this more fundamental question requires planning at a broader level. As Marquis (9) has pointed out:

Research planning can be carried out at different levels. I would like to distinguish three levels which I will call experimental design, program design, and policy design. Experimental design is the planning of a single specific project. Program design is the planning of an integrated set of projects focused on a central problem. Policy design is a new word for over-all planning of the

distribution of effort among programs, areas, or fields.

The field of evaluation in health education is particularly weak in the program and policy design levels of research planning. If studies of evaluation are planned in terms of these broader frameworks, the evidences that accumulate will gradually satisfy the long-term as well as the short-term needs of health education and provide a sound basis for program planning. Is it reasonable to anticipate that from such studies might eventually develop a science of health education?

The problems involved in measuring personal and social variables might lead some people to say that the effectiveness of health education in changing behavior cannot be measured. This defeatist viewpoint has no basis in fact. To cite Thorndike, "Whatever exists at all exists in some amount" (10) and can, therefore, be measured. As the programs of health education are more clearly conceptualized, effective means can and will be developed to evaluate them.

### Summary

Basic concepts that should be considered in evaluating any health education program are:

1. Planning evaluation studies of health education requires adequate exploratory fact-finding.

2. The broad purposes of an educational program must be broken down into specific, concrete, definable goals before evaluation can be undertaken.

3. The interrelationship between the findings on the achievement of specific goals must also be considered.

4. The measure of success of an activity is how well it achieves its intended purpose.

5. Concrete evidence that an objective has been achieved is the only realistic criterion for measuring effectiveness.

6. The method used to evaluate a program should be especially selected or constructed for that program in terms of the goals of the program.

7. The measures used to evaluate health education seldom start from absolute zero.

8. Evaluative measurements are nearly always indirect measures of the effects of the program on behavior.

9. Each practical evaluation should be planned to satisfy long-range needs of health education while meeting the immediate project requirements.

### ACKNOWLEDGMENT

The author wishes to acknowledge the assistance of Dr. Mayhew Derryberry, chief of the Division of Public Health Education, and of members of the division who contributed to the development of this paper.

### REFERENCES

- (1) Calver, Homer N., Derryberry, Mayhew, and Mensh, Ivan N.: Use of ratings in the evaluation of exhibits. *Am. J. Pub. Health* **33**: 709-714 (1943).
- (2) Franzen, Raymond: Physical measures of growth and nutrition. School health research monograph No. 2. New York, American Child Health Association, 1929, pp. 1-17.
- (3) Guilford, J. P.: Psychometric methods. New York, McGraw-Hill Book Company, 1936, pp. 1-19.
- (4) Cantril, Hadley: The why of man's experiences. New York, The Macmillan Company, 1950, pp. 59-78.
- (5) Derryberry, Mayhew, and Weissman, Arthur: Evaluation of exhibit material. Read by Derryberry before the Science Museums Section of the American Association of Museums, Columbus, Ohio, May 15, 1941. (Not available for distribution.)
- (6) Knutson, Andie: Evaluating A. P. H. A. exhibits. *Am. J. Pub. Health* **39**: 1027-1035 (1949).
- (7) Cantril, Hadley, Ames, Adelbert, Jr., Hastorf, Albert H., and Ittelson, William H.: Psychology and scientific research. I. The nature of scientific inquiry. II. Scientific inquiry and scientific method. III. The transactional view in psychological research. *Science* **110**: 461-464; 491-497; 517-525 (1949).
- (8) Thorndike, Edward L., and others: The measurement of intelligence. New York, Teachers College, Columbia University, 1928.
- (9) Marquis, Donald G.: Research planning at the frontiers of science. *Am. Psychologist* **3**: 430-438 (1948).
- (10) Thorndike, Edward L.: The nature, purposes and general methods of measurements of educational products. *In* The measurement of educational products. National Society for the Study of Education, 17th Yearbook, Bloomington, Ill., 1918, pp. 1-190.