DOCUMENT RESUME

ED 128 384 TM 005 506

AUTHOR Sjödahl, Lars
TITLE Evaluation of Training.
PUB DATE [Apr 75]
NOTE 12p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Educational Planning; *Educational Programs; *Evaluation Methods; Forestry Occupations; *Job Training; Measurement; Measurement Goals; *Measurement Techniques; Norms; Program Development; Program Planning; Rating Scales; Test Reliability; Test Validity

ABSTRACT

The main purpose of this paper is to draw attention to some facts and ideas that perhaps can help to identify problems or fields for development and research within the evaluation of training. Topics for group discussion are preceded by material on some basic concepts of evaluation and educational measurement. The ratio scale, the interval scale, the ordinal scale, and the nominal scale are given as examples of kinds of scales used in educational measurement; the problem of norms is discussed; potential purposes of evaluation or educational measurement are outlined; and some characteristics of a good measuring instrument are explained. The author also defends the inclusion of evaluation as an integral part of a model for planning and carrying out educational programs. (BW)

EVALUATION OF TRAINING

Lars Sjödahl

Sjödahl, L. Evaluation of training. Didakometry (Malmö, Sweden: School of Education), No. 47, 1975.

This paper is a summary of lectures held at the seminar on pedagogic and organizational problems of forest worker training, Zollikofen 22-26 April 1974, and has been published in Rechsteiner, K. (Ed.) Seminar on pedagogic and organizational problems of forest worker training, 1974, pp 25-37.

The selection of content and the presentation of ideas have been made with the intention to stimulate discussions around possible research and development work within evaluation of forest worker training. The subject matter is arranged under the following headings:

1. Kinds of scales used in educational measurements
2. The problem about norms
3. The purpose of evaluation or educational measurement
4. Some characteristics of a good measuring instrument
5. Evaluation seen as an integrated part of a model for planning and carrying out educational programmes.
6. Subjects for group discussions.

Keywords: Evaluation, forest worker training, educational measurement

EVALUATION OF TRAINING

Summary

The aim of this paper is not primarily to describe an ideal model for evaluation of training systems or training results. The selection of content and the presentation of ideas have been made with the intention to stimulate discussions around possible research and development work within evaluation.

Evaluation always means some kind of judgements. It is true that all judgements do not need to be quantitative in order to be informative and useful and it is the author's opinion that evaluation data should include both qualitative and quantitative descriptions. However, it is important to remember that progress in research is often due to better methods for quantifying observations. Therefore, this paper starts with some fundamental facts about measurements and norm systems.

I may seem obvious that both the evaluation process and the evaluation results should be of use to those people engaged in the training system being evaluated. The fact is on the other hand, that many teachers, for example complain about the lack of relevance and utility of much of the research going on within the field of educational measurement. The purpose of evaluation or educational measurements is therefore taken up for analysis in this paper.

No measurements are better than the instruments used in the measuring procedure. Drawbacks and supposed advantages with different kinds of measuring or assessment methods are often discussed among students, parents, teachers and educational planners without much attention given to possibilities to improve the situation by a better understanding of the measuring procedure and by training those responsible for carrying out the measuring operations. Some characteristics of a good measuring instrument are therefore mentioned in this paper.

The paper concludes with a discussion of evaluation seen as an integrated part of model for the planning and carrying out of educational or training programmes. A suggestion is made that the scope of evaluation should be broadened to include all the behavioural domains regarded as important for the individual's vocational or professional role as well as his social and personal development.

INTRODUCTION

This paper is written from the point of view that progress within any field is made not only by application of skills and knowledge already learnt but also by innovations and identification of new problems. The main purpose of this paper is to draw attention to some facts and ideas that perhaps can help us to identify problems or fields for development and research-work within the evaluation of training. The text is arranged under the following five headings:

- Kinds of scales used in educational measurements
- The problem about norms
- The purpose of evaluation or educational measurements
- Some characteristics of a good measuring instrument
- Evaluation seen as an integrated part of a model for planning and out of educational programmes.

Kinds of scales used in educational measurements

The ratio scale (absolute measurements)

This scale is characterized by an absolute zero and equal units along the scale. The measures obtained with a ratio scale are often called absolute or fundamental measurements. Measures of weights, lengths, areas, angles, etc. all conform to the ratio scale. Only with this kind of measurement can we in an absolute sense make assertions of the nature: "A is twice as heavy as B" or "I spent half the amount that you did". It is only when we have equal intervals and an absolute zero that we can put figures in relation to each other and interpret ratios in an absolute sense.

The interval scale

The interval scale has equal units and an arbitrary zero, i. e. the zero point on an interval scale is a matter of convention. Because the zero point is arbitrary, relations between positions along the scale cannot be interpreted as ratios in an absolute sense although they can be stated in terms of the distance (i. e. number of scale points) between the positions. Thus with an interval scale one cannot state in an absolute meaning that a peron's attitude is twice as favorable as that of another person, just as one cannot state that 20°C is absolutely twice as hot as 10°C. The value of the ratio is relative to the zero point. By moving the zero point (an operation that is permissable

because the zero point is a matter of convention) we also changed the value of our ratios.

The ordinal scale

The ordinal scale is characterized by an arbitrary zero and probably unequal units (intervals). An example of an ordinal scale is the Moh's scale of hardness, which is applied to minerals. With the help of that scale the minerals can be arranged according to the ability of one mineral to scratch another.

The nominal scale

The nominal scale is a classification into categories, between which there are no quantitative relations. The categories are just different in some qualitative respect, e. g. men and women, students studying subject B.

THE PROBLEM ABOUT NORMS

Without any ready-made instruments with equal units and an absolute zero we must try to solve the problem about finding norm systems when using marks and constructing scales for measuring the student's attitudes, knowledge or skills. Just as we always have some kind of points of reference (absolute or arbitrary zero, equal or probably unequal intervals) in physical measurements, we must also have reference points in (educational) behavioural measurements. From a normative point of view we can consider the following kinds of marking or educational measurements.

Individual marking

If every teacher is left to himself in constructing and applying a norm system in his marking we can talk about individual marking.

Relative marking

When all teachers agree upon using some relative performance level, e. g. the average point in the class, as a reference point we get some kind of relative marking. The reference point can in this case be of three kinds:

a. The point of reference is the student's earlier achievement or his assessed potential qualifications. According to this norm system, two objectively equally good achievements could be given different grades. The result obtained by one student might be very good with regard to his potential abilities or earlier achievement level. With the same relative way of marking, the result of another student may be very poor. This kind of marking is called student-relative marking.

b. The point of reference is the average achievement in the single class or group and a certain percentage distribution of the marks is agreed upon in advance to be applied to every single class. This kind of marking is called group- (or class) -relative marking.

c. The point of reference is the average achievement within the classes belonging to the same level or stage and a certain percentage distribution of the marks is decided upon in advance to be applied to the whole stage (not to the single class). This kind of marking is called stage-relative marking.

Absolute marking

When the normsystem consists of "objectively" stated requirements that state what the student has to do to get a certain mark we talk about absolute marking.

It is the author's conviction that many development and research problems have to do with the relative merits and drawbacks connected with different kinds of scales and norm systems.

THE PURPOSE OF EVALUATION OR EDUCATIONAL MEASUREMENT

Educational measurements are always carried out for some purpose. When we talk about purpose we mean the intentions held by test constructors and teachers (instructors) about how to make use of the measurements or evaluation results. There are at least five broad evaluation or measurement purposes, each of which requires, to a ceratin extent, a different approach when we construct measuring instruments.

Systems evaluation

When evaluation data are collected with the aim of illuminating how the total training system under consideration functions, we talk about systems evaluation.

Evaluation for individual prognosis

Evaluation is often carried out to create a basis for marks which are later used as prognostic instruments in selection of applicants for jobs or further training.

Evaluation used for diagnostic information

During training, both students and teachers need evaluation data to guide their further efforts. The sub-goals and sequence of the training activities need to be kept in mind when constructing diagnostic instruments.

External evaluation

Evaluation can also be carried out "in the field", i. e. after the training programme is finished, in order to test the relevance of the training for the job. This kind of evaluation will thus give an opportunity to check the relevance of the training goals as well as the efficiency of the training process.

Evaluation used for research- and development purposes

Evaluation data can also be collected for research purposes. The possibilities for a scientific approach to many problems within education and training is seriously limited by lack of measuring instruments and techniques for data collection.

SOME CHARACTERISTICS OF A GOOD MEASURING INSTRUMENT

The requirements of a good measuring instrument vary to some extent depending on the measuring purpose and the use made of the results. The following three characteristics are usually mentioned in connection with construction of instruments for prognostic use.

Differentiating power

Any measuring instrument must be constructed with the group to be studied in mind. If the tasks (the items of the instrument) are too difficult most of the subjects get low points (scores). Many perhaps have the same total and the differences between the other subjects may be very small. If the test is too easy then most of the subjects

get high totals. Whether the instrument is too difficult or too easy a poor spread is obtained in the results. The differences obtained between the individuals can in such a case be so small, that they are due mainly to chance. As a rule differentiating power is achieved if the test contains tasks of varying degrees of difficulty. A good instrument must have a level of difficulty suitable for the group of students it is going to be applied to.

Reliability = Freedom for random errors

Whenever we measure something errors due to chance influences of varying kinds come in to a lesser or greater extent. The reliability of a measuring procedure can be examined by studying the stability of the measuring results against variations in some of the factors included in the concept of chance.

Validity

A measuring instrument may have good differentiating power and satisfactory reliability but still be of no value if the measurement results cannot be used for what they are intended to be used for. A good measuring instrument must thus apart from being differentiating and reliable also be relevant. The technical term for that characteristic of a good test is "validity". To find the extent of the validity of a measuring procedure we compare the measurement scores with scores from some kind of fairly reliable assessment of the behaviour which we assume our measuring instrument predicts. Progress within research and development work is to a great extent hampered by lack of reliable validity criteria.

EVALUATION SEEN AS AN INTEGRATED PART OF A MODEL FOR PLANNING AND CARRYING OUT EDUCATIONAL PROGRAMMES

A widely used model for planning and carrying out educational or training programmes encompasses the following three main components:

a. Goal seeking and goal description
b. Carrying out the educational or instructional process
c. Evaluation and revision

Education is here seen as one of three main components which linked together constitute what is called a technological model for educational planning. These main components can be broken up into

subparts each of which has its own research and development problems. Educational evaluation is usually confined to observations, measurements and judgements about educational products (=changes of student behaviour) in order to compare achieved results with pre-specified objectives. This rather narrow way of looking at evaluation can be supplemented with the following points of view:

- The educational and instructional process is usually so complex that many other effects than those pre-specified by the curriculum will be achieved. These unanticipated effects on the students may be of great interest if we knew something about them - still they are likely to be completely overlooked if the evaluation study is restricted to those goals prescribed in the curriculum. Methods should therefore be developed to register even "products" that are not specified in advance.

- Even the operating system, i. e. the instructional - and learning process, should be evaluated. Predescribed goals and recommended methods are usually redefined or changed by the individual teacher when implemented in practice. Many of these unforeseen changes or deviations from the "ideal pattern" may turn out to be valuable innovations worth greater attention. Methods should therefore be developed to observe and describe the educational and instructional process and to study cause-effect relations within the instructional system.

- Instruction, teaching and learning always proceed in a learning milieu. This milieu is of two kinds. Both teachers and students work in a social-psychological as well as in a material environment. Methods are needed to describe the learning milieu in relevant aspects and also to study milieu-factors as cause or effect variables.

- Despite common knowledge that individuals are all different and the widespread adherence to the principle of individualization, most evaluation studies deal with group results. The traditional evaluation approach should therefore be supplemented with individual case-studies. There is thus a need for developing methods or guidelines for writing such case-studies.

- The scope of evaluation should be broadened to include all the behavioural domains regarded as important for the individual's vocational or professional role as well as his social and strictly personal development.

  Without any pretence of being complete the following list of behavioural domains can be given:

  a. Knowledge of facts
  b. Complex cognitive behaviour like problem solving, analysis, synthesis etc.
  c. Social skills
  d. Manual and motor skills
  e. Affective or emotional reactions like giving priority to values, aesthetic perception and judgements
  f. Creativity
  g. Attitudes
  h. Personality traits

The author is well aware of the controversial issue about making affective and emotional behaviour an object of educational evaluation but still thinks that the scope of evaluation should be put to thorough discussion.

SUBJECTS FOR GROUP DISCUSSIONS

Discussion 1

Discuss within the group which kinds of scales are used for evaluation of forestry training in different subjects. Sum up the discussion by making a list of the subjects and corresponding types of scales.

Example:

Subject	Evaluation method	Type of scale
Silviculture	Written examinations	ordinal
	Ratings of practical exercises	ordinal
	Time used to plant a certain amount of plants	ratio

Discussion 2

Discuss within the group the relative merits and draw-backs with the following kinds of normsystems in marking and evaluation. Which kind of normsystem is used in your country in forestry training? Is it possible to use more than one single normsystem within one and the

same training system?

1. Individual marking
2. Student-relative marking
3. Group- (or class-) relative marking
4. Stage-relative marking
5. Absolute marking

Discussion 3

Are the marks given to the students in forestry training actually used later as instruments in selection of applicants for jobs or further training or are other selection criteria used? For which subjects and jobs are marks most needed as tools for selection?

Discussion 4

Both students and teachers need evaluation data to guide their further efforts. Discuss within the group the need for diagnostic tools (rating scales, examinations etc.) within forestry training. Give some example of how diagnostic evaluation is used in some subjects and try to list those conditions that make it more important to develop diagnostic instruments in some subjects than in others.

Discussion 5

Are there some subjects, where external evaluation is more important than in others? Discuss within the group how external evaluation should be carried out and how the results should be used.

Discussion 6

To what extent are the forestry teachers in your country trained to construct measuring instruments that satisfy the requirements of reliability, validity and differentiating power? Describe what the student-teachers do when they acquire the skill to construct good measuring instruments.

Discussion 7

A detailed model for planning and carrying out educational or training programmes has just been presented. Discuss within the group from which components in the model the most urgent research and development problems should be drawn.

Discussion 8

When seeking the training goals how do we make certain that the stated goals are valid and not perhaps subjective opinions of a few experts?

The problem can be split up in the following three questions:

1. Which methods and instruments do we use when we try to find the training goals?
2. How do we distinguish between a valid (=relevant) and an invalid (=irrelevant) training goal?
3. How often do we need to up-date our training goals?

Discussion 9

Discuss within the group the scope of evaluation in forestry training. Try to rank the behavioural domains listed below according to their relevance to forestry training. Which are the difficulties in deciding if any one of the behavioural domains below should be included in a programme for evaluation of forestry training.

Behavioural domains

a. Knowledge of facts
b. Complex cognitive behaviour like problem solving, analysis, synthesis etc.
c. Social skills
d. Manual and motor skills
e. Affective or emotional reactions like giving priority to values, aesthetic perception and judgements
f. Creativity
g. Attitudes
h. Personality traits