doc. 395

Reunión
Interamericana de Bibliotecarios y Documentalistas Agrícolas

– San José, Costa Rica, 10 - 14 Abril, 1978 –

• SECRETARIA GENERAL
Biblioteca
Universidad de Costa Rica
Ciudad Universitaria
San Pedro de Montes de Oca
Costa Rica

• SECRETARIA DE AIBDA
c/o IICA-CIDIA
Apartado 74
Turrialba, Costa Rica

CODIGO I.3

THE EVALUATION OF COURSES AND OTHER EDUCATIONAL OFFERINGS IN THE FIELD OF DOCUMENTATION

F. W. Lancaster
Professor
Graduate School of Library Science
University of Illinois
Urbana, IL 61801
U.S.A.

CODIGO I.3

5 a. REUNION INTERAMERICANA DE BIBLIOTECARIOS Y DOCUMENTALISTAS AGRICOLAS, San Jose, Costa Rica, 10-14 April, 1978

THE EVALUATION OF COURSES AND OTHER EDUCATIONAL OFFERINGS IN THE FIELD OF DOCUMENTATION

F. W. Lancaster
Professor
Graduate School of Library Science
University of Illinois
Urbana, IL 61801
U.S.A.

Various approaches to the evaluation of educational programs are described. A distinction is made between formative evaluation and summative evaluation. Procedures are described for: (1) the evaluation of participant reaction, (2) the evaluation of learning acquired, (3) the evaluation of behavioral change, and (4) the evaluation of program results. Validity, reliability, and universality of evaluation results, efficiency of evaluation procedures, and application of evaluation results are also dealt with. The paper is based on the author's Guidelines for the Evaluation of Training Courses, Workshops and Seminars in Scientific and Technical Information and Documentation (UNESCO, 1975).

Introduction

Evaluation procedures should be built into the educational process from the early planning stages of a particular program. As indicated by Hampton (1), "evaluation must occur before, throughout and beyond the entire program." Any program for education and training in documentation needs, therefore, to be evaluated at various stages or levels, as follows:

(1) evaluating the need for various types of activity in order to be able to establish meaningful priorities;

(2) evaluating the goals and objectives of proposed activities to determine if they are, in fact, realistic;

(3) evaluating plans for the achievement of stated goals and objectives;

(4) evaluating educational activities while in operation;

(5) evaluating educational activities in retrospect to determine if they did in fact achieve their stated objectives.

This paper, which is based largely on guidelines I prepared for UNESCO (2), relates to the fourth and fifth stages of evaluation identified above. The purposes of this evaluation are twofold: (a) to improve the quality and relevance of programs while they are in progress, and (b) to provide data that will allow the organization responsible for planning the program to decide how successful a particular activity has been, how far it has contributed to the attainment of stated objectives, and what might be done to improve educational offerings of this type in the future.

Throughout this paper, "course" is used generically to refer to any type of educational offerings in the field of scientific documentation.

Approaches to Evaluation in Education

The case for evaluation in education has been very well stated by Hampton (1):

> "Evaluation is one of the most neglected aspects of programming in continuing education today. Yet it has the potential of being one of the most powerful administrative tools available to educators. Most program planners do a respectable job of planning and carrying our their plans to a high degree of proficiency. At this point, however, the program development process tends to break down, for too often planners fail to review their program adequately before directing their attention toward planning another program. Information derived from reliable appraisals of educational efforts is essential to the educator in order to rationalize decision-making with respect to improving future programming efforts. Formal and informal judgements made about the program should logically evolve into a plan of operation, which ultimately may affect program process, teaching methods and program accomplishments".

Hampton insists that evaluation should be regarded as an integral part of the educational process rather than just an adjunct. Evaluation is an ongoing activity and it should be built in from the very beginning (i.e., from the initial planning stages). Knox (3) identifies four questions that need to be answered in making decisions relating to evaluation: (1) who needs to know how effective the program is, (2) what do they need to know, (3) who can best provide this information, and (4) how can it be obtained most effectively and efficiently?

There are many ways in which approaches to the evaluation of education have been categorized in the literature. Perhaps the most important distinction is that between summative evaluation and formative evaluation, a distinction that appears to have been made first by Scriven (4). Summative evaluation is essentially retrospective. At the end of a course an evaluation is conducted to determine how successful it has been. Summative evaluation has value in that, if properly conducted, it will reveal failures or limitations of the course and suggest ways in which this or similar courses might be improved

in the future. Summative evaluation can lead to improvements in future educational endeavors but, clearly, it cannot improve the course that has already been completed. Formative evaluation, on the other hand, is conducted while the course is in progress, the objective being to improve this particular educational experience before it is completed. To take a simple example, suppose we have a course that lasts for three weeks. We evaluate the ongoing course at the end of the first week and again at the end of the second. On the basis of feedback from participants, we may be able to make changes to the course in order to make it a more meaningful educational experience in the second and third weeks. That is, we may change the teaching approach or the direction or emphasis of the course content. In the case of a course of one week duration we can obtain daily reaction from participants. It is obvious that formative evaluation of this kind is more feasible with courses of relatively long duration. There is little possibility for the formative evaluation of a one-day seminar. Formative evaluation is not necessarily purely subjective. It is perfectly possible to have a formative test of learning acquired by students. Such a test might, for example, determine whether or not the students can demonstrate the achievement of certain behavioral objectives established for a course.

Whether it takes formative or summative form, the major purpose of evaluation in education is, according to Steele (5):

> "to determine the effects of teaching under given conditions on the knowledge, attitudes, and behaviors of those being taught to provide a basis for improving, justifying, or discontinuing the teaching activity".

Evaluation can be conducted from a number of different viewpoints, the most important being:

1. that of the participants;
2. that of the lecturers;
3. that of those responsible for planning, administering or funding the program.

The teaching staff itself should be most interested in formative evaluation, the object being to improve the quality of program content and their own teaching methods. Steele (5) has stated the need for formative evaluation in this way:

> "Not only must a constant check be kept on how participants are reacting, but evidence of negative reaction must be fed back into the program immediately so adjustments can be made".

Course participants will also be concerned with formative evaluation; they will want to provide feedback to the teaching staff in order, where necessary, to change the direction or emphasis of the course and thereby make it more responsive to their own requirements. Participants will also want to know how well they are progressing in a course (in terms of their own objectives or those set by the teaching staff) and how they might be able to apply the learning gained. As Knox (3) points out, the participant may also want to discover what he does and does not know relative to the subject of the course - so that he may concentrate on what he does not know. Potential participants in a program, on the other hand, are likely to be most interested in summative evaluations of previous courses organized by the same body and in the same subject area or one closely related. Employers may also find such evaluations useful in deciding whether to send participants to a particular course and, if so, which people to send.

Planners and administrators of educational programs, on the other hand, are likely to want evaluations that are more comprehensive and far-reaching. Although they may want to see evaluation results from individual courses, they will also need to evaluate complete programs and they are likely to be concerned with

the total impact of the program on all participants.

Evaluation may be formal or it may be informal. Informal evaluation of an educational offering occurs all the time. A lecturer will solicit feedback from participants at a workshop or seminar, over coffee, at lunch, or on other occasions in which he mixes with students. Even if he does not actively solicit evaluative data of this kind, he is likely to receive it gratuitously. Whether or not they are asked to do so, students (consciously or unconsciously) will conduct their own evaluation of any course or workshop they attend. That is, they will make their own judgment about quality and relevance of the content, the success of the approaches to teaching, and so on. By mixing with students an independent observer may be able to obtain a fairly clear idea of how the course is going. The teacher, also, is likely to sense how well a course is going and will be evaluating himself as it progresses. In contrast to this informal or casual evaluation, formal evaluation implies a deliberate effort on the part of the organizers to get some assessment of the success of a course. A formal evaluation will involve some "standardized" approach to data gathering, usually by the use of a questionnaire, interviews with the participants, or some controlled comparison or test.

Another distinction worth making is that between internal evaluation and external evaluation. Internal evaluation is performed by those who conduct the educational program and it is usually formative. External evaluation is performed by individuals somewhat removed from the program. External evaluation is mostly summative.

There is still another classification worth mentioning: the distinction between subjective and objective evaluation. Subjective evaluation is more easily accomplished than an objective study. Subjective evaluation is based on <u>opinions</u>, of students, of teachers or of an independent observer. Objective evaluation, on the other hand, attempts to move away from opinion, pure and simple,

and to come up with an assessment that is more standardized and perhaps more quantifiable. An obvious example of an objective evaluation is one in which the success of a course is measured by testing the knowledge or abilities of the students before and after their participation. Some standardized test is applied before the course and again when the course is completed, the purpose being to measure the change in the students as a result of their participation. Presumably, if the educational experience has been successful, the students will achieve significantly higher "scores" on the second test than on the first. A variation on this is the use of some standard test applied to two matched groups of students, each group having been exposed to a different method of presenting the same material, the object here being to compare the success of one approach with that of the other.

One of the clearest expositions of approaches to evaluation is that given by Hampton (1), who identifies four possible "steps":

(1) evaluating reaction of participants;

(2) evaluating learning acquired;

(3) evaluating behavioral change;

(4) evaluating program results.

Evaluation of reaction (of students, lecturers, observers) is easiest to accomplish. Such evaluation is completely subjective, although the data may be gathered in a systematic way and in a consistent form. These data may also be quantifiable in some sense (e.g., 80% of the participants were satisfied with the approach used in teaching). Evaluation of learning, on the other hand, can only be achieved by some objective procedure, usually some form of test. More difficult still is the evaluation of behavioral change in the participants. This goes beyond learning as such into the application of the learning acquired. It is possible for a student to "learn" a particular skill (in the sense that he can pass some test which may be based on memorization) but still not be able

to apply it in a practical situation. One approach to the evaluation of behavioral changes is by the measurement of on-the-job performance of an individual before his participation in some educational program and again some time after his participation. Preferably this evaluation should be objective. Conceivably, however, it could be subjective, based on the opinions of his peers or supervisors. Another possible method of measuring behavioral change is by the use of tests of problem solving or decision-making abilities. Program evaluation differs from the types mentioned earlier in scale rather than in approach or form. The evaluation of an education program is of concern to those who plan and administer it. Program evaluation implies the existence of a set of program objectives. The evaluation is conducted to determine how well these objectives have been met. Clearly, a complete program evaluation could involve studies of reaction, learning or behavioral changes, or all three of these, depending on what the objectives of the program happen to be.

Hampton also points out that we can categorize evaluation activities by the stage at which these activities are performed, as follows:

(1) pre-course evaluation, (2) ongoing evaluation, (3) terminal evaluation, and (4) follow-up evaluation.

These evaluation stages are closely related to and identifiable with the various "approaches" mentioned earlier. Ongoing evaluation is likely to be formative while terminal evalution is, by definition, summative. Follow-up evaluation is likely to be used in the measurement of behavioral changes, while the ongoing evaluation and the terminal evaluation are likely to be based on measurements of reaction and learning. Pre-course evaluation may be conducted for the purpose of evaluating learning or it may be used as one element in a more complete program evaluation (i.e., some form of pre-course evaluation,

in this case, may be conducted in order to establish meaningful objectives for the program).

Evaluation of Reaction

The most important aspect of this form of evaluation is presumably the reaction of the students attending the course. Studies of participants reaction will tend to be subjective. That is, we use various methods to determine the opinions of the participants relating to the course in general and, possibly, certain specific features of it. At the most general level, an evaluation of reaction seeks to determine how "happy" the students are with the way a course is progressing or with the way the course was conducted. In fact, the type of data collected in this form of evaluation has been referred to as "happiness data" or as a "happiness index".

Evaluation of student reaction has definite value. As Hampton (1) points out:

> "It is important to know how people feel about the programs they attend, for it is reasonable to expect that participants who enjoy a program are more likely to obtain maximum benefit from it".

Knowles (6) has stated that:

> "On the whole, this kind of feedback is most useful in providing a general sense of trends in morale and satisfaction, but it frequently turns up specific and practical suggestions for improvement in the general program or in specific activities; and it may reveal problem points that call for deeper evaluation".

"Reaction data" can be gathered for the purpose of formative evaluation or they can be gathered for summative evaluation. Both formal and informal procedures can be applied. At a more informal level, the teaching staff can ask the students for their unstructured, "off the cuff" impressions of the course. With a relatively small group, this could possibly be achieved through an informal discussion with the entire group at the end of each day of meetings. At an even

more informal level, it is quite likely that useful reaction data can be obtained through discussions, between students and teaching staff, over coffee, at lunch, and during organized social events. An independent outside evaluator can obtain reaction data in a similar informal way, by observing the progress of a course and by generally "mixing" with the students. The outside observer may, in fact, be in a much better position than the teaching staff to gather informal reactions in this way. Participants who may be reluctant to criticize an instructor to his face may be more candid when asked for their opinions by some apparently impartial observer.

A more formal approach to the collection of reaction data will use some form of structured instrument for data gathering. Usually this will be a questionnaire completed by each student, probably anonymously, although interviews may be used in place of the questionnaire. If interviews are used, it is important that they be conducted in some consistent manner, following an interview guide of some kind. Questionnaires, although widely used and accepted as survey instruments in social science research, are criticized by some investigators, usually for two major reasons:

(1) questions may be misinterpreted by respondents and it is sometimes difficult to know whether or not the respondent has interpreted a particular question in the way the designer of the instrument intended;

(2) there is sometimes some doubt as to whether a respondent has answered truthfully or accurately and there may be no convenient or practical way to check the validity or accuracy of a response.

These objectives are unlikely to be too serious in the situation under consideration here. The number of people attending most programs will be sufficiently small that it will be possible to have someone in attendance while questionnaires are being completed with the group. This person (one of the lecturers perhaps) will be available to "interpret" the questionnaire to participants and to answer any questions they may have on how to complete it.

The question of veracity or validity of responses is unlikely to apply to a student evaluation of a course. This is a problem that is more likely to apply to a situation in which the respondent is, in some sense, evaluating himself. For example, there may be a tendency for a respondent to overestimate the number of journals he reads, the number of hours he spends in reading, or the number of his own publications. This is a matter of prestige and of the desire of the respondent to appear "in the best light". There is no reason to suppose, however, that the participant in a course would have any reason to be untruthful or otherwise inaccurate in the completion of a questionnaire - the nature of the questions asked virtually precludes the possibility of this problem occurring.

Interviews seem to have three major advantages over questionnaires as methods of data gathering:

(1) the presence of the interviewer tends to ensure that all questions are correctly interpreted by the respondent;

(2) it may be possible, by means of "probing" questions, for the interviewer to check on the accuracy of responses;

(3) the interviewer may be able to collect unsolicited observations from the person interviewed. Data unanticipated in the interview may thus be collected.

We have already indicated that the first two of these advantages are unlikely to be very significant in the evaluation of programs in this field, and the third benefit does not seem sufficiently important to warrant the use of interviews in place of questionnaires. Interviews are more expensive and time-consuming. They require scheduling of participants, which may not be at all easy to arrange, particularly at the end of a course when most of the attendees are likely to be anxious to leave. Moreover, interviews cannot be completed anonymously, unlike questionnaires, and, for reasons stated earlier, they would

require use of an independent interviewer. Members of the teaching staff should not conduct such interviews: they are unlikely to get completely candid responses and they may, perhaps completely unwittingly, influence the responses by the way they pose the questions. For formal measurement of student reaction, then, the printed questionnaire is likely to be the preferred instrument for gathering data.

To allow meaningful evaluation, on the basis of reaction data, it is important that some statement of the objectives of the course be available to participants. One important facet of the evaluation will be the students' assessment of how well the course has met its stated objectives. If no stated objectives exist, or are not available to students, the possibility exists that they will evaluate the course in relation to goals that the organizers never intended. While it will also be important to learn of the personal objectives of students, and their judgement on how far these personal objectives have been met, a wide discrepancy between the objectives of the students and the objectives of the organizers would tend to indicate a failure in the advertising and promotion of the course or, possibly, the selection of students to participate.

But evaluation should not be solely on the basis of stated objectives. A course may have benefits to the particpants that were not anticipated by the organizers. In some programs, for example, the value of the social contact among students may be greater than the value of the formal presentations. It is possible that a course might receive a rather poor score in relation to its stated objectives but still be a valuable educational experience for some other reason. It is important, therefore, that the evaluation of any educational enterprise should be sufficiently open-ended that it will take account of benefits that are apparent to the students although not anticipated by those responsible for planning and organizing the endeavor. Parenthetically, it is worth mentioning that certain educational experiences may also have undesirable

consequences that were unanticipated by the organizers or the teachers. For example, it is conceivable that a course on "literary appreciation" may cause certain students to hate reading, especially is the course requires the student to read a large body of literature in which he may have little or no interest. Harmful effects of this kind are more likely to occur in the basic education of children or young people. They are unlikely to occur in the continuing professional education of an adult population.

As stated earlier, studies of participant reaction may be used in both formative evaluation and summative evaluation. It may be worthwhile at this point to consider the kinds of questions that may be asked of students in each of these types of evaluation.

Let us first consider the subject of formative evaluation and the kinds of information we need to gather, by questionnaire, at the end of each day or at the end of each week for the duration of the course. The following are probably the most important categories:

1. The importance and relevance of the subject matter covered up to this point. If the course can be divided up into a series of discrete sections, we will probably want an assessment of the relevance of each.

2. The quality of the teaching in terms of how the material has been presented. If a number of different lecturers are involved, we may wish the students to rate the performance of each, particularly if the same teaching staff is to perform again later in the course or in future courses.

If a number of different teaching approaches or teaching materials have been used in the course (e.g., films, transparencies, demonstrations, lectures, discussion groups), we will certainly want the students to indicate how successful each has been as a means of presenting the information.

3. The novelty of the information presented. Except in the case of a "refresher" course, a student will normally attend an educational offering in order to learn something new. We should therefore be interested in discovering how much the student is really learning (i.e., what proportion of the material is new to him). Both relevance and novelty are important in the evaluation of teaching. The subject-matter of a course may be relevant to a student's interests but not new to him or it may be new to him but not directly relevant to his needs.

4. The "level" of the material presented. It will be important to know, as a course is progressing, whether the material presented is at a level appropriate for the particular audience addressed. It should not be too simple. The students must not feel that they are being "talked down to". Simplicity may also be related to novelty - the students may be learning nothing new because the lecturers have underestimated their prior knowledge of the subject matter and are treating the material at too basic a level. Neither, of course, should the material presented be too complex for the audience (i.e., it should not be "over their heads"). This matter of level of treatment is extremely important in formative evaluation. We must know, early in the course, whether or not we are "reaching" our audience. There is no point in waiting until the end of the course only to discover that the majority of the participants have been completely bored, either because the material has been redundant or because they were unable to follow it.

With the type of audience that can be expected for most programs, in this field, it may be extremely difficult to conduct a course at a level that will be appropriate to all. It may be necessary to assume, at the outset, a minimum of knowledge on the part of all participants and to use the early part of the course to bring all participants up to about the same level, even if this means that some of them will find this early material redundant.

5. How well the student feels that he or she is progressing toward satisfying his or her own objectives in attending the course.

6. An indication of which features of the course, as conducted so far, have been most valuable or interesting and which have been least valuable or interesting.

7. If outside projects or readings have been used in the course, the student assessment of the value and relevance of these experiences.

8. Any suggestions the students have as to how the course might be changed, in later days, to make it a more valuable educational experience.

9. The overall rating of the course, on some scale, by all participants.

10. Any other observations or comments the students would like to make.

A formative evaluation does not necessarily have to be as highly structured as we have suggested. There may be something to be said for a more free-form approach. Woodward and Yeager (7), for example, used daily logs in which students record general impressions of each day in an anecdotal, free-response form. In this they were guided only by very general headings such as "problems encountered". Moreover, a formative evaluation does not necessarily have to be done on a daily or weekly basis. Instead, a formative questionnaire may be developed for each section of the course and administered when each section is completed.

In all of this data gathering it will be important to discover why various responses are made. If students express dissatisfaction we must learn the precise nature of this dissatisfaction if we are to make changes designed to improve the course. Thus, we must discover exactly why a student feels he is not making satisfactory progress in the course, which parts of it he has found too complex, which parts repetitious, and so on. This implies that the questionnaire should be designed in such a way that, whenever the student

indicates some form of dissatisfaction, he is asked to describe the precise nature of this dissatisfaction.

In the above discussion we have noted the kinds of data we will need to collect in order to conduct a formative evaluation of some educational offering. These data are summarized in Table 1. The tabulation is undoubtedly incomplete. In particular, each course or workshop may have some special features that require emphasis in the evaluation questionnaire. Nevertheless these data seem to be those of greatest overall importance and they should be generally applicable in the formative evaluation of most of the activities with which organizers of courses in this field will be involved. Some examples of questionnaires used in formative evaluation are given in the UNESCO Guidelines (2).

In the summative evaluation of the course we may want to ask some of the same questions that were used in its formative evaluation, this time addressed to the course as a whole rather than just a part of it. We are also likely to include some types of questions that we did not use in formative evaluation. Some of the major questions to be answered in a summative evaluation are presented in Table 2. Most are self-explanatory. At this point it will be necessary to ask students to view the course in its entirety as an educational experience, including various factors that may have contributed to an atmosphere conducive to learning. These factors will include physical arrangements (the lecture and meeting facilities and hotel or other accommodations) and organized social events. In this summative evaluation we are concerned with looking back on the course, to determine how successful it was in toto and to identify any failures or problem areas that may have occurred. On the basis of this experience we may be able to make valuable changes to the course in the future and we can draw on this experience in the planning and execution of future courses in other subject areas.

An important element in the summative evaluation will be the determination of how well the objectives of the students have been satisfied. It will be important to determine, for each student, what his or her objectives were in attending the course, and how far these objectives were actually met. In this connection, it might well be desirable to use some very brief "pre-course questionnaire" (perhaps incorporated into the actual application form) to determine student objectives before the course begins. At the end of the course, we can present to each participant the objectives he specified before the course began. He can then be asked (a) to modify his statement of objectives if he feels that, in retrospect, it was too narrow, too broad, or otherwise inappropriate, and (b) to judge the success of the course in meeting his original or revised objectives. It will also be desirable to include a statement of the objectives of the course as prepared by the organizers of the course and to ask students how far they feel that these general objectives have been met.

As in the formative evaluation, it will be important in summative evaluation to discover <u>why</u>, especially in situations in which the student has indicated dissatisfaction with some aspect of the program. Some examples of questionnaires used in the summative evaluation of courses are given in the UNESCO Guidelines (2) and in Slater (8).

Table 1

Types of Data to be Collected in the

Formative Evaluation of a Course

1. Importance and relevance of subject matter.
2. Quality of the presentation
    a) by individual instructor
    b) by form of presentation used.
3. Novelty of the information transmitted.
4. Intellectual "level" of the material presented.
5. Students' assessment of their own progress.
6. Most and least valuable features of the course so far.
7. Value of outside projects and/or readings.
8. Student suggestions for improvement of the course.
9. Overall student rating of the course.
10. Any other observations the student wishes to make.

Table 2

Some Important Questions to be Answered in the Summative Evaluation of a Course

1. Have student's objectives been met? Have overall course objectives been met?
2. How valuable has the course been to participants in terms of their own professional goals? What were the main benefits gained?
3. How do students rate the quality of the instruction
   a) by individual instructor
   b) by type of learning experience (lecture, discussion, practical sessions, and so on)?
4. Size of group participating--too large, too small, about right?
5. Was the level of treatment of the subject matter at the "correct" level for the majority of the audience?
6. Was coverage of subject matter comprehensive or were there notable gaps in coverage?
7. Were the lectures and/or other educational experiences well integrated into a meaningful and complete program, or were they disjointed and overlapping?
8. Was most of the material presented "new" to most of the participants, or was it repetitive and redundant?
9. Which features of the course were most valuable and which least valuable to those attending? Another way of putting this is in the form of features that participants would like to add to or delete from the course.
10. Were the physical and related arrangements (lecture halls, accommodations, meals and/or coffee breaks, social events) satisfactory?
11. Would participants recommend this course to their colleagues and would they attend another, similar course if it were offered?
12. Any other ways the course could be improved?

In a course of fairly long duration, in which a somewhat wide range of related topics is discussed, it will be necessary for the lecturers to develop a set of behavioral objectives for the course, with one or more behavioral objectives associated with each segment of the course. If, at the conclusion of the course, each student is asked to indicate how far he feels these objective have been met, we can legitimately regard this as a part of our evaluation of reaction. If, on the other hand, we develop some "demonstration of achievement" for each of the objectives, and use some objective testing procedures to determine how far the objectives have been met, we are really measuring the learning of the students, and this form of evaluation then falls legitimately in the category "evaluation of learning acquired". In Table 3 are presented, by way of an example, some specific behavioral objectives, together with demonstrations of achievement, that were developed for the Unesco International Training Course on Information Retrieval and Information Retrieval Systems, held in Katowice, Poland, in August 1974.

So far we have discussed only the reaction of students. While student reaction is obviously of great importance, it is not the only reaction data of interest. Whenever it can be arranged, it may be highly desirable to have some independent observer present at some or all of the presentations. Such an observer can serve two functions: (1) to mix with students and obtain feedback and reaction that may be more "candid" and impromptu than the reaction data obtained by the teaching staff through more formal procedures, and (2) to observe the teachers in action and thus to arrive at his own assessment of the quality of the course, both in terms of its content and of the methods of presentation. This second function may be regarded as a form of "peer review". It requires, therefore, the use of an individual who is knowledgeable in the subject-matter covered in the training program and at approximately the same level of seniority and professional experience as the members of the teaching

staff themselves. It may be desirable to provide this outside observer with some type of checklist on which he can record, in standardized form, his observations on course content, teaching methods, student interest, and whatever else we may be concerned about.

Another "reaction study" will involve the teaching staff. It will be valuable to have the lecturers evaluate the course from their point of view, and to prepare a summary of this evaluation once the course in concluded. A good lecturer will constantly be evaluating himself. He is likely to recognize the fact that certain material was not presented as clearly or as completely as he would have liked, that a different sequence of presentation would have improved the situation, that certain types of information could be better presented in an alternative way (e.g., by student readings and group discussion instead of straight lecturing), and that certain topics might be omitted entirely in the future, because they were redundant, of only marginal relevance or interest, or because they were clearly boring to the audience. The lecturers' evaluation should also include their observation on the students - their quality, degree of interest in the subject-matter, the intelligence of the questions asked, their diligence (e.g., in completing background reading and other assignments), and their general suitability for participating in a course of this type. It is possible that a course may fail to meet part of its objectives because some of the students did not have the background necessary to benefit fully from the experience. This could be a significant problem in the case of a program that may be attended by participants from many different countries, with widely divergent backgrounds, and varying levels of competence in the language in which the course is offered. In any case, the lecturers' evaluation of the course, and of the student body, is likely to be an important element in a complete evaluation program and such an evaluation may have great significance

in the advertising of future courses and in establishing appropriate criteria for the selection of candidates. The teaching staff should also evaluate the "environment" of the course: the adequacy of the facilities, the helpfulness of the local organizers, and other factors that may add to or detract from the educational aspects.

Knowles (6) has mentioned that instructors have great shortcomings as observers because they "are personally involved in the outcome of the evaluation, so that it may be difficult for them to be objective. They may tend to overlook instances in which desired changes are not being produced and to emphasize minor successes". On the other hand, it could be argued that the reverse situation might also be true. Some instructors may be too sensitive, too self-critical. They may find faults where none really exist. Whatever the limitations of the instructor as an evaluator, however, it is clear that he has an important role to play. His input will be analysed and interpreted along with the input from students, independent observers and other individuals who are involved in some way with the educational experience. It is important that a complete evaluation should be based on input from a number of individuals representing different levels of involvement and points of view.

Table 3

SAMPLE OF

BEHAVIORAL OBJECTIVES FOR A COURSE IN INFORMATION STORAGE AND RETRIEVAL

Objectives	Demonstration of Achievement
1. Understand the input and output operations involved in modern computer-based IR systems.	1. Be able to describe the functions of such a system in terms one of your own students would understand.
2. Recognize the advantages of on-line and off-line IR systems.	2. Be able to enumerate these advantages and discuss them clearly and concisely.
3. Understand how search strategies are constructed and understand the Boolean logic involved in such strategies.	3. Be able to conceptually analyse a request for information, represent the relationships graphically by Venn diagrams, reduce the conceptual analysis to a set of index terms and logical operators, and reduce eventually to a simple equation.
4. Understand the distinctions between pre-coordinate and post-coordinate retrieval systems and the advantages and disadvantages of both types.	4. Be able to categorize any system as pre-coordinate or post-coordinate. Be able to write a paper clearly distinguishing the two types, with examples, and comparing the pros and cons of each.
5. Understand the manual and semi-mechanized post-coordinate systems that were the forerunners of modern machine-based systems.	5. Be able to set up simple card files to demonstrate the peek-a-boo, uniterm and edge-notched card principles.
6. Recognize the principal types of tools or services by which a scientist may keep current in the literature of his subject field.	6. Be able to write a paper describing the current awareness services that you would provide were you appointed to the post of technical librarian or information officer in an industrial organization.

Evaluation of Learning Acquired

The evaluation of the learning acquired by students is more difficult than evaluation on the basis of student reaction only; it requires the use of more objective procedures. One possible approach, mentioned in the preceding section, is the use of behavioral objectives, with accompanying "demonstrations of achievement", as the basis for measuring the learning imparted. Some "demonstrations of achievement" can be tested objectively. For example, to show understanding of the hierarchial and cross-reference structure of a thesaurus, the student can be asked to construct a sample of four or five thesaurus pages in a subject area of his choice. A complete curriculum may be constructed in the form of a collection of behavioral objectives.

Another, somewhat similar procedure would involve the use of a brief test administered to students before they participate in the course. Such a test could be made part of a pre-course questionnaire, which is also used to determine the student's expectations concerning the course. The test would be completed anonymously (to avoid possible embarrassment on the part of students, who might feel uncomfortable if they were unable to answer many of the questions posed). This type of test has value for two purposes:

(1) the teaching staff will gain a better idea of what the students, as a group, already know and what they do not know;

(2) the same set of questions can be used, when the course is completed, to test the student again and thus measure the change in their knowledge, at least as far as this set of questions is concerned.

This technique of pre- and post-testing of student knowledge is likely to be most easy to apply where the questions can be made reasonably factual.

In actual fact, there is some possibility of bias associated with the technique as described above. If the lecturing staff prepare the set of questions there is a danger that, consciously or unconsciously, they will give particular

emphasis to these questions in their actual presentations. If the same questions are then used at the end of the course the "evaluation of learning" may be somewhat biased in favor of the course since it is possible that other questions, although equally important, might not be answered nearly as well by the group of students. One way of avoiding this possible bias is to have the questions composed by an outsider experienced in this subject field (perhaps the person who was mentioned as an independent observer in our discussion of the measurement of student reaction), his questions being based on the course outline and its behavioral objectives. This set of test questions would be answered by students, to the best of their abilities, at the very beginning of the course, and would again be used at the end of the course. They would not, however, be seen by members of the teaching staff until after the course is completed.

This technique, although an improvement on the one first mentioned, also has some slight possibility of bias. It is possible that the students themselves, remembering the questions asked in the pre-course questionnaire, will themselves concentrate on these topics in their study, possibly to the exclusion of other topics of equal or greater importance. This could be true even when the student is given no indication that he will be tested again at the end of the course, a condition which is essential to the conduct of an evaluation of this type.

The possibility of bias in the evaluation results could be reduced by a slightly more complex design. This would involve a "cross-over" test in which the set of questions, as well as the group of students, is divided up into two sets. Question set A is administered to student group A before the course and to student group B after the course, and vice versa for question set B, as shown below:

Student group	Pre-course	Post-course
A	A questions	B questions
B	B questions	A questions

It could be argued, however, that such a test might bias the results against the course because A students might again concentrate, in their study, on A questions and thus do unnecessarily poorly on the B questions at the end, and vice versa for the B students. It is also likely that students from both groups might discuss the test questions with each other, although this danger could probably be minimized by asking the students to treat the test questions as confidential.

It can be seen, from the above discussion, that it is somewhat difficult to arrive at a test design, for this situation, that has no possibility of bias in one direction or another. A better design would probably be one involving a control group of people who have not taken the course but who are otherwise well "matched" with the student group in terms of other characteristics, especially in their educational background and level of experience. The use of a control group in this way would eliminate the need for the use of pre-course questions. The two groups would simply answer the same set of questions when the course is completed, and the two sets of results can then be directly compared, the assumption being that the student group should get significantly better scores than the control group. This type of measurement of learning, although possible within a university environment, would be extremely difficult to implement with most courses, where potential groups of participants are much less "captive".

Although we have taken pains to point out the possibility of various biases in approaches to the measurement of learning, these biases may be less serious in practice than the discussion implies. It is quite likely that the crossover test, as described above, would give a reasonably reliable indication of how much learning has actually occurred during a particular course, especially if the questions are composed by an independent evaluator.

The use of an independent evaluator for this purpose raises problems of its own. It is obviously extremely important that this evaluator should have a very clear understanding of what the course is intended to accomplish, otherwise his questions may test for information that the lecturers themselves consider outside the scope of the course. Alternatively, his questions may test in areas that the teaching staff considers of peripheral importance, to the exclusion of more important areas, or he may use terminology that differs from the terminology adopted by the lecturers, thus causing great confusion among the students.

Let us try to summarize the situation relating to the evaluation of learning acquired:

1. Two possible approaches can be used:
   (a) the development of specific behavioral objectives for various segments of the course and the use of "demonstrations of achievement" while the course is in progress, and
   (b) the use of a before and after test, again based on the behavioral objectives of the course.
2. If a before and after test is used, the students should not be told in advance that they will be tested at the end of the course.
3. It is highly desirable that a different set of questions be used for the "before" and "after" portions of the test.
4. This can be achieved in one of two ways:
   (a) the use of two sets of questions, of approximately equal complexity, one used before and one used after the course, all students answering all questions, or
   (b) a crossover test of the type mentioned earlier.

5. The lecturers can develop their own set of questions provided they are careful to ensure that the entire set of questions ranges over the complete subject-matter to be covered in the course.
6. Alternatively, the question set can be devised by an independent investigator and the test applied by him.
7. If an independent evaluator is used he must work closely with the lecturers before the course begins and there must be clear understanding between them on terminology, on the scope of the course, and on its behavioral objectives.

Clearly, it is easier to test learning acquired in a course if its main purpose is to impart factual information or if it is designed to teach a particular skill (e.g., the production of a thesaurus). In particular, a test of learning will be easiest to apply and to grade if it can be reduced to a set of questions of a multiple choice or a true/false type. This is not always possible. In fact, a course may be designed not to impart factual information but to deal with problem-solving or decision-making situations. It is more difficult to test the impact of a course on the problem solving or decision-making skills of participants but it is not impossible. One possible technique has been described by McGuire and Babbott (9). The test was devised to assess the abilities of medical students in taking care of a sick person. Each student is given a medical case history of the patient along with several possible diagnoses or courses of action. The student must choose among the alternatives presented and each choice branches into other choices. Eventually, the patient will either recover, be referred elsewhere, or will die. A panel of medical experts is used to rate each choice at each decision point on a five-point scale, ranging from "clearly indicated" to "clearly contra-indicated". An efficiency score, a proficiency score and a composite score are calculated for each student. The efficiency score represents the proportion of the student's

responses that are beneficial to the patient and the proficiency score represents the percentage agreement between the student's choices and those of the panel of experts. The composite score combines the efficiency and proficiency scores. Although this test was designed for the health care situation it is clear that the technique could be modified to apply to other types of decision-making situations. It might therefore have applicability to various courses in the field of scientific documentation.

In conclusion, the evaluation of learning acquired is worth doing, and seems to have some relevance to the evaluation of programs in this field. The situation is, however, somewhat complicated by the fact that we are not dealing in a subject area in which standardized tests exist or, indeed, in an area in which easily measurable skills are dealt with. It is relatively easy to measure such skills as typing ability and proficiency in translating from one language to another. It is much more difficult to test for learning acquired in an area related to the transfer of scientific and technical information. But we must not ignore this aspect of evaluation simply because it presents some difficulties.

Evaluation of Behavioral Change

If evaluation of learning is difficult, evaluation of behavioral change in students is even more of a problem. In this phase of evaluation we are concerned with the long-term effects of an educational program. This goes beyond learning as such into the application of the learning acquired from a particular course. The obvious consideration in this evaluation is to find out how the students have benefited, in the long run, from some program. Benefits would most likely relate to the professional advancement of the student: promotion, increased responsibility, successful movement into a different type of position (e.g., into teaching or research) and so on. Hampton (1) suggests that one way of accomplishing the evaluation of behavioral change is

by the "systematic appraisal of on-the-job performance before and after the educational experience" and that such an appraisal should be made "by the person receiving the education, his superiors, his peers, and his subordinates".

It seems quite likely that organizers of courses in this field will be interested in this level of evaluation for some at least of their educational offerings. The UNESCO course conducted at Katowice in 1974, for example, had as a short-term objective the training of teachers, and prospective teachers, in the subject area of information retrieval. The longer-term goal was the further dissemination of the knowledge acquired in the course. This goal would be achieved if the course participants were able to use some or all of the course material in their own teaching, or in the planning of curricula, when they returned to their own countries. It is clear that the long-term success of this course could only be measured by some type of follow-up study, conducted perhaps six months to a year after the end of the course. A follow-up questionnaire can be used in this situation to determine whether or not the course participants have been able to make use of the material presented in their own teaching and, if they have used it, with what degree of success. At the same time, the participants can be asked to look back on the course and to assess its value within the longer-term perspective.

With the kinds of individuals and situations with which we are dealing in this particular subject field, it may not be appropriate to attempt formal on-the-job performance assessments, using peers or superiors of those participating in the programs. Nevertheless, there may be situations in which it will be possible and appropriate to conduct some form of follow-up to gain a further and different view of the results of a program. For example, some agency of a national government may nominate or sponsor one or more students to participate

in a particular course. It may be desirable and feasible to ask this agency, some months after the completion of the course, to provide feedback on the progress and performance of these individuals. In this way, we can get some idea of how satisfied the various national governments are with the programs we are developing and offering in this important subject field.

For most courses, it will be worthwhile to at least conduct a follow-up with the students. This follow-up, by means of a questionnaire, will be conducted to obtain each student's longer-term view on the value of the course, its impact on his own professional development, and the extent to which he has been able to apply the knowledge acquired.

Evaluation of Program Results

The evaluation of program results is concerned with the assessment of a complete educational program, which may involve a whole series of courses or other educational endeavors. The organization responsible for planning and presenting the program should be interested in discovering how successful it has been as a complete entity. This is a level of evaluation that is of broader scope than the type of evaluation considered so far. As an example, we can evaluate as distinct units the several courses that comprise a complete curriculum in an academic department (e.g., a school of library or information science). It is conceivable that each course could be quite successful but that the entire program fails to achieve its objectives, possibly because it is incomplete or because it emphasizes the wrong things. The program may be successful in imparting basic technical skills and thus equipping the students to perform adequately in beginning professional positions. However, it may fail to provide the professional philosophy or theoretical underpinnings that will allow the student to "grow" in his profession and thus be capable of

accepting increased managerial responsibility. If the objective of the program is merely to instruct in basic skills it can be considered successful. If, on the other hand, the program has the long-range objective of equipping the student to develop professionally, to adapt to change, and to continue his learning, it may have failed completely. It is quite clear from this that we must have program objectives as well as course objectives, and that we must develop criteria and procedures whereby the results of the program can be evaluated against the program objectives. Evaluation of the program is the responsibility of the program planner rather than the individual teachers although these teachers, along with the students, may clearly have an important role to play in the assessment of the program.

A series of general objectives should be developed for any program in the area of education and training. For any specific educational activity within this program it will be important that precise objectives, relating to the overall program objectives, be developed. These objectives must clearly deal with who is to be trained, what information is to be imparted, and what the end results of the programs are intended to be. The educational activities themselves must obviously be planned with these objectives clearly in mind. In relation to the stated objectives a number of important decisions must be made, decisions concerning how the objectives can be met as efficiently and economically as possible. Questions to be considered will include the following:

(1) What type of educational offering (short course, long course, workshop, seminars, or various combinations of these) is likely to be most appropriate to satisfy the program requirements?

(2) How many of these are needed to satisfy the program objectives and how frequently must they be presented?

(3) Where should the training programs be held in order to have the maximum possible impact at the least cost?

(4) What individuals, types of individuals, or groups are best qualified to implement a particular program?

(5) What type of student should be admitted, how many students should be admitted, and how should the students be selected?

(6) What language or languages should the program be held in?

The above list is intended only as a sample of the types of questions that must be considered in the planning of training activities. The important thing to recognize here is that evaluation must be integrated into these activities from the very beginning and that it must occur at various stages in the complete process. The following sequence is recommended:

(1) Establish objectives for a particular program;

(2) Evaluate the objectives. Are they really the objectives we wish to attain? Are they reasonable and viable? Modify the objectives if they fail to stand up to this examination;

(3) Develop plans for a training program that is likely to satisfy these objectives as efficiently as possible, addressing all of the various questions mentioned earlier;

(4) Develop criteria by which the program can be evaluated in terms of the achievement of its objectives;

(5) Develop procedures for program evaluation;

(6) Have each element in the program evaluated. By "element" we mean individual workshop, seminar, course, or whatever other means is used to give the education or training needed. Evaluation of a program element will include reaction evaluation (formative and summative), evaluation of learning acquired, and evaluation of behavioral changes in the students;

(7) Analyze and interpret the results of these evaluation activities. This should be a continuing and current activity. Programs may be slightly

modified, expanded, drastically changed or abandoned completely on the basis of results gathered through these evaluation activities. It is, of course, important, that the evaluation data be considered in relation to the criteria developed earlier; i.e., the criteria we identified as significant in measuring the degree to which the program objectives have been met;

(8) on the basis of all the evaluation data available, consider whether or not the objectives have been satisfied, conducting whatever further surveys (e.g., follow-up studies with students, employers, sponsoring agencies) may be needed to complete the overall evaluation of program results;

(9) identify weaknesses or failures and the causes of these. Use the knowledge thus gained in the planning of future, improved programs.

Conclusion

One important matter, mentioned earlier, is that of the adequacy of the evaluation information that is collected. Knox (3) identifies five facets of "adequacy" in this context: (1) Validity, (2) Reliability, (3) Universality, (4) Application of results, and (5) Efficiency.

In terms of validity the important consideration is to ensure that the data collected do in fact tell us what we want to know. This implies that the evaluation, as well as the educational offering itself, must have defined objectives. That is, we must first carefully specify what it is we want to learn about the course and must then design evaluation procedures that will gather the data we need to draw our conclusions. We must be certain that our evaluation procedures, collectively, gather all the data we need and, equally important, that we are not collecting superfluous data for which we have no particular use. Also, we must be sure that the questions we ask are questions that the respondents are competent to answer. If we are looking for completely frank responses, we may be more likely to get these if questionnaires are completed anonymously. Likewise, the anonymous approach may be preferable in any tests of learning acquired by students. A student may feel less threatened if he is not required to put his name on a test of this kind.

Reliability relates to the reproducibility of evaluation results. A questionnaire, or other evaluation instrument, may be considered reliable if it gives approximately the same results however many times it is applied in an identical situation. We could test the reliability of an evaluation questionnaire by applying it twice under exactly the same conditions (i.e., using the same group of students at exactly the same point in a course). This could be done, for example, by having the questionnaire completed before and after lunch on a particular day, but this does not seem a very practical or desirable approach. Usually we must be satisfied with less. At the very least, we must do everything

we can to ensure that the questions we ask are quite clear and that they are not susceptible to various interpretations. Closed-ended questions (i.e., those requiring a selection from a list of alternatives) are likely to get more consistent responses than open-ended questions. Standard scales are likely to produce quite reliable responses because these have been tested and proven with many previous audiences. It is desirable that a questionnaire be pretested before a course by asking various people, having roughly the same background as the actual participants, for their interpretation of the meaning of the questions. A question that is always interpreted in the same way, by members of this pretest group, may be regarded as unambiguous and reliable. A question that is variously interpreted will require modification for purposes of clarification. As more courses are held by a particular organization, and more evaluations are conducted, it will be possible to identify questions or types of questions that appear to produce reliable responses. Wherever possible, it seems desirable to develop fairly standardized evaluation instruments - questionnaires that can be applied, perhaps with minor modification, to several courses. This has two advantages: (a) reliability is improved by the use of questions that have been tested in the past, and (b) we can compare the impact of various courses if the evaluation methodology is held constant.

This brings us logically to the next characteristic, that of universality. Each educational offering within a particular program will have some unique features that we may want to evaluate. But the various programs will also have many things in common. If the same approach is used to the evaluation of the common elements it will be possible, as Knox has pointed out, to combine findings in order to develop improved guidelines for the planning and implementation of courses. Analysis of responses to common elements would allow the planning body to

identify characteristics of courses or participants that, in the words of Knox, "are associated with high satisfaction or high achievement and those associated with low satisfaction or achievement". It is strongly recommended that a sponsoring organization should encourage the development of standard approaches to the evaluation of courses, workshops and seminars. Each set of evaluation instruments should include both common elements and elements that may be needed to evaluate the unique features of a particular program.

The characteristic application of results obviously relates to the use that is to be made of the evaluation data. We must be quite sure that we collect data in such a way that we can apply it to answer all the questions that we would like to answer about a particular program. To take a simple example, we may want to know what effect the amount of previous professional experience has on the participants' reaction to a course or on the degree to which they seem to have benefited from it. Quite obviously, if forms are completed anonymously, we must ask respondents to indicate the number of years they have been active in the field. In this way we can separate the responses of the more experienced participants from those of less experience. This highlights once more the need to prepare evaluation specifications; that is, to define very carefully in advance what it is that we want to learn about a particular course. It seems extremely important, then, that those responsible for planning a course should prepare a set of specific questions for which they would like to obtain answers. The evaluation should then be designed to answer these questions as reliably, thoroughly and efficiently as possible.

Efficiency relates to the way data are collected. It is the responsibility of the evaluator to see that the needed data are collected in the easiest possible way from those who are in the best position to provide it. The evaluator must avoid redundancy: he must not collect data that he will not use and he must not collect the same data twice (unless he feels he need to as a check on reliability).

Questionnaires should not be used to gather data that are readily available in other ways (e.g., the application forms of participants). The data-gathering procedures should interfere as little as possible with the conduct of educational programs. We must avoid evaluation "overkill" at all costs. The evaluator must also consider the best time to collect various data as well as the best way of collecting it. It is also important to plan the data collection in such a way that reduction, analysis, and interpretation of the data are facilitated. Because the majority of documentation courses will involve relatively small numbers of participants, it is likely that the data reduction can be handled quite adequately by hand. If very large quantities of data are collected, however, serious consideration should be given to the possibility of collecting these data in such a way that they may be manipulated by standard data processing techniques.

References

1. Hampton, L.S. Evaluating continuing education programs. Adult Leadership (Mount Morris, Illinois), vol. 22, no. 3, September 1973, pp. 105-119.

2. Guidelines for the evaluation of training courses, workshops and seminars in scientific and technical information and documentation. Prepared under contract for UNESCO by F.W. Lancaster. Paris, UNESCO, August 1975. SC.75/WS/44.

3. Knox, A.B. Approaches to conference evaluation. Working paper prepared for the Conference and Institute Division Meeting, National University Extension Association, Lincoln, Nebraska, 29 April-2 May 1962. Available from the Office for the Study of Continuing Professional Education, University of Illinois, Urbana, Ill.

4. Scriven, M.S. The methodology of evaluation. In: R.W. Tyler et al (eds.). Perspectives of curriculum evaluation, pp. 39-83. Chicago, Rand McNally, 1967. 102 p., illus. (AERA Monograph Series on Curriculum Evaluation, 1).

5. Steele, S.M. Contemporary approaches to program evaluation and their implications for evaluating programs for disadvantaged adults. Syracuse, N.Y., ERIC Clearinghouse on Adult Education, 1973.

6. Knowles, M.S. The modern practice of adult education. New York, Association Press, 1970. p. 224. 384 p., illus.

7. Woodward, J.P.; Yeager, J.L. Evaluation of programs to train educational R & D personnel. Pittsburgh, University of Pittsburgh, Learning Research and Development Center, April 1972. 26 p. ED 064 376.

8. Slater, M. Short course assessment and evaluation in the library/information field. London, Aslib, 1974. Aslib occasional publications no. 15.

9. McGuire, C.H.; Babbott, D. Simulation technique in the measurement of problem solving skills. Journal of educational measurement (East Lansing, Michigan), vol. 4, no. 1, Spring 1967, pp. 1-10.