Formative Evaluation:
The TVOntario Perspective

Patricia Parsons
Anne-Marie Lemire

Abstract: Formative evaluation is undertaken at several educational media producing institutions in North America. The article describes the formative evaluation process at TVOntario, the provincial educational television network of Ontario, and provides examples to illustrate the points outlined. Its main objective is to describe the many ways of interacting with audiences.

First presented are the objectives of formative evaluation and the kinds of information that can be sought, such as the level of appreciation and the intellectual impact of a program. Next, a brief overview of methodology for formative evaluation is given, followed by an outline of the various production stages when an evaluation can be undertaken. Descriptions of recent evaluations, with findings, are briefly noted. Finally, the article discusses how results are analyzed and reported to the production team and management.

Formative evaluation can be loosely defined as "verifying a product with the target audience in its developing stages in order to provide feedback to revise and improve the product." The term "formative evaluation" originated in 1967 (Scriven, 1967) when two kinds of evaluation were distinguished: one that can take place during a product's developmental stages and another that measures the effectiveness of a product after its completion. There is, however, documentation of this activity for instructional media as early as 1921 (Lashley and Watson, 1921). At this time American World War I training films were becoming available for general viewing, and government wanted to obtain early audience reaction. The practice has now been incorporated structurally into the research and evaluation undertaken at several organizations.

This article presents an overview of the formative evaluation process at TVOntario. TVOntario is a provincial educational network that develops television programs and educational materials for the schools and general public. There are three research departments at TVOntario: Development Research, Market Research, and Project Research. Development Research is involved with long term planning, while Market Research concentrates on audience research and large scale mail surveys. Project Research is the department most closely linked to the production process and it is here that formative evaluation has been undertaken at TVOntario since 1974.

This article focuses specifically on the different stages of production when formative evaluation can be undertaken. It does not discuss in depth the use of different methodologies

Patricia Parsons and **Anne-Marie Lemire** are Research Officers for TVOntario, Toronto, Ontario. Both undertake research and development projects and service TVOntario market areas including ascertainment studies, formative evaluation, summative evaluation and special research projects involving new technologies.

and their credibility. Its main objective is to describe some of the many possible ways of interaction with members of the target audience.

THE OBJECTIVES OF FORMATIVE EVALUATION

At TVOntario, formative evaluation is not mandatory and is usually undertaken at the request of the production team. Uncertain about various aspects of the production, the production team wants to obtain feedback from the audience they are trying to reach. Those not involved in the technical production, but who have concerns - such as managers and administrators - may also request formative evaluation. Given the cost of television production, it has become more and more important to justify decisions made during the creation of a program. By discovering errors before the product is ready for broadcast, one can, in part, assure the presentation of a quality product and, in part, avoid spending large sums of money that could be better used elsewhere. At TVOntario, an educational concern is added to this financial one. As well as measuring audience reaction to the program, it is also necessary to verify to what point the educational objectives have been met. A formative evaluation should measure the level of appreciation for a product (a program, script, etc.) and its intellectual impact on the audience.

Level of Appreciation

A major concern of producers is that their program be interesting to watch. It is not, however, enough that a program be interesting. It must be interesting enough that the viewers will be motivated to watch the whole program and want to watch future programs. Different elements can influence the level of interest a program arouses.

(a) Subject: The fact that one is interested in a subject considerably influences one's level of appreciation. For formative evaluation, it is therefore necessary to identify participants' interest in a subject before and after the screening of a program. The treatment of a subject is also an important element that can change participants' interest in a subject.

(b) Style, Formats: As the same format can be effective for cultural information but ineffective to explain social phenomena, it is necessary to reexamine the effect of format for every new production. This part of the evaluation of format also includes aspects such as rhythm, setting, etc.

Intellectual Impact

The intellectual impact of a program can be measured by the amount of information retained and by the level of comprehension of the content. For a program created for classroom use, it is easier to evaluate retention of information. First, students are more familiar with being evaluated and, second, program content is more directly linked to a curriculum, making it easier to categorize. Students' learning is therefore easier to measure.

However, it is often "delicate" to evaluate the comprehension and retention of an adult audience. Adults feel insecure more easily in an evaluation situation. Indirect methods, such as relating concepts from a program to their personal experience and asking general questions about program content, can be used at these times.

In evaluating the intellectual impact of a program we generally investigate the following areas:

(a) The amount of information: Audiences often remark that the program was interesting but they were unable to remember all the information. Unlike print, television has a fixed pace and viewers are normally unable to stop and review or reflect on points as

they are presented. The amount of information in a program should sustain viewers' attention while assuring comprehension of the concepts presented. At TVOntario we have found, as those at other institutions have noted (Bates, 1984,1985), that one of the most common problems in all types of educational programming is "too much information, too fast."

(b)   The level of information: One purpose of an evaluation is to verify that the content of the program is presented in language easily understood by viewers. Concepts introduced must also be suitable for the target audience and readily comprehensible.

(c)   Depth of information: In our experience, this aspect is often sacrificed to amount of information. As one example, in an evaluation of a recent pilot TV program for a series on new scientific discoveries, viewers were very interested in the three items presented in the half hour program, but would have preferred one topic covered in more depth.

Beside the level of appreciation and intellectual impact, formative evaluation also endeavours to respond to any other questions of the production team. For example, participants could be asked what they think of a mini-series that production has in mind or what ideas they have for future programs in an existing series.

For all of the above objectives and to find out very specific concerns the research team begins the evaluation process by first meeting with members of the production team. All the necessary information, such as their questions and areas of concern, the objectives of the series and the program, and timelines, is obtained. Getting a clear definition of the program's objectives allows the researchers to select an appropriate methodology. While working on a program, the production team can often lose sight of its original objectives. This exercise is thus beneficial for both the production and the research teams. Timelines are also a factor in determining the extent of the evaluation. Unfortunately, at TVOntario the evaluation process often begins very late in the production schedule. Very little time is available to the evaluation team who must often undertake a study and report results in a period as short as three weeks.

In consultation with the production team, the evaluators must also determine who the sample audience is for testing purposes. A learning system or educational media product often has two or more target audiences: a primary audience, a secondary audience, and those of the general public interested in a topic. A learning system about the use of computers, for example, might have several target audiences: those who have a computer at home, those who are enrolled in a course, or students in secondary schools. The mix of people recruited for an evaluation session, the sample audience, can vary. Sometimes an evaluation is conducted only with a primary target audience, while for other products several types of audiences can be included.

After the sample audience for the study and the objectives of the evaluation have been defined, the evaluation team prepares a proposal for the study and presents it to the production team and to others involved (for example, the manager of adult programming, if the program being evaluated is for adults). The proposal briefly describes the context of the evaluation as well as the methodology suggested. It is reviewed by those requesting the evaluation and revised if necessary.

METHODOLOGY

At TVOntario qualitative methods are often used for formative evaluation. The holistic approach adopted allows the researchers to see interrelationships among the different elements that make up a program. By not manipulating any variables, the qualitative

method also provides an overall sense of a viewing situation in a natural setting. The reader is referred to the work of Bogdan & Taylor (1975) and Huberman (1981) to obtain arguments favoring the utilization of qualitative methods. The most frequent criticisms formulated (Sadler, 1981; Miles, 1979) cite the impossibility of generalizing to the population, the use of a tool that is only more or less structured and the contamination that can occur in group discussions or interviews. Recent research on this subject (Research Communications Ltd., 1985) seems, however, to indicate that participants do not allow themselves to be influenced by the remarks of others and that, in fact, remarks generated in discussion are congruent with results of questionnaires. Examination of our own data also leads us to believe that this halo effect only rarely occurs and usually with the type of participant who would not give his or her opinion anyway. Research from Research Communications Ltd. seems also to show that the results obtained from a small sample with characteristics representative of the target population are similar to results obtained from a larger representative sample. Regarding the criticism concerning the lack of structure in methodological design, the research team at TVOntario is currently developing an evaluation grid adaptable to evaluations for various educational materials while still maintaining some rigor in the analysis of results. This grid would allow a systematic and rapid comparison of verbal comments and written questionnaire results within and between groups. Qualitative and quantitative data are not, in our view, incompatible. While remaining aware of the possible limitations of qualitative methods, we feel it should be recognized that they permit the collection of information that allows formulation of recommendations and concrete suggestions.

The specific methodology used in each evaluation differs according to the nature of the product, its target audience, and its development stage. To simplify in the extreme, a basic model can be formulated to serve the purposes of many formative evaluations. In general, an evaluation session would be divided into four parts: a prescreening questionnaire; the screening of the program; a post-viewing questionnaire; and a group dicsussion. However, for each product evaluated, different modifications can be made to obtain the information requested.

Production State for Evaluations

The following is an outline of the different pre-production, production-in-process, and post-production stages during which research and evaluation can enter into play. The list is not exhaustive but provides an idea of what we at TVOntario have undertaken. The only elements that limit evaluations are budget restrictions, time, and imagination. Some methodological aspects are briefly described.

Concept testing. This can be done when producers or project teams know they are going to do a program or a series on a certain topic, but are uncertain how much emphasis to place on what, or want to know more about how special segments of the population think before proceeding to the scripting stage. For instance, TVOntario has recently produced a series on the future of work (Karam & Duggan, 1983). The developers of the series requested the help of formative evaluators to explore the potential audience's reactions to material that might be included in the first program and in the series.

Small discussion groups were organized with participants from varied ages, educational levels and occupations. A written questionnaire was used to stimulate discussion and to record quantitative data. The first section of the questionnaire presented the basic ideas of the series in the form of 11 questions. The second section included questions based on the ideas for program 1. After completing their questionnaire, the participants took part in a discussion where they were encouraged to exchange their opinions on the suggested topics. The participants gave their reactions to a brief description of the program's major possible

themes. By compiling and comparing the quantitative data and the verbatim comments, the research team found major difficulties with the outline proposed and made a series of recommendations.

The storyboard or graphic stage. Before camera work, it can be useful to test drawing or pictures with members of the target audience. For example, a children's series producer was developing a sequence for a preschoolers' television program using a set of graphics that would be shown on air with a nursery rhyme as a voiceover (Duggan, Parsons & Karam, 1981). When reviewing what his graphic artist had done for the sequence, he became concerned that the younger members in his audience, the 4-year-olds, might be frightened by drawings of a mouse that had elongated teeth, sharp ears, bright black eyes, and a long twisted tail. The evaluators showed the drawings to a group of 4-year-olds, in individual interviews, using a brief story as introduction. The children were then asked probing questions about their feelings about the mouse. A sizable proportion of the children tested did express some dislike for the mouse, saying he was ugly, and looked "scary". The producer instructed the graphic artist to "soften" the look of the mouse and the producer could then feel more confident that irate partents would not be complaining to the network about a program frightening their child.

The outline stage. Before completing a full script, a writer often likes some feedback before fleshing out the details of the story. He or she may be dealing with a particularly sensitive subject and wondering what approach to take, or may be unfamiliar with the target audience's thinking about the topic.

Again, in the same preschool series, the developers requested help from evaluation for a program they wanted to produce on the topic of death. The writer had completed an outline for a story about the death of a butterfly, and before writing the complete script, he wanted to know if his treatment would disturb young children. In this instance, instead of testing ideas with members of the target audience (which would be extremely delicate, and perhaps unethical, because we could frighten children), we engaged a child psychologist to consult for the program. She read the outline, and with the writer, producer, and researchers, discussed the fears of young children and the feelings that could be explored at the end of the program. As a result of this, the writer altered the script considerably, deleting certain scenes and adding an ending that included a celebration of life.

The script stage. After a complete script or several scripts have been written, the developers often feel ready to have it reviewed by an objective team. Although this stage has the advantage of a complete story to present to an audience or reviewers, the disadvantage is that, if problems turn up, the writer is faced with a considerable amount of work in revamping the story or theme. This misfortune did in fact occur during the development of a series prepared at TVOntario for a teenage audience (Bailey, 1983). A manager, in reading several completed scripts, had certain reservations and requested an evaluation before further scripts were written or production began. We realized that a teenage audience may not be interested in reviewing scripts and answering a questionnaire. Instead, we chose to rewrite the scripts in play format, so they could be more easily read aloud in highschool classes, with several students taking part. In this way, an entire class heard the script and was able to become involved with the characterizations. The students then gave their views about the characters and how realistic they felt the situations to be. Since teenagers are often the most difficult audience to target for, the data collected were extremely informative. In this instance, the teenagers were critical of aspects of the plot as well as the characterizations.

After actual production takes place, there are several points at which formative evaluation can be undertaken.

The rough-cut stage. When shooting has been completed and the production team is ready to assemble the programs, the producer may be uncertain about elements in the

programs and want to obtain reaction before assembling a final pilot program. One program is put together, in a format that will be used for the entire series. The audiotrack may not be completely done and some sequences may not be finished as the producers envision them for the final product. At this stage, formative evaluation serves to give feedback on various features of one test program, so that all the programs in the series can be planned better. We conducted a "rought cut" evaluation for an adult learning series on the Middle East (Karam & Winsor, 1984). A "typical" program was assembled and shown to audiences. Because of certain segments, the audience showed some confusion as to whether the series was exploring the current situation or recounting the history of the area. Some visuals, such as maps, were also felt to be unclear.

The pilot program stage. This is the stage at which producers and management at TVOntario traditionally request formative evaluation. A pilot program can be a trial or sample program made specifically to test audience and management opinions and this is true for most of the pilot programs produced for the large North American private networks. In educational broadcasting, where funds are more limited, a pilot program is usually the first program in a series, and may be kept for broadcast if acceptable to the target audience, or may be changed in part, after formative evaluation. At TVOntario the most difficult barrier to successful formative evaluation at this point is the production deadline for the rest of the series. Evaluators often only have a brief period to test the program and report results back to the production team before the scripting and production planning are too advanced to incorporate changes recommended from the testing.

Pilot testing can often alert production people to inconsistencies, errors, or faulty assumptions that develop during the rush of putting a program or package together. At TVOntario we have often found many elements forgotten, neglected, or unchecked. For instance, in a preschool pilot we noted that print appeared on the screen without the characters saying the words aloud. The writers had forgotten that preschoolers — the target audience — can't read (Duggan, Parsons & Karam, 1981). For the same series, we discovered that quizzes, with direct questioning of the audience, elicited greater participation and attention than elaborately produced studio sequences.

The test instruments used for a pilot evaluation vary according to the audience and the participants recruited. For example, with a teenage audience, instead of having group interviews led by one person, we formed task groups during which participants responded to a series of questions. The ultimate objective of this task group was to revise the program from the elements already used and make a new version. In each group one member had the responsibility to see what participants said for each question. In the analysis of results one must be conscious of the fact that the participants are not experts in television and that their suggestions do not take into account factors which are not known to them such as budget, timelines, and technical concerns.

After three or four programs. After the rough cut or pilot stage, when the developing team has gone on to produce several programs, a producer may request formative evaluation to find out if an audience reacts more favorably to certain elements in the program after changes were made according to a pilot evaluation. For a family-oriented science program, the production team wanted to verify if they were on the right track in a later program in the series, after a pilot evaluation had indicated that the information in the program was presented in a confusing manner (Windsor, 1983).

After one season. If a series continues for more than one season, management or the production team may be contemplating some changes and wish to obtain reaction to them before committing themselves to production with these new variables. A new host may be contemplated, or less location shooting and more studio sequences proposed. We recently investigated audience reaction to a movie show that had been on air for many seasons. The

production and management team requested formative evaluation to help them decide between the continuation of location shooting for interviews and straight studio "living room" interviews. An interesting finding was that although the viewers appreciated the visual variety of location shooting, they found that the surroundings disrupted the flow of information presented in the interviews. They learned less from location interviews than from studio interviews.

ANALYSIS OF RESULTS AND REPORT PRESENTATION

When evaluation sessions are completed, the questionnaire results are compiled statistically. The calculations done at this first stage are generally frequencies and crosstabulations with characteristics such as lifestyle and general appreciation, level of education and comprehension. Given the small number of participants, more elaborate statistical calculations often would not give any additional pertinent information at this stage.

Tape recordings of group interviews are also transcribed. A systematic analysis of their content provides complementary information to the quantitative data. Supplementary information about the elements of the program that were not included in the questionnaire is also obtained.

Preliminary report. Following this first analysis, a preliminary report is presented to the production team from one to two weeks after the evaluation sessions. At this meeting statistical data and a content analysis of the discussion are briefly presented. Some recommendations and interpretations are then formulated. The meeting essentially serves to present information for the questions that particularly interest the members of the production team. If other results seem interesting, they are also outlined. The emphasis is placed on the oral presentation and the written report usually constitutes several pages. Diagrams, graphs, and other visual material are used to make the presentation of results easier. Often, transcripts of discussion group tapes or actual discussion tapes that seem particularly representative are left with the producer. Typed lists of responses to open-ended questions can also provide the rich detail that producers enjoy. If the request for evaluation has come from management, results are usually first presented to the producer or production team and, then, at another meeting - with the producer present - presented to management.

Draft report. After the preliminary reporting, a draft report is written that includes as detailed an analysis of the results as the researcher feels necessary. Demographic information may be looked at more closely or specific questions addressed. At this point, more interpretation is included and verbatim comments of test participants are added if these can help provide insight and more interesting reading.

Final report. The draft report is revised after feedback or criticism from readers. Often, after disucssion with producers or developers, more recommendations for the project are included. A problem that frequently arises at this point is report distribution. Most developers and producers do not relish having even the slightest weakness in their project discussed in a report that is circulated to upper management. At TVOntario, an abstract of the report is circulated to upper management, and often does not include the details of sessions. The full report is usually available to upper management upon request.

CONCLUSION

What happens after the results of formative evaluation have been analysed and presented

depends on the producer, the evaluators, the budget, the program, the series, and the time available. If, for example, after a pilot evaluation, results show that one character has received a very negative reaction, and parts of the program are found to be confusing, a decision to reshoot that character's segments and reedit certain sequences may be made. On the other hand, the pilot might be left as is and the evaluation recommendations taken into consideration in the planning of the rest of the series. Not only can the production gain from an evaluation but the methodology used can also be refined. Through the years the process of formative evaluation at TVOntario has undergone changes and developments. After more than 11 years of experience in this area, we at TVOntario are convinced of the effectiveness and usefulness of formative evaluation. We know that formative evaluation does not guarantee the improvement of educational media products, but feel it is a tool that is available for those willing and able to use it; a tool to help make a product that answers more precisely the needs of the audience.

REFERENCES

*Bailey, I,. (1983). *Formative evaluation of career planning* (Report No. 9 -1983). Toronto: TVOntario.

Bates, A. (1984). *Broadcasting in education: an evaluation.* London: Constable.

Bates, A. (1985). *Conditions for successful learning from broadcast.* Speech delivered at TVOntario, Toronto.

Bogdan, R., & Taylor, S. J. (1975). *Introduction to qualitative research methods.* New York: John Wiley.

*Duggan, K., Parsons, P., & Karam, B. (1981). *Today's special formative evaluation of two pilot programs* (Report No. 5-1981).   Toronto: TVOntario.

Huberman, A. M. (1981). Splendeurs, miseres et promesses de la recherche qualitative. *Education et Recherche, 3* (3), 233-249.

*Karam, B., & Duggan, K. (1983). *The Future of Work: Formative Evaluation* (Report No. 11-1983). Toronto: TVOntario.

*Karam, B., & Winsor, S. (1984). *The middle east: formative evaluation* (Report No. 27-1984). Toronto: TVOntario.

Lashley, K. S., & Watson, J. B. (1921). A psychological study of motion pictures in relation to venereal disease campaigns. *Social Hygiene, 7,* 181-219.

Miles, M. B. (1979). Qualitative data as an attractive nuisance: the problem of analysis. *Administrative Science Quarterly, 24,* 540-601.

Research Communications Ltd. (1985). *Formative research: believe it or not?* Paper presented at the Annual Conference of the Association for Educational Communications and Technology.

Sadler, D. R. (1981). Intuitive data processing as a potential source of bias in naturalistic evaluations. *Educational Evaluation and Policy Making, 3,* 25-31.

Scriven, M. (1967). The methodology of evaluation. *American Educational Research Association Monograph Series on Curriculum Evaluation, I.* Chicago: Rand McNally.

*Winsor, S. (1983). *Not another science show: a formative evaluation of the program "Up, Up, and Away"* (Report No. 22-1983). Toronto: TVOntario.

*Note: References preceded by an asterisk are internal documents of TVOntario which not available for circulation.