DOCUMENT RESUME

ED 124 166 IR 003 556

AUTHOR Parer, Michael S.
TITLE Formative Evaluation and Instructional Television; the Need, Problems, and Instruments.
PUB DATE Mar 76
NOTE 9p.; Paper presented at the Association for Educational Communications and Technology Annual Conference (Anaheim, California, March 28-April 2, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Educational Television; *Evaluation Methods; Evaluation Needs; *Formative Evaluation; Speeches
IDENTIFIERS AECT 76

ABSTRACT

Though instructional development models stress the importance of formative evaluation, techniques for the formal evaluation of video tape programs remain primitive. The attention, attitude, and adoption levels achieved can be assessed by use of the following evaluation instruments: (1) expert appraisal of script; (2) special questionnaires; (3) story board; (4) content analysis; (5) stop frame; (6) observation--personal or video; (7) eye movement; (8) demand for audio-video; (9) real time program analyzer; (10) eye contact; (11) distractor; and (12) multiple screen. This monograph provides a brief overview of each of these instruments and discusses their use in formative evaluation. (EMH)

# FORMATIVE EVALUATION AND INSTRUCTIONAL TELEVISION
## THE NEED, PROBLEMS, AND INSTRUMENTS

AECT, International Division,
Anaheim, California, USA
1976





by
Michael S. Parer

Project Co-Director, Literacy Instructor Training Via TV (LIT-TV)

Senior Video Coordinator and Instructional Developer
Center for Innovation in Teaching the Handicapped
Indiana University, Bloomington, Indiana, USA

Former Director of the Migrant Education Television Project for The Australian Government and Producer of "You Say The Word", an instructional television program for migrants.

In 1970 the Australian Federal Government, through its Department of Immigration, initiated an instructional television (ITV) project for migrant education and integration. Over a period of 25 years the Department had developed the Situational Method to teach migrants of fifty different nationalities. During a two year period, 80 one hour television programs of "You Say The Word" were developed, produced, and transmitted on the commercial WIN-TV network. The budget for the first year was $123,000, rising to $188,000 in 1973, and $800,000 in 1976. A major increase in budget was sought for the second year to allow for experimentation and evaluation, but this was declined. Programs were transmitted at 9:00 a.m. on Saturday, 11:00 a.m. Wednesday, plus a 15-minute segment each morning at 9:00 a.m. Currently they are on 12 commercial networks. The series is aimed to teach migrants English, to tell them about their new homeland, and to give Australians an appreciation of the new settlers. There have been surveys of the size and composition of the audience that have produced favorable results. The viewing audience, according to the first surveys, included 11,000 migrant women who had been untouched for 25 years by any other of the Department's education programs. Further, 19.6% of the total target audience watched the program. This is phenomenal success in education and television terms. Currently the program scores in the commercial ratings. However, evaluation has yet to be done on the content, talk back technique, production pace, style, and the English as a Second Language progression. Feedback is the basis for an improving ITV. The enormous initial effort to master the techniques and technology of the new medium usually leaves little time for systematic formative evaluation. This lack of feedback is tragic. Constant evaluation plus continual pretesting of ongoing experimental productions are the very life-blood of successful TV.

It was to seek some definitive answers to ITV evaluation that I came to The Division of Instructional Systems Technology at Indiana University. There, within the Center for Innovation in Teaching the Handicapped working as video coordinator and currently as project director, I have had a chance to explore the possibilities and recognize the limitations of formative evaluation in ITV.

The past two decades have witnessed dramatic progress in video technology. Today, small units may be equipped with TV production facilities for $50,000. Only a decade ago the cost was $250,000. This same period has seen a vast development in instructional development theory as it relates to the production of mediated instructional materials. As a consequence the potential of TV for instruction is now within the reach of vast new areas from business, to industry, to insurance, to special education, there is, in fact, almost no limit. Yet this magnifies a problem.

All instructional development models stress formative evaluation as essential, yet the instruments of formative evaluation in instructional TV remain as crude as a 1914 byplane in an age where sophisticated technology transmits live TV pictures from Venus. This lack of efficient formative evaluation instruments to implement the instructional development theory presents a major hurdle in ITV. Some institutions, such as the Childrens' Television Workshop, the Agency for Instructional Television, and the Center for Innovation in Teaching the Handicapped (CITH), have refined particular instruments for their own use, but the direct applicability to other TV projects generally poses difficulties in adaptation to new content areas, different target populations, and new program formats.

At CITH the instructional development process, using the 4-D model (Thiagarajan, Semmel, & Semmel, 1974), is being applied to the design, development, and production of two federally funded projects. The first, "Choose a Curriculum Package," is a video module within the "Tips for Teachers" series (1975) of preservice and inservice training for special education teachers. The second is a series of video workshops in the Literacy Instructor Training Via TV (LIT-TV) Project (1976) to train adult basic education teachers and paraprofessional tutors in basic literacy methods, who in turn will teach adults who function at the lowest literacy level to read.

Since these video programs are products of the instructional development process they are assumed to have been through the formative evaluation cycle on numerous occasions to guarantee a high level of effectiveness. However, as noted above, the formative evaluation instruments cannot yet guarantee any certainty that they are returning the critical information necessary to make decisions to reproduce elements of the program. While the academic and theoretical writings of such people as Komoski, Scriven, Weiss, and Wholey are necessary and valuable, there remains a vast gap between their theorizing and the actual application to TV programs. The psychological process in education is complex and difficult to categorize even within a single teacher-pupil interaction. Research related to the interaction of a learner and mediated package is equally difficult, but television has certain demands and limitations that makes its evaluation even more unusual and complex. For example:

A. In television, each frame contains an almost infinite number of messages with movement, audio, music, facial expressions, laugh track, color tones, and so on. These change to new combinations with such rapidity that the variables are nearly impossible to identify and control. Consequently, the external validity of research findings have limited value, and efforts to develop a science of instructional TV have met with only limited success.

B. Adults and children are exposed to the commercial TV networks' sophisticated production, visual pace, entertainment expectations, and ability to stir deep emotions. These overtones can lead to harsh evaluatory judgments and comparisons with ITV programs. Many factors such as the seeking to cause cognitive growth or attitudinal change and a relatively small production budget, all too often make ITV progams appear to fall far short of commercial television productions. These features require some form of multivariate instrument, which has led to the design of specialized evaluation methods.

C. ITV is not produced for entertainment alone and should not be evaluated for the mere retention of program content. ITV evaluation must also account for:

1. Attention--Is the viewer prepared to sit and watch the program?
2. Attitude--Is the viewer stimulated to positively accept the program suggestions and favorably modify his/her attitude?
3. Adoption--Is the viewer who watches an instructional television program left with the resolve that would predict an expected adoption of the program's recommendations?

This presentation outlines the various formative evaluation instruments that were considered for use with these two CITH projects. These evaluation instruments are:

1. Expert appraisal of script
2. Special questionnaires
3. Story board
4. Content analysis
5. Stop frame
6. Observation - personal or video
7. Eye movement
8. Demand for audio-video
9. Real time program analyzer
10. Eye contact
11. Distractor
12. Multiple screen

1. Expert Appraisal of Script

The assessment of a script's program content by the media experts is normal within the instructional development process. However, television experience has shown that such appraisal should only be done once, as experts tend to become hypercritical and tend to polish, add, and make reservations. Such experts ideally should have some professional familiarity with the television media. Niceties of refinement may be well accepted in a print or even an audio format, but television has limitations in the volume of narrative that can be used. When the narration continues to exceed 80 words per minute, the program tends to become a traditional ITV talk show. Given audience sophistication, this style of production requires a high degree of motivation to retain the attention and interest of the viewer.

2. Special Questionnaires

Questionnaires are a rather highly sophisticated form of communication. Pre- and posttests are normal in educational experiments and are widely used in television with a combination of (a) the Osgood semantic differential, (b) the Likert attitudinal scale, (c) Scriven's goal free evaluation, (d) a selection from psychological tests, and (e) demographic data. Questionnaires can be especially designed for lower socioeconomic status (SES) groups and young children. For example, smiling faces and frowning faces can be used as

a two-point scale. With simple questions this nonverbal style of questionnaire is useful. However, for these lower SES groups, behavioral measures are far better than those meeting a verbal or visual literary form.

3. Story Board

The assumption behind story board evaluation is that the mind can make a television story out of a series of still pictures or drawings by putting motion in between. With sophisticated audiences a considerable degree of insight can be obtained at this formative evaluation level with this technique. The story board instrument is set-up by taking the major action scenes from the script and writing narrative beneath each picture. Thus, the story board appears like a comic strip. Television production is so expensive that it is worth this minimal effort to ask certain kinds of questions by running a story board test. This is one of the least expensive forms of formative research. A greater degree of simulation to television can come by a professional producing the audio track and video taping the stills so that the viewer listens to the audio and sees the pictures on the television screen. This can then be tested on a sample of the target population using a TV monitor. Using this technique the visuals in the story board can be experimentally manipulated and changed to several versions for a low dollar expenditure.

4. Content Analysis

This method is a straightforward content analysis of the audio and the video track to see whether each element is compatible and supports the other; to see if the visuals are illustrating what the audio is saying, or if one segment of the program is inconsistent with another. Within instructional television, visuals often do little to enhance the message. Radio or audio cassette could accomplish the instructional task equally well or better than poorly visualized and inconsistently scripted television programs.

5. Stop Frame

This technique is useful to check the comprehension of the viewer. The television show can be played to a certain point and then stopped so that the image is frozen on the monitor as on a blackboard. The subject can then be asked to answer the question and, if necessary, physically go to the screen and point it out. This research can avoid the post-viewing analysis of comprehension and give indications in actual real time of curriculum comprehension. For example, if three minutes into a program there is a segment dealing with a particular concept, then thirteen minutes later on a further segment concerned with the same concept, the program could be stopped at the former and the latter to produce a pretest-posttest within the single program. This avoids having to show the entire program and then rely on memory to discriminate a pre- and posttest analysis. This method can reveal precisely when actual learning takes place during the program and, in turn, give the producer and script writer far greater control in programming.

6. Observation

There is a lot of common sense observation to be used in formative evaluation of television. The actual watching of expressions on faces, movement, interaction with others viewing the same program, is a simple but insightful

form of formative research. An example of such observation, of watching children in a small group setting to see to what extent they mimicked or modeled, arose in Sesame Street. The black actor, James Earl Jones, recited the A B C's with a ferocious look. Each letter was dramatically pronounced under simulated emotional stress. After a brief exposure period, a pattern was established. First, the children modeled the actor and repeated the alphabet after him; second, the children recited the alphabet simultaneously with the actor; and third, the viewers anticipated him and used the television visual as a cue to jump ahead of the actor with the alphabet letters. This unintended effect of the producer was a good example of feedback from this most simple of all formative evaluation techniques. A further refinement of this method is to set-up an instrument to assess, over an extended period, the degree that a group of people pay attention to the TV screen. This may be done by setting a mirror behind the viewers that reflects the TV screen. A 16mm camera can be placed in an upper corner to take still pictures each two seconds, and thus capture a record of who is viewing the screen. The film is then analyzed through the special athletic-projector with a slow and stop frame. This method gives a good indication of the attention held by the TV program as to how much the unobtrusive measure should be reinforced so that viewers are made aware that they are being observed.

7. Eye Movement

Traditional usage in captioned and foreign films has the print normally placed in the lower center part of the screen. However, an instrument with a beam pinpoint of light into the eye of the viewer is reflected onto the screen. This beam plus screen is photographed, and when played back, the beam of light indicates exactly where the eye is focused. Results have revealed that the eye is not always on the print, but more often on the face, and that the center lower portion of the screen is not a good place to attract attention. Further studies regarding the placement of print and the eye movement among poor, medium, and good readers led to the following conclusions: (a) words appearing out of the mouth using a profile are the most effective, (b) poor readers struggle with words and fail to read them and thus, can give the producer a guide as to how long words should be left on the screen, (c) there is a need to induce left to right reading as illiterate students viewed haphazardly and do not naturally follow the English convention of reading left to right, and (d) there is a certain length of time visual material should be left on the screen for optimum viewing.

8. Demand for Audio-Video Stimulus Response

This instrument grew out of the school of behavioral psychology where action that is rewarded tends to be repeated. The logic behind the experiments were the the tests with pigeons trained to press bars in return for reward pellets. This is applied to people in television research. The audio track is permanent, but to maintain the visual image the viewer had to continually press a button. Thus, if the reward for the effort was not satisfactory, the tendency was to not press the button and leave the screen blank. On the other hand, if the viewer was interested or found the visual information rewarding enough to make the response worthwhile, the button was continually pressed. The device can make both audio and video available only with effort.

The behavioral interests of this instrument can be developed by varying the amounts of effort or pressure required to bring on the audio and video. For example, at one extreme the viewer could be on a high geared bicycle and have to peddle vigorously to obtain the picture or sound track. This measure by-passes questionnaires and self-reports.

9. Program Analyzer

This is an instrument to obtain feedback and viewer judgments in real time while the program is in motion. A light and bell are placed on top of the TV monitor. The viewers are each given a sheet of paper with consecutive squares and asked to vote positive or negative with a cross or minus. At pre-determined intervals the light is switched on for a period of, say five seconds, and the viewer is asked to make a simple judgment to a constant question. For example, Do you find this enjoyable? Is this of professional interest to you? Is this new information?

Experience has shown a remarkable case of constancy across varying samples, and also the increasing data base gives more and more revealing information and the basis for comparison. CBS now is highly sophisticated in interpreting profiles because they have many years experience and data. It is important to stress what particular information this program analyzer reveals. The question asked is, "Do you like what is going on at this moment?" It's important not to overgeneraiize. For example, the viewer is not being asked, "Would you give up this program to watch "Upstairs, Downstairs"? This is critical because the most significant factor in the success or failure of television programs is the competing show. When the information is collected, graphs are plotted for the program and the dips and peaks give indications of programs' strengths and weaknesses which require reproduction or additional adjunct material.

10. Eye Contact

This is simply an extension of an observation technique for viewing within classroom situations. A team of observers views students, with each observer responsible for four students. At predetermined intervals the observer plots the eye contact of the student to the screen. This may be done directly or by video taping the audience and analyzing it later. Such work has been done with infrared film to study the reactions within the theater. By recording the audio track, a record can be kept of the precise position within the program to which the audience is responding.

11. Distractor

The distractor technique of formative evaluation was pioneered by Edward Palmer who was in developmental psychology at the University of Oregon. Publicity on his research came to the attention of Joan Cooney at the Childrens' Television Workshop. Palmer was consulted about the applicability of this technique through the new innovative series that was later to be known as Sesame Street. Palmer was hired, and this distractor was the major formative evaluation technique of the series. It is now adapted by NBC for pretesting their children's programs. A child--one at a time--is placed in front of a TV monitor on which the test material is played by a video tape recorded. At an angle of 45 degrees a portable rear screen projector is placed at a slightly greater distance than the TV monitor. A circular slide bank of beautiful

scenery shots is automatically advanced each 7½ seconds with the corresponding click as the slides change. An observer is placed in the room fully visable to the child to observe the time interval that the child is distracted from the program by the slide. The child's distraction is scored on the following scale:

0 = the child not looking at the distracting slide at all
1 = watches the distracting scene less than half the time
2 = the child watches it more than half time
3 = the child views the distracting screen all the time.

The child has no inhibitions or qualms about being distracted, so this technique provides a good measure of how well the TV material holds attention. The eye is drawn to the visual change which is clearly within the child's visual field. It is impossible not to see the changing visual stimulus. But the question asked is whether the program is so all-absorbing that the viewer will override the stimulus and stay with the TV program. Experiments in perception have shown that the brain can block out diversionary stimuli. One study showed that when the brain of a cat was monitored to the ringing of a bell, the nerve impulse peaked as the bell sounded. However, when a mouse was let loose in the room and the bell was kept ringing, the nerve stimulus continued to be impulsed and continued to register, but it was not diverted by the cat from the bell to the mouse. The bell sounded and created the impulse but was literally not heard. The paralled logic is that, if the child is engrossed in the TV screen, he will perceptually block out the diversionary distractor screen. Testing is undertaken for approximately 15 minutes of a TV program. This material is tested on 10, 20, 30, or 40 children. This is expensive data collecting, and samples must be kept simple and small. The data collected gives a numerical score each 7½ seconds which is then plotted on a graph. The attention span is plotted against the time. Suppose there were twenty children. With a maximum score of three for each at any given interval, the maximum total would be sixty, which is converted to 100 percent. This is then plotted on the attention graph. After all intervals have been plotted, the graph is then back-timed and the program elements noted. Thus, in the final form data is produced so that it is analytically useful, showing where the attention rose and fell and telling whether some segments are not working and should be entirely revised. If there is a segment of high impact that is necessary to communicate but has low attention, then this implies it should be sandwiched in between two items of high appeal, making this instrument a diagnostically useful device for guiding editing changes. This level of appeal of any set of slides is unknown, but this problem is overcome by using the same set of slides for every set of tests. Thus, if a scale of distraction is from 40 to 90, this becomes the constant. Material is then tested against the relative appeal, and each test material is measured against the constant. The rationale is that in the home there are constant distractions and the appeal of the program must overcome these distractions. It is methodologically impossible to replicate the real world distractions--such as a freshly baked apple pie or a puppy licking a child's hand. Rather than replicate the wide range and informality of real world distractions, this instrument stimulates a standardized distraction. Since all tests are run against this standardized distraction, it is immaterial that the initial distractor is of an unknown degree of distractability. It remains constant, and everything is measured against it.

This instrument is ideal for three to five year-old children who have no necessary commitment to follow long plots and who need materials to hold their attention. The implication is that the rate of production, editing, music, and so on must be geared toward holding that attention.

This instrument has no verbal instruction, and no verbal facility is required on the part of the child. It is a behavioral measure for very tough research subjects, namely, three to five year-old children.

12. Multiple Screens

With this instrument suggested by Dr. Keith Mielke, the viewer is placed before four television monitors. Four programs are screened simultaneously. At least two are programs of known ratings. The viewer has an audio control box with which he may select the audio track of one program. At regular intervals, say 30 or 60 seconds, the television programs are randomly switched. If the viewer wishes to continue watching a particular program, he must readjust the audio selector. This is a technique where it seems theoretically possible to predict ratings by placing test programs in the context of programs with known ratings. The technical difficulty of this may be overcome by randomly editing each program in advance and at a predetermined duration so that the programs are contained on the master tape for each particular monitor.

Conclusion

This brief overview of a smorgasbord of formative evaluation instruments does not pretend to be exhaustive. In fact, this field requires professionals who are as creative as the TV script writers and TV producers in dreaming up new instruments to actually capture the objective response of the TV viewer.

Given the strong move today towards accountability within education and government institutions, it may be hoped that the evaluation budget will be increased. This would allow for the hiring of top personnel to turn their creativity and energy to refining the above instruments and discovering new ones. Perhaps one of the main reasons that ITV has been so sluggish in education is that the interaction of viewer and TV is so difficult to categorize; while in business industry, ITV has boomed because its value and cost-effectiveness is clearly visable (i.e., if via a TV program errors in the accounting department decrease by thirty-five percent). The need for more exact formative evaluation instruments is one of the major challenges facing the areas of Instructional Development and Instructional TV.

© Michael S. Parer
CITH
2508 East Tenth St.
Bloomington, IN 47401
(812) 337-5847