Evaluation of the development, dissemination and effectiveness of mass media health programming

Brian R. Flay

Abstract

Many mass media health promotion programs seem to fail because of inadequate attention to several levels of evaluation research, possibly during product development. A rational sequence of evaluation research is suggested and described. Pre-production research includes planning research, concept testing, and pretesting of rough forms of the product. Audience testing research is suggested post-production but prior to dissemination—it includes small-scale assessments of message/product acceptability and efficacy. Post-dissemination research includes implementation, effectiveness, and process evaluation. It is argued that greater emphasis be placed on formative (pre-production) research and audience testing (particularly efficacy testing) than is usually the case.

Introduction

It seems to be inordinately difficult to develop and disseminate effective mass media health promotion programs. At least this is the conclusion that must be reached when the many evaluations of mass media health programs are reviewed. Reviews find consistently that most evaluations of media products are unable to detect significant or meaningful changes (Atkin, 1979; Flay, 1981; Flay, diTecco and Schle-gel, 1980; Griffiths and Knutson, 1960; Lau et al., 1980; Mendelsohn, 1968, 1973; Solomon, 1982; Wallack, 1981), with the notable exception however of the use of mass media for smoking cessation (Flay, 1987). Most of the emphasis in past reviews and analyses has been on finding ways to improve media products to increase their effectiveness, or on the need for improved outcome or summative evaluation (e.g. Ball, 1976; Flay and Cook, 1981; Haskins, 1970; Towers, Goodman and Zeisel, 1962; Wilde, 1975). However, it seems that many media products have failed because they were not tested adequately during development (McCron and Budd, 1981) or because they simply did not reach the target audience [e.g. most Public Service campaigns (Flay and Sobel, 1983)]. Such outcomes suggest that more emphasis needs to be placed on planning research, pre-production testing, audience testing of a finished product before it is disseminated, and the evaluation of how well the final product reaches its intended audience and is accepted by them, than on the determination of the effectiveness of the disseminated product.

Evaluating the worth of any media product is a process that is best carried out in a sequence over an extended period of time. It can involve any number of steps, but one possible sequence is as follows (Table I). Prior to message development or production, planning research can help determine the most appropriate target audience, and the needs, perceptions and language of that target audience. Such information will provide first clues to the properties of a successful media product. During message or product development, formative evaluation might include tests of concepts being considered for use in a message or product, and the pretesting of a rough form of the final product.
**Table I. Phases of evaluation**

<table>
<thead>
<tr>
<th>A. Pre-production</th>
<th>B. Post-production but pre-dissemination</th>
<th>C. Post-dissemination</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Concept testing</td>
<td>5. Efficacy</td>
<td>7. Effectiveness evaluation</td>
</tr>
</tbody>
</table>

Formative

Summative

Once a media product has been developed, audience testing of the acceptability and the short-term efficacy of the media product are desirable precur-sors to complete distribution/implementation/dissemination/marketing. Why go to the expense of disseminating a product if it will either (i) not be accepted by media gatekeepers and therefore, not get aired; (ii) not be accepted (i.e. attended to, liked by the target audience); or (iii) not produce the desired effects (e.g. knowledge, attitudes, intentions, self-efficacy) even if it does reach the intended audience?

Once a media product is disseminated, three further levels of evaluation are possible. Evaluation of the success of the dissemination or marketing strategy is desirable. This level of research will be referred to as implementation evaluation. It also serves the purpose of determining the conditions under which the product was distributed/implemented, the proportion of the target audience it reached and the type of audience it reached. Effectiveness evaluation is designed to determine the effectiveness of the media product as disseminated. It seems obvious that it is not worth attempting effectiveness evaluation until there is some confidence that the product (i) is acceptable to the target audience, (ii) is efficacious if it reaches them, and (iii) has been successfully disseminated. Finally, it is sometimes desirable to determine why a product was or was not effective, or how it produced the observed effects. This is called process evaluation.

Under ideal circumstances, the levels of evaluation introduced above might be undertaken in the order in which they have been presented. In reality, of course, one or more levels may be omitted or skipped, although usually at the expense of the interpretability of results from the highest level attempted. More detail about each of the levels of evaluation presented above is provided in the sections below.

**Pre-production research**

Pre-production evaluation research provides information that helps determine the form of the final product — hence formative evaluation (Cambre, 1981; LaRose, 1980; Mielke and Chen, 1981; Palmer, 1974, 1981; Scriven, 1967). At least three distinct levels of pre-production research are possible: planning research, concept testing and message pre-testing.

**Planning research**

Many media products have failed to gain the attention of the intended target audience, even if they were aired on appropriate channels at reasonable times. Planning research can fulfill the very important functions of gaining insight into the needs, perceptions and language of the target audience, and thereby minimize the probability of failure by not reaching the intended audience. This level of research can involve exploratory interviews with groups from the target audience, in-depth interviews with individuals representative of the target audience and small-scale quantitative surveys.

We know that media programming is most effective at reinforcing existing trends in attitudes and thereby possibly causing additional small changes in behavior (i.e. the concept of canalization, Lazarsfeld and Merton, 1948). Shifting norms may sometimes be an objective. For these objectives to be met successfully, it is very important to establish current norms and design the media messages to be within their ‘latitude of acceptance’ (Sherif, Sherif and Nebergall, 1965). Determining more definitively whether or not the proposed messages are within an acceptable range is the task of the next levels of evaluative research.

The costs of planning research are not excessive, and are well worth every penny spent. No one should
be seduced into relying on the judgements of ‘experts’ in place of carrying out this level of research. Unfortunately, expert judgements only occasionally correlate with ultimate message effectiveness (LaRose, 1980).

As planning research helps establish which approaches and concepts to try, further formative research helps choose among them and refine them (concept testing), and tests the target audience’s reactions to proposed messages (message pretesting). These phases of research are also very important, in that they should help avoid program/campaign failure due to inability to attract the attention of the target audience and have them accept the message.

**Concept testing**

Concept testing may be viewed as an extension of planning research in that similar subjects and methods are used. An additional useful method may be ‘central location intercept’ interviews, where interviewers ‘hang-out’ in locations frequented by members of the target audience and interview randomly selected subjects. The major difference between concept testing and planning research is that questioning is much more focused in the latter. The objective is to assess the acceptability to the target audience of several concepts developed on the basis of information collected during the planning research. Results from concept testing will lead to choosing from among several concepts and refining of the chosen concepts.

**Message pretesting**

Once concepts have been refined, research at a further level of detail is still necessary to ensure (or at least increase the probability) that the final media products will be acceptable (relevant and believable), comprehensible, and memorable to the target audience. This requires that a sample of the target audience be exposed to a rough form (animatics, photomastics or rough live action) of the final form of the proposed product. The objectives of this level of research are to: (i) assess the level of recall and comprehension of the message; (ii) determine its relevance and believability; and (iii) check reactions to any controversial or sensitive elements (appeal). The appropriateness of production techniques (e.g. use of voice-overs versus direct messages, whether or not music should be used, and if so, what type) may also be assessed during this phase.

Two major methods of data collection are most often used in message pretesting research: audience-response analysis (Baggaley and Smith, 1982; Palmer, 1981) and focus-group discussions (Axelrod, 1975; Higginbotham and Cox, 1979; Merton, Fiske and Kendall, 1945), preferably with the same subjects. For youth-oriented projects, schools provide an ideal location and captive audience for this purpose. Some university communication departments have audience-response analysis systems that would also be ideal for audience testing. The Health Message Testing Service (Office of Cancer Communications, 1984) at the National Cancer Institute and the National Heart Lung and Blood Institute has developed a standardized approach to message pretesting, and their approach is recommended. Recent findings from the analysis of about 60 tested public service announcements (PSAs) about what characteristics improve effectiveness are also of value in message development (Freimuth et al., 1986; Office of Cancer Communications, 1983). It is noteworthy that the valued characteristics of PSAs for health managers are similar to those for commercial managers (Ogilvy and Raphaelson, 1982). Cambre (1981), LaRose (1980), and Palmer (1981) provide succinct discussions of approaches to formative evaluation of media products.

**Audience testing of the final product**

While the results of message pretesting will often provide valid data about the final product, it would be misleading to expect this in every case or in any particular case. Final production decisions, or even substantial changes made in response to pretesting results, could easily change the effectiveness of the produced media product. It is, therefore, recommended that the final product be subject to the same (or a very similar) type of audience testing as the rough form.

The value of audience testing of a final media product lies in two areas. First, how the target audience will respond to the product when it is distributed in the ‘real world’ can be determined. This is known
as acceptability, and is concerned with audience attention to and affective reactions to the product. Second, audience testing can determine the likely effects of the product on the target audience’s awareness, knowledge, attitudes, intentions and self-efficacy. This is known as efficacy, in that it concerns product effectiveness under ‘optimal’ or ‘captive’ conditions (Flay, 1986); product effectiveness in the real world can subsequently be determined by summative evaluation.

Audience testing of final messages can be viewed as a form of process evaluation that will later contribute to the understanding of findings from summative evaluation. Results from audience testing could also be used in a formative way to improve future media products.

Post-dissemination

Once a program has been disseminated, evaluation might be designed to find how well it was disseminated, what effects it had, and how or why it produced the observed effects or did not produce any or only some of the desired effects. These three levels of evaluation may be considered as implementation evaluation, effectiveness evaluation, and process evaluation, respectively.

Implementation evaluation

Implementation evaluation addresses the question of how well, and under what conditions the program/campaign was implemented and who it reached (Johnston, 1981; Liethwood and Montgomery, 1980; Patton, 1978; Williams, 1976). As such, implementation evaluation also contributes to evaluation of dissemination or ‘marketing’ procedures (Manoff, 1985). It is a highly recommended precursor to effectiveness evaluation. The interpretation of effects data can be very difficult without some knowledge of the level and conditions of implementation (Johnston, 1981; Flay, 1986; Patton, 1978).

The methods of implementation evaluation fall into two groups. A relatively inexpensive, but limited, approach is to rely on Broadcast Advertisers Reports (BAR) data. These data provide information on which products were aired by which stations, at what times (and how often at what dollar value), over how long a period. A more expensive approach (actually an adjunct to BAR data) involves surveys and/or interviews of station managers (or other appropriate ‘gatekeepers’) in selected cities. A very useful potential of this approach is the ability to determine what other related media or community activities were implemented at the same time (either coincidentally or as a result of the program/campaign).

Field et al. (1983) have provided an exemplary implementation evaluation of a mass media campaign. They evaluated the implementation of the 1982 Alcohol Abuse Prevention Campaign of the National Institute on Alcohol and Alcohol Abuse. This campaign consisted of 12 Public Service Announcements (PSAs) targeted at women and youth. Using BAR data, they found that PSAs could have reached 85% of the nation’s TV households an average of 45 times per day (across all stations). However, only 6% of these airings were during prime time (8.00 p.m.-11.00 p.m.) and only 20% between 5.00 p.m. and 11.30 p.m. Even these lead to over-estimates of reach and frequency in that only local stations aired significant numbers between 5.00 p.m. and 11.30 p.m.; the networks aired only 3% during this time period. Thus, local situations aired more PSAs during prime time, but national networks tend to have most of the audience during those hours.

Primary findings from the Field et al. study were that effective dissemination was associated with (i) well-planned and carefully executed approaches to media gatekeepers (usually public service directors at each station); (ii) extensive involvement of community resources such as volunteers, state and local authorities, private organizations, schools, political and government leaders, experts, celebrities and other media.

Effectiveness evaluation

Effectiveness evaluation addresses the questions of the effects of the program/campaign on the target audience. There are numerous possibilities for effectiveness evaluation. One useful approach is to group them into three commonly used paradigms (Flay and Cook, 1981); (i) the ‘advertising paradigm’, which is usually inexpensive and is sen-
sitive to small effects, but usually focuses early in the causal chain so that the observed effects are often of low policy relevance; (ii) the ‘monitoring paradigm’, which uses existing archives or data recording systems to focus on more ultimate impacts that are of greater policy relevance; while also inexpensive, this approach tends to be relatively insensitive to true effects and easily produces false-negatives; (iii) the ‘experimental paradigm’, which can improve upon the low policy relevance of the advertising paradigm and the low sensitivity of the monitoring approach; this approach is often, however, more costly and time consuming.

**Advertising research approaches**

Probably the least expensive approach within this paradigm is to have specific questions inserted into regular ongoing surveys. A more expensive approach, but one that yields much more information, is a special survey of a representative sample of the target audience. An in-between option is to survey only people who request further information as a result of a particular media presentation, as most evaluations of mass media smoking cessation programs have done (Flay, 1987).

Any survey can include a range of questions, from awareness and recall, to attitudes, intentions and actual behavior (which can range from requesting more information, to discussions with peers or family, to changes in relevant behaviors). Decisions need to be made about the number of waves; post-test only, pretest -post-test, or multiple pretests and post-tests. Each option obviously costs more but also provides more information, which is more easily interpreted.

**Monitoring approaches**

Probably the least expensive effectiveness evaluation is of this type. All that is required is a simple record of the number of people requesting further information as a result of a media presentation. (Note that if written materials are to be a component of a campaign they deserve evaluation at all levels discussed in this paper, just as any other media product does.) An ability to match any request to a particular media presentation, or a particular wave of a campaign, increases the value of this inexpensive option. Then at least one can conclude that some messages are of greater value than others, at least in terms of stimulating a demand for more information. This approach has been used extensively in evaluations of the US Cancer Information Service (Stein, 1986). The approach can be improved still further by running different components of a campaign in different parts of a community at different times – see experimental approaches below.

Other monitoring approaches involve using existing archives or recording systems. Some examples are sales of cigarettes, police or school system records on drug-related incidents, hospital emergency room admissions that involve alcohol, existing questions in relevant ongoing surveys. Each of these obviously has its own problems (Cochran, 1978) and probably should not be used alone. Note too, that most archival records will focus on ultimate behavioral outcomes, and so may not be highly relevant to the realistic objectives of the proposed campaign. Most archival records can be subjected, however, to some quite rigorous analytical methods (e.g. time-series analysis can sometimes detect a shift as small as 5% – Hennigan et al., 1982).

**Experimental approaches**

Experimental approaches involve exposing some groups or locations (e.g. communities, cities) to a media product or campaign, and to exposing others (e.g. the Stanford and North Karelia community projects – Maccoby and Farquhar, 1975; Maccoby et al., 1977; McAlister et al., 1980; Puska, 1981). Samples of individuals from each group are then surveyed, and any differences between the groups at post-test are attributed to the intervention (media product or campaign). Of course, such an approach is fraught with difficulties (Campbell and Stanley, 1966; Cook and Campbell, 1979), the more so the larger the aggregated units being compared, for example in cities, communities, schools (Flay and Best, 1982; Flay and Cook, 1981). In some instances, the costs of implementing an interpretable experimental summative evaluation may argue against its use. Instead, media product developers and evaluators might sometimes be satisfied with assessments of acceptance, efficacy and implementation. Certainly, in
most circumstances, no experimental effectiveness evaluation should be performed until acceptance, efficacy and implementation have been shown to be adequate (Flay, 1986).

Process evaluation

Too few evaluations of mass media products have provided us with any information on why they were or were not effective. Without such information a coherent body of knowledge cannot develop. Each of the above levels of evaluation do, of course, provide us with some of this information. However, there needs to be more frequent use of them and a concerted effort to analyze and report results. In addition, additional process data can be collected relatively inexpensively during summative evaluation. Information on viewing levels, reactions of viewers to the media materials/messages, presenters, and context, and many different potential levels of effect (e.g., awareness, knowledge, attitudes, intentions) could inform future media product developers and researchers.

Conclusions

As the order of these notes suggest, I advocate a systematic approach to media product evaluation (and development), rather than a concentration at any particular level. Results from one level then can be used as a source of information about questions worth asking at the next level. For example, there would be little sense in doing an expensive experimental summation evaluation of a product that audiences find unacceptable or that did not reach the intended target audience, or that was not efficacious under optimal or captive conditions. These points argue for greater emphasis being placed on formative evaluation (McCron and Budd, 1981) and assessment of efficacy than is usually the case (Flay, 1986). They also suggest that assessment of efficacy be as scientifically valid as possible. Thus, experimental evaluation of media products will usually be much more useful and cost-efficient if carried out in the audience testing stage than in summative evaluation. On certain occasions it may well be prudent, therefore, to be satisfied with audience analysis and implementation evaluation of wave one of a campaign, and then consider effectiveness evaluation for a second wave. This approach should ensure the greatest value for expended resources, both in the quality of resulting programming and in the interpretability of evaluation data.

Acknowledgements

This paper was prepared with support from the National Cancer Institute (Grants CA 34622 and CA 38268) and the National Institute on Drug Abuse (Grant DA 03468).

References


Flay, B.R. (1986) Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. Preventive Medicine, 15, 451-474.


Received on 2 September 1986; accepted on 4 November 1986.