SECTION 4

Design Research and Evaluation

The final test of whether or not we have effectively bridged the applicability gap comes when designers and behavioural scientists critique pre-construction design proposals. How should the design be evaluated? And how should the results of that evaluation be incorporated into the design? The fifteen papers in this section represent efforts to fuse the disparate and often contradictory viewpoints that result when deisgners invite such scrutiny by behavioural scientists.

Four of the papers examine theoretical and systems approaches to design evaluation. The remainder focus on problems within specific design areas: the design of wards, wings, or buildings of medical or educational institutions; the design of neighbourhoods and community facilities; design for the special needs of the handicapped, elderly, or other special groups; and the design of special environments.

The first three papers form a "how-to" primer on good design research: thinking through the problem, framing alternatives, programming a solution and assessing the outcome. Campbell discusses evaluation of the design programme. Walkey, Carley, Roberts, and Bancroft detail a process for the development of planning options--specifically for regional correctional services. And Ostrander and Connell examine post-construction evaluation, arguing that such work is most valuable when it provides feedback to actual decision-makers.

The second group of papers draws on findings of specific research in special settings. In effect, these authors are saying: If we closely examine what people do in a setting, we can find clues that will lead to more effective and more humane designs for that type of setting. Kerpen, Marshall, Whitehead, and Ellison are concerned with an economical re-design of large, aging mental hospitals. Pendell and Coray used behavioural mapping and survey techniques to compare nursing unit designs in four hospitals. They then demonstrate how such data can provide feedback to administrators. Phelps and Baxter provide revealing data on how academics use their time and space. They then discuss how their data can be used in evaluating progress toward departmental goals and in assessing users'

satisfaction with available facilities. In another study of academic environments, Ochsner examines attitudes and activity patterns in a student dormitory.

The next group of papers is concerned with involving the user more directly in design decidions and their implementation. Brower, Stough, Gray, and Headley report and evaluate a variety of interesting programmes in which inner-city residents are encouraged to initiate and operate various recreational and maintenance programmes in neighbourhood parks. Bender examines how the poor in Bogota, Columbia, create and maintain their own "squatter settlements". And Pressman examines a variety of problems faced by residents of new, small, isolated Canadian communities. He also makes a number of recommendations that could benefit government agencies who are increasingly involved in planning such communities.

A final group of papers examines the needs of special users or users of special environments. From survey research data, Newcomer derives design standards on how far the elderly are able to travel to obtain needed services. Steinfeld, Schroeder, Bishop, Aiello, Andrade, and Buchanan critically review available research on standards for the accessibility of a building to disabled persons, and they promise much-needed new data to help improve these standards. Orleans discusses designs that can help normalize the institutional environment to which the mentally retarded are often subjected. And Culjat reports a study of a new community in the Canadian Arctic where climatic influence is so intensified.

The problems and hypotheses of the applied research discussed in the papers of this section have implications, of course, for designers in EDRA. But, it is also worth pointing out that they have a value to the behavioural scientists in EDRA as well. Practical probelms and the hypotheses developed in doing research on such problems provide a stimulus to basic researchers and, in some cases, even a practical test of some of their favoured notions. In other words, traffic on the bridge over the applicability gap goes in both directions. This collection of papers illustrates the numerous problems encountered by those who attempt to integrate research and application. These papers also illustrate, however, that lines of communication exist and that such communication and interaction between the two emphases can be fruitful to both.

EVALUATION OF THE BUILT ENVIRONMENT:

LESSONS FROM PROGRAM EVALUATION

David E. Campbell

Department of Psychology
University of Kansas
Lawrence, Kansas 66045

ABSTRACT

For some 20 years now, behavioral scientists have been building up a body of knowledge in the area of program evaluation. This knowledge has found application recently in the many federally-funded social action programs--most of which now require some assessment of their success and efficiency. In recent years, evaluation has become a major concern in the area of environmental design. Planners are now concerned with the performance of built environments with respect to the behavioral and cognitive goals of their designs. It would be wise for such design evaluators to ask what advice is available from the program evaluators who have had a number of years of experience in evaluation research. Social action programs and planned environments are both intended to influence behavior in specific ways. The problems of evaluation in assessing the behavioral effects of social action programs and built environments are similar in a number of ways. This paper reviews these similarities. Important questions are raised and some specific advice is given to design evaluators concerning the purposes of evaluation, design goals, data collection plans, measurement techniques, cooperation, and dissemination of findings. This advice should help environmental evaluators to avoid some of the many pitfalls that have plagued program evaluation attempts.

1. INTRODUCTION

Before discussing the lessons to be learned from program evaluators, we should be clear on just what program evaluation is. Program evaluation refers to use of the scientific method in an effort to determine the effects of a treatment or program, especially with respect to the intended effects (Struening & Guttentag, 1975; Suchman, 1967; Weiss, 1972a; Wortman, 1975). The program itself can be any systematic or formal procedure for changing humans in some way. Examples of programs are Head Start (an educational program for lower-socioeconomic children), the American Cancer Association's anti-smoking campaign, the New Jersey Negative Income Tax Experiment (to encourage employment), and the death penalty (to discourage crime). Each of these has specific objectives in changing human behavior. The program evaluator's task is to determine how well the program is working. This generally involves defining the goals of the program, gathering data according to some plan to ascertain whether the goals are being obtained and why, and communicating the results to the initiating decision-makers. Evaluators use the scientific method as their model in carrying out this task. Unfortunately, the application of experimental controls and scientific measurement techniques is far more difficult when conducting an evaluation in the "field" than when testing a hypothesis in the research laboratory. Program evaluators have encountered a number of problems in the conduct of their work that threaten to discredit the results of their evaluations. Through the years, they have also discovered approaches to their work that can minimize the inherent problems involved (Caro, 1971; Suchman, 1967; Weiss, 1972a,b).

Why should the problems and solutions of program evaluators be of concern to us in design research? Because frequently, we are involved in attempts to evaluate the behavioral impact of built environments. I anticipate that our concern with such behavioral evaluations will increase during the next 10 or 15 years. To a large extent, the evaluation of environmental designs is similar in procedure to the evaluation of social action programs. Not only are the methods similar, but the problems involved are similar. I refer here not to basic research in the design area but to evaluation research--for example, assessment of building performance in terms of user behavior. If we can anticipate the problems that we will encounter in our efforts at design evaluation, then we will have a chance at saving much wasted effort---especially when ways to minimize or circumvent these problems have been demonstrated.

It should be clear at this point that I refer to evaluation of a design's impact on human behavior--on what people actually do. I am far less interested in cognitive processes such as how people perceive their surroundings and how satisfied they feel with where they spend their time. Still, I recognize the value of such internal dimensions in helping to explain why building users behave as they do. It should be clear that this paper does not address evaluations which do not require any assessment of user behavior. Thus assessment studies of structural strength, compatibility of materials, and aesthetic preferences are specifically excluded from this discussion. My concern here with user behavior implies an important assumption--namely, that the design of setting does influence what the setting's inhabitants do. Different possible designs can result in different patterns of user behavior as a direct function of the design itself. Previous research has offered only weak support for this assumption (Gutman, 1966; Michelson, 1970). However, evidence for environmental influence is mounting for such behaviors as accidents and conversation, and for such user groups as the

aged and the handicapped. Clearly, this area is far from fully researched.

Discussion of evaluation occurs occasionally in the design literature (Lang, Burnette, Moleski, & Vachon, 1974; Sommer, 1972; Zeisel, 1975) and examples of design evaluations appear from time to time (Heyward, Rothenberg, & Beasley, 1974; Trites, 1969). But there has been virtually no airing of the issues separating successful from unsuccessful evaluation, and the closely related literature on program evaluation has received too little attention in discussions of design evaluation. A major purpose of this paper is to stimulate discussion of the relevant issues from the field of program evaluation. We have much to gain in the way of improved evaluation.

2. WHO NEEDS EVALUATION AND WHY?

One of the first concerns of the evaluator should be to determine who needs evaluation and for what purpose. This is important information simply because the type of evaluation that will be appropriate depends on the level of organization requesting it and the kind of decision involved. One may say that this is a trivial admonition. But reflect for a moment on texts you have encountered giving advice on research methods. You probably encountered information on how to operationalize measures, what methods to use in gathering data, what research plan to follow in timing the data collection, how to choose the sample of subjects, and how to analyze and summarize the resulting data. But you most likely encountered no information on how the conduct of the research depends on who needs it. In evaluation research this information is critical. Failure to attend to this information could result in weeks or months of careful evaluative research producing a report that will never be read by those who asked for it. Suppose a federal agency requests evaluation of the use of open-classroom schools. A report is needed detailing the degree to which the open-school buildings facilitate the activities of the open-education program. The federal agency funding the project may simply need to know the extent to which its objectives are actually being met in using tax money earmarked for schools facilitating a particular type of educational program. Suppose, on the other hand, that the evaluation is requested by the architectural firm that is designing the school building. The firm may be less interested in goal attainment and more interested in information that can be used in the future when a similar project is to be designed. They may be most interested in cumulative information that provides them with information on what design assumptions regarding behavior seem to be valid and which ones should be changed or discarded. Consider a third case. Suppose the evaluation is requested by the school system itself. The school administrators are committed to providing education to the community. If the school design falls short of planned objectives, the administrators can't shut down the schools and send the designers back to the drawing boards. They have to make the best of the existing situation--good or bad. Consequently, they may require evaluation data that provides them with information on what modifications need to be made to the existing facility or program to result in efficient use of the newly-built school building.

It should be clear here that the form an evaluation should take depends on who needs it and how the results are to be used. We can distinguish here between two types of evaluation: summative and formative (Scriven, 1967). Summative evaluation refers to an overall comparison of the extent to which a program, or in our case a building, is fulfilling its intended goals. Basically, the evaluation consists of a comparison of what is actually occurring with what should be occurring. In the case of formative evaluation, on the other hand, the problem would not be simply to assess how well a building "works". Rather, the problem would be to understand the nature of the fit between building and user so that this fit can be improved by making modifications or adjustments to the building. Formative evaluation would be most appropriate where buildings are purposely made with future adjustments in mind based on evaluative information, or where a few buildings are to be constructed as a pilot project before building others. One example of the former case is the practice of leaving walkways around campus buildings uncompleted until students have indicated where the walkways should be placed by wearing paths in the surrounding grounds.

Identifying the uses of evaluative information in advance can save wasted effort. Usually the results of the evaluation are required for some sort of design decision. The decision may be whether to modify the design for future buildings intended to serve the same function as when a few homes of a housing development are built on a pilot basis. Or the decision may concern whether or not to remodel an existing facility and what specific changes to make. But program evaluators can attest that evaluators are often called into service for entirely different reasons. Sometimes a program administrator needs the evaluators to find evidence justifying the future existence of an ineffective program. The analygous case in the design profession may involve a design firm which wishes to hire evaluators to provide data of a positive nature justifying efficiency of a design in hopes for future federal contracts. In such a case, the design firm may be uninterested in how effective the design actually is; it may just want some scientific-looking evidence that make the design appear affective. Obviously, the design firm would not use or even want a report from the evaluators documenting ways in which the design fell short of its goals concerning inhabitant behavior. It must be stressed here that the evaluators should find

out in advance whether negative results are permissible. If not, then the evaluators might be wise to turn down the job. Other reasons for an evaluation include attempts to find fault with a design regardless of its actual value. For example, a political group on seeing inner-city residents relocated as part of an urban redevelopment program might wish to see an evaluation study carried out that will discredit the redevelopment and prevent future relocation of other inner-city residents. Besides attempts to bias evaluation results in a positive or negative direction, occasionally evaluators find that their results are to be ignored altogether. Suchman (1967) has called this "posturing", using evaluation to assume the pose of scientific research and give the appearance of professional sophistication. It is to the evaluator's advantage to spend some time carefully probing the actual reasons for requesting an evaluation before beginning to design the project itself. Generally, this must take the form of frank interviews with individuals involved in the building design and use at several levels.

3. WHAT ARE THE BEHAVIORAL OBJECTIVES OF THE DESIGN?

One of the trickiest problems in the program evaluation area is that of defining the goals of the program. Goal definition must come before any attempt to determine whether the program goals are being met, or even when and how to measure them. The problem is that the individuals who design and administer social action programs often do not have clear and specific objectives in mind. For example, they may have only vague notions of improving the condition of the program's target population in any of a number of ways. It is the task of the evaluator to meet with the program developers and try to get them to specify the program objectives in such a way that measures can be made of how well the program is meeting these objectives. In the case of design evaluation, the situation is little better. Room labels are frequently as close as the designer comes to specifying just what behaviors are expected to occur in a proposed building. Further, there is no set of accepted principles in the design profession making explicit the various structural features that facilitate or impede certain desirable and undesirable behaviors. Instead of a set of known design principles, each designer is left with intuition and previous experiences, different intuitions, and rarely any systematic information to guide them in their decisions concerning the relation of structural features to behavior. As a result, the evaluator is confronted with a challenging task defining the behavioral objectives of a built environment. The problem is further confounded by the frequent use of visual rather than verbal modes of communication in the design profession. Perspective drawings and mat-board models are ineffective in describing the behavior of building users. When people appear in such visual communication aids, the behavior is all too often passive sitting, standing, or walking--hardly an exhaustive display of the behavioral repetoire intended of users.

It is generally necessary for the evaluator to meet several times with those requesting the evaluation to define the goals in such a way as to make them measureable. To say that a building design should encourage efficient performance by its inhabitants is inadequate. There are too many ways to define efficiency and no indication of what sort of performance the evaluator should examine. Thus, the evaluator's task often involves an effort to get the objectives of a program (or building) redefined in the direction of greater detail and specificity. Often, this means moving toward behavioral definitions. In the case of efficiency as an objective, one must decide what is efficient behavior and how to distinguish it from inefficient behavior.

There is a problem that may develop in working with designers to define the goals of a built environment. Under pressure from the evaluator, the designers may actually be making up the goals on the spot. When the behavioral objectives are decided upon after the design is completed, one can expect less committment to these post hoc goals than would be the case if the goals were foremost in the designer's mind throughout the design phase of the project. If things don't go well during the evaluation--if the environment does not seem to be performing as intended--then it is all too easy for the designers to change the goals to fit the data. They can simply say: "That isn't what we intended. Actually we meant for the users to act just as the evaluator's data show they are acting." Given this possible state of affairs, would it not be better to simply dispense with objectives altogether? Why not just go into the environment-user system after construction is completed and see how the environment is performing? Not only is the problem of goal definition avoided, but the evaluator is free to note unintended effects of the environment as well. For example, the evaluation might focus on variables such as accidents and aggressive behavior as well as more positive behavioral dimensions. Avoiding goals altogether is undesirable because it leaves the evaluator open to the accusation that the evaluation missed all the intended (and often beneficial) effects of the built environment. Also, the lack of objectives leaves the evaluator with an almost infinite array of possible variables to examine in assessing the effects of the environment. Of course, the evaluator can supplement measurement of intended effects by measurement of other variables deemed important based on theory or previous research. For example, one might include in the evaluation of a radial-plan hospital variables such as nurse travel time between nursing station and patients--a variable for which there are intended effects. But in addition, the evaluator might study variables related to general quality of

the patient's life such as variety of activities engaged in and number of different settings entered.

4. CHOOSING A PLAN FOR DATA COLLECTION

At first thought, the actual task of collecting data on user behavior seems fairly straightforward. One need merely observe how an environment is being used, then compare its use with a list of intended uses. Then, the evaluation report is made favorable or unfavorable depending on whether user behavior falls short of or meets the behavioral objectives. Unfortunately, the situation is not quite so simple as this for a number of reasons. The problem is that one needs to know not only what users are doing but why they are doing what they are doing. This brings us into the messy area of causality--what causes people to behave as they do? If the environment is being evaluated, then we need to know whether it is causing people to behave as intended. But when we observe user behavior, we can't tell whether the behavior is caused by certain aspects of the built environment or by other factors, such as the type of people who are users, the users' past experience with other settings, the particular set of rules or program for user behavior, or what. The situation is confounded all the more because we know that behavior often has more than one cause. What we do is influenced not only by where we are, but by who we are with, what we think is expected of us, how we feel, and a number of other factors. All may interact to influence our actions. This state of affairs leads statisticians to speak in terms of the proportion of variance in a given behavior that can be attributed to each of several possible determinants. It leads design evaluators to lie awake at night and ponder why the world can't be simpler. But all is not lost. We do know that sometimes the physical environment has clear effects on human behavior. The construction of stairs can have much to do with the frequence of slips and falls of users. The placement of a church collection box can influence the amount of money contributed by worshippers. The evaluator must plan the schedule for measuring user behavior in a way that will make as clear as possible the role that the physical setting plays in influencing user behavior. Depending on when and how the necessary measurements are made, one can obtain evidence as to how a built environment is influencing user behavior.

We can illustrate this point by reference to the two most frequent evaluation problems, that of evaluating a newly built environment and that of assessing the effects of a remodeling project. The task of evaluating a newly built environment is most difficult. Once people begin using the environment, any data gathered is beset by the already mentioned difficulty of deciding what behavior is influenced by the physical setting and what isn't. One way to handle this problem is to interview users, simply asking them how the building helps or hinders their attempts to behave as they wish. Sometimes people are accurate at identifying the determinants of their behavior, but only sometimes. Another way to assign causal status to structural features is to locate other environments similar in function and observe how they are used. The comparison of user behavior in the evaluated building with other similar buildings can aid in determining whether users act as they do because of building features or because of other reasons. For example, if a comparison building differs structurally but has a very similar user population (according to variables such as age, education, and socioeconomic status), then it is less likely that differences in building use are due to the type of user. Similarly, if the program, the set of rules by which people know what they are expected to do, is similar in the comparison building, then differences in behavior are less likely a function of the program and more likely a function of the physical setting. Ideally, the evaluator would locate comparison settings that are similar to the evaluated setting in every way except for features of the physical structure. If everything else is the same, then one can ascribe differences in behavior with relatively more confidence to the physical setting itself. Gutman and Westergaard (1974) used this approach in their study of academic research buildings. Assuming similar occupant populations and similar tasks, the different responses to the buildings were more likely actually causally related to the physical features of the buildings. A major task for the evaluator in using this comparison-group research plan is to locate similar environments, obtain permission to gather data in them, and obtain the needed funds from the persons who are paying for the evaluation. The added cost for increased confidence in causal relationships can be considerable.

The second type of research problem involves assessing the impact of a remodeling project. In this situation, one needs to gather data before and after remodeling and see if the changes in behavior meet what was expected. Assuming that the user population does not change appreciably during remodeling and that the behavioral program remains constant, then changes can be considered likely to be caused by the design change. Increased confidence in this assertion can be made if the assessments of building use are made at several times before and after the remodeling. If the behavior changes only at the time of remodeling, one can be even more sure of the effects of the remodeling. Even with this type of research plan, it is best to obtain a comparison group--in this case, a similar setting which did not experience remodeling, or at least not at the same time. Behavioral changes that occur following remodeling in the evaluated setting should not occur at the same time in the comparison settings. Further plans and experimental designs for data-gathering are available in publications by Campbell and Stanley (1963) and by Bechtel

(1974). The need for careful attention to the choice of plan for data-gathering cannot be over-emphasized in design evaluation. With the designers implicit assumption that the physical design does influence its inhabitants, it is all too easy to observe user behavior and glibly ascribe it to the design--ignoring the myriad of other possible causes which can provide competing explanations for user behavior. Unless the design evaluator is willing to argue forcefully for a good research plan, the final evaluation report is in danger of giving a very misleading picture of the behavioral performance of a built environment. Perhaps user behavior will be accurately described, but its ties to the specific design features will be left in serious question.

5. CHOICE OF MEASURES

Choosing the appropriate measuring instruments is especially difficult in design evaluation because this field is so new. However, there are some considerations from the area of program evaluation that should be heeded. One of the most important things to do in program evaluation is to specify the program. Generally the actual program varies considerably from the intended program. Similarly, the final built environment may be quite different from its original design due to a number of factors such as availability of materials and construction costs. Therefore, measures of user behavior should be accompanied by measures of the built environment.

The actual measures made will depend to a large extent on the goals or objectives of the design. But whatever the goals, the measurement phase of the evaluation will probably involve some sort of systematic observation of user behavior. This might take the form of trained observers using the behavior mapping technique (Ittelson, Rivlin, & Proshansky, 1970) and possibly having a sample of the user population fill out activity records of some sort (Michelson & Reed, 1975). One question confronting the evaluator is whether to use direct measures of behavior (like direct observation) or to rely on cognitive measures (satisfaction with an environment). The answer to this question is that both can be useful. One particular role that cognitive measures can play is in helping to explain user behavior. To some extent, we behave as we do because of our intentions and perceptions. Thus the evaluator can learn much by asking inhabitants (by means of interview or questionnaire) about their beliefs and intentions concerning their activities in the setting of interest.

There is a particular danger in the use of cognitive measures. They are relatively easy to use--far easier than direct observation in many cases. Probably one of the easiest ways to gather data is to slap together a questionnaire of items that seem reasonable and pass it out to the user population. Since such cognitive techniques are so easy, evaluators are likely to dispense with the more difficult behavioral measures. They can obtain data on how people respond to a setting in terms of satisfaction with the environment, then assume that behavioral performance varies directly with satisfaction. The assumption is that satisfied people perform well and dissatisfied people perform poorly. Unfortunately, this relationship rarely holds up when it is checked carefully. In fact, it has been shown that cognitive measures and the behaviors that follow from them frequently appear to be independent dimensions. (Mischel, 1968; Wicker, 1969) For this reason, the evaluator must carefully validate any cognitive measure that is intended to provide indirect data about user behavior. For example, if one wanted to use user estimates of their own activities in a setting (a cognitive dimension), then for at least a sample of users, measures of their behavior must also be made directly (a behavioral dimension). This allows the evaluator to check out the accuracy of user estimates and make sure they provide data on actual user behavior. Potential methods have been summarized by Michelson (1975) and are described in previous EDRA proceedings.

Even when measures are carefully validated in the early stages of an evaluation project, each measure has its shortcomings. Observation sometimes influences the behavior of the observed persons, and questionnaires suffer from all sorts of biases such as central tendency, halo effect, and leniency error--not to mention low reliability. Because of the inherent imperfection of current measures in evaluation projects, the evaluator is well advised to use more than one measure for each dimension where possible. Hence, behavioral mapping can be supplemented by time budgets and interviews to get a more complete and accurate picture of user behavior. Examples of the use of multiple measures can be found in Heyward's study of three playgrounds (Heyward, et al., 1974) and Van der Ryn's study of college dormitories (Van der Ryn & Murray, 1967).

6. COOPERATION

One problem that particularly plagues program evaluators is that of gaining the cooperation of program administrators and staff (Aronson & Sherwood, 1967). It is hard for the academician who does research in laboratories with college-student subjects to appreciate the difficulties that arise in field studies of an evaluative nature. The lack of cooperation shown by program administrators may be partly justified. It is certainly true that evaluation reports frequently report only that social action programs have had no measureable impact despite the thousands of tax-payer dollars that go to support them. Thus the evaluator is seen as a very real threat to the program's continued existence. As design research becomes more commonplace, evaluators may encounter a lack of

cooperation with two groups--the designers and the environment users. It is possible that designers would resent the evaluator's presence because the evaluator poses a threat of sorts. After all, the designer believes in the design. The designer's reputation is at stake if the design proves to be a failure--the more so if the design involves expensive construction. The evaluator serves as a reminder that the design may not perform as expected and the evaluator would be just the one to bring any shortcomings to light. This places a burden on both parties to understand the position of the other. Certainly the designer should realize that the evaluator can provide useful information that can be applied to future designs, making them more effective than they would be otherwise. The evaluator needs to have a healthy respect for the somewhat vulnerable position of the designer. The literature on the differing viewpoints of the behavioral scientist and the architect can be of use in helping to sensitize the evaluator and designer to each other's positions (Altman, 1973; Gutman, 1972). It can also be valuable for the evaluator to make sure that the evaluation data is of use to the designer. This may require special effort in gathering data of interest to the designer. Another important way to encourage cooperation is to give the designer as active a role as possible in designing the evaluation plan from the outset.

The other group that may not always be completely sympathetic and helpful to the evaluation project is the design user. Argyris (1968) has discussed some of the ways in which research subjects respond in an undesirable manner to the requests of the researcher. The situation of the design user is similar to that of the research subject. The user can perceive the evaluator's attempts at gathering data as an unpleasant burden from the bothersome task of filling out forms to the invasion of privacy aspect of direct observation. Further, it is easy for the evaluator's intentions to be misread. Industrial studies have shown that people in their jobs may see the evaluator as a lacky of the management, gathering data preliminary to increasing the employees' workload. Design users may show similar suspicion of the evaluation. The best way to gain the cooperation of the users is to involve them in the evaluation project at an early stage. This may take the form of working with representatives of the user population. For example, if one wanted to evaluate the ill-fated Pruitt-Igoe housing project in St. Louis soon after it was built, one might worry about the cooperation of a transient group of lower-socioeconomic inhabitants. It would be wise to recruit some Pruitt-Igoe residents to explain what was needed and to find out how best to contact a sufficiently large sample of residents for the evaluation. Another way to encourage cooperation is to make sure that the residents receive something of value to them for cooperating in the evaluation project. For example, people could be paid a few dollars for keeping time budgets. If children were to be observed in the play areas, then parents could be given feedback on their children's behavior. It takes some imagination on the evaluator's part to find ways to reward all who give something of their time and effort to provide evaluation data. Yet the return in cooperation and better data for the evaluator and for future projects with the same people makes the effort worthwhile.

7. HANDLING THE RESULTS

In most academic research, handling the results of a study is almost routine. The investigator presents the findings in a paper at the next professional meeting and sends a report in a standard format to the appropriate journal. Handling the results of an evaluation study is not quite so simple. The purpose of an evaluation study is usually to provide data needed in making certain decisions--in the case of design evaluation, decisions about future similar designs. However, evaluation reports frequently find their way onto storage shelves or into files without having any impact on decisions of any kind. An interesting example of how evaluation data can be handled is reported in Weiss (1972c). In 1940, the RAF Bomber Command was using aerial photography to evaluate the effectiveness of bombing raids. Photos were examined for aircraft that had reported their bombs on target. The photographic evidence indicated that only 25% of these "on target" aircraft even got within 5 miles of the target. One officer responded to this evaluation data by refusing to accept the report. Apparently it was less threatening to discredit the evaluation than it was to accommodate some undesirable and upsetting findings. Program evaluators can provide many more examples of rejected evaluation findings. This relates back to the discussion of why the evaluation is being conducted. If only positive results are acceptable, it is best to find this out ahead of time before wasting precious energy on a serious evaluation effort. It cannot be emphasized too strongly in this regard that one must find out at the outset who needs the evaluation results and what they intend to do with them. There are other ways in which the results can be made more acceptable. If it is possible to set up a situation in which several alternative designs are compared simultaneously, then the results may be more palitable. For example, one might design and evaluate several student dormatory arrangements on a pilot basis before embarking on design and construction of the rest of the dormatories that are scheduled to be built. Picking the best of several plans is less threatening to the designer's ego than placing a single design on a spot on a single good--bad continuum.

The actual mode of presenting evaluation reports can influence the impact of a study. The evaluator usually must present the evaluation findings to a group of decision makers who are less

knowledgable about research techniques and statistical analyses. Thus the report should not be phrased in terms of research jargon. It is best to present a summary of the findings with charts, graphs, and photographs to illustrate the major points. The details on methods and analysis can be presented in a set of appendices for those who are interested. When presenting to designers, the evaluator must be aware that the design profession is familiar with visual displays as a means of communication. The factor should be made use of by presenting the findings on the designers terms through visual displays. Afterall, the goal in dissemination of results is to communicate effectively, not to achieve satisfaction with an elegant and sophisticated research report. The designers may be impressed by scientific elegance, but they won't use the data.

Another important point in presenting the findings is to draw out the implications of the findings clearly. Researchers are used to maintaining a cautious position with respect to their data. Since one can never be absolutely certain of the generality of one's findings, the researcher is usually reluctant to state the implications of the findings in strong terms. The evaluator, on the other hand, cannot afford the luxury of presenting data and leaving the reader to draw the conclusions. The cost of doing so would probably result in an evaluation without any substantial impact. For this reason, the evaluator should consider the implications of the findings and make very clear recommendations based on the report. The evaluator may even have to become an advocate of the report's findings if the evaluation effort is to have a meaningful impact.

One particularly vexing problem about evaluation data is that evaluation reports too often do not get published--hence, they do not become readily available to others. Since the information does not get out, it is difficult for the results of evaluation to be cumulative. Design evaluation is unlikely to result in a costly change to an already built environment. The value of the information lies in its impact upon future building projects. Thus, it is important that evaluation findings become available to others in the profession. There is no easy solution to this problem. Journal editors are often reluctant to accept the results of evaluation studies on the grounds that they may be of little relevance to theory development and pertinent only to the evaluated setting. Other possible outlets should be explored, including selected bibliographies such as those compiled by the Council of Planning Librarians and computor information storage facilities. (See Sommer, 1972)

8. SUMMARY AND CONCLUSIONS

With the increase in evaluative studies involving built environments as their target, we can anticipate a rediscovery of many of the problems common to program evaluation research in education and sociology. Environmental publications such as the EDRA proceedings have produced a number of discussions of the use of behavioral research methods in design research. However, only a few of these publications have specifically focused on evaluation research in built environments. (e.g., Gutman & Westergaard, 1974; Sommer, 1972; Zeisel, 1975) Only a few of the issues peculiar to evaluation research have been aired thus far. It is hoped that this paper has served to present some of the other considerations that the design evaluator should ponder before embarking on an evaluation project. The importance of finding out who needs the evaluation and why has been stressed. The reasons given include the fact that evaluators are often called into action for reasons other than that someone needs assessment data on a built environment. Sometimes the reasons for the evaluation are clearly political. Formative and summative evaluation were proposed as two possible types of evaluation that may be required. The need for and difficulty of obtaining measurable design goals was discussed since evaluation generally requires some comparison with what occurred and what was intended to occur. Some of the difficulties in choosing a plan for data collection were covered. Comparison or control environments were advocated as an aid in determining what role the physical environment plays in influencing user behavior. The reader was advised to avoid unvalidated cognitive measures when the goal of evaluation is to assess user behavior, not just user satisfaction. Problems of lack of cooperation in evaluation research were presented. Including designers and users in evaluation decisions was advanced as one technique for eliciting cooperation. Finally, the problem of presenting evaluation results in such a way that they will not be ignored was noted along with suggestions for handling this problem; these included attention to presentation format, knowledge of who will use the results, and making clear recommendations that follow from the data.

This paper is not meant to be an exhaustive coverage of all the pertinent issues in program evaluation of which design evaluators should be aware. The author has prepared an annotated bibliography that will serve as a guide to the program evaluation literature for those in the man-environment research (Campbell, 1975). Attention to the already extant literature in program evaluation should have a beneficial impact in making future design evaluation studies of maximum use to environmental researchers and to the design profession.

References

ALTMAN, I. Some perspectives on the study of man-environment phenomena. Representative Research in Social Psychology, 1973, 4, 109-126.

ARGRYIS, C. Some unintended consequences of rigorous research. Psychological Bulletin, 1968, 70, 185-197.

ARONSON, S. H., & SHERWOOD, C. C. Researcher vs. practitioner: Problems in social action research. Social Work, 1967, 12, 89-96.

BECHTEL, R. B. Experimental methods in environmental design research. In J. Lang, C. Burnette, W. Moleski, & D. Vachon (Eds.), Designing for human behavior: Architecture and the behavioral sciences. Stroudsburg, Penn.: Dowden, Hutchinson & Ross, 1974.

CAMPBELL, D. E. Evaluation research in man-environment systems: An annotated bibliography. Unpublished manuscript, 1975.

CAMPBELL, D. T., & STANLEY, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

CARO, F. G. Readings in evaluation research. New York: Russell Sage, 1971.

GUTMAN, R. Site planning and social behavior. Journal of Social Issues, 1966, 4, 103-115.

GUTMAN, R. The questions architects ask. In R. Gutman (Ed.), People and buildings. New York: Basic, 1972.

GUTMAN, R., & WESTERGAARD, B. Building evaluation, user satisfaction, and design. In J. Lang, C. Burnette, W. Moleski, & D. Vachon (Eds.), Designing for human behavior: Architecture and the behavioral sciences. Stroudsburg, Penn.: Dowden, Hutchinson & Ross, 1974.

HEYWARD, D. G., ROTHENBURG, M., & BEASLEY, P. R. Children's play and urban playground environments: A comparison of traditional, contemporary, and adventure playground types. Environment and Behavior, 1974, 6, 131-168.

ITTELSON, W. H., RIVLIN, L. G., & PROSHANSKY, H. M. The use of behavioral maps in environmental psychology. In H. M. Proshansky, W. H. Ittelson, & L. G. Rivlin (Eds.), Environmental psychology: Man and his physical setting. New York: Holt, 1970.

LANG, J., BURNETTE, C., MOLESKI, W., & VACHON, D. (Eds.). Designing for human behavior: Architecture and the behavioral sciences. Stroudsburg, Penn.: Dowden, Hutchinson & Ross, 1974.

MICHELSON, W. Man and his urban environment. Reading, Mass.: Addison-Wesley, 1970.

MICHELSON, W. (Ed.). Behavioral research methods in environmental design. Stroudsburg, Penn.: Dowden, Hutchinson & Ross, 1975.

MICHELSON, W., & REED, P. The time budget. In W. Michelson (Ed.), Behavioral research methods in environmental design. Stroudsburg, Penn.: Dowden, Hutchinson & Ross, 1975.

MISCHEL, W. Personality and assessment. New York: Wiley, 1968.

SCRIVEN, M. The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), Perspectives of curriculum evaluation. AREA Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand McNally, 1967.

SOMMER, R. Design awareness. Corte Madera, Calif.: Rinehart,Press, 1972.

STRUENING, E. L., & GUTTENTAG, M. (Eds.). Handbook of evaluation research. Beverly Hills, CA: Sage, 1975.

SUCHMAN, E. A. Evaluative research. New York: Russell Sage Foundation, 1967.

TRITES, D. K. Radial nursing units prove best in controlled study. Modern Hospital, 1969, 112, 94-99.

VAN DER RYN, S., & MURRAY, S. Dorms at Berkeley: An environmental analysis. Berkeley, CA: Center for Planning and Development Research, University of California, 1967.

WEISS, C. H. Evaluation research. Englewood Cliffs, N. J.: Prentice-Hall, 1972a.

WEISS, C. H. (Ed.). Evaluating action programs: Readings in social action and education. Boston: Allyn & Bacon, 1972b.

WEISS, C. H. Utilization of evaluation: Toward comparative study. In C. H. Weiss (Ed.), Evaluating action programs: Readings in social action and education. Boston: Allyn & Bacon, 1972c.

WICKER, A. W. Attitudes vs. actions: The relationship of overt and behavioral responses to attitude objects. Journal of Social Issues, 1969, 25, 41-78.

WORTMAN, P. M. Evaluation research: A psychological perspective. American Psychologist, 1975, 30, 562-575.

ZEISEL, J. Sociology and architectural design. New York: Russell Sage, 1975.