CHAPTER 10

THE ASSESSMENT OF LANDSCAPE QUALITY: MAJOR METHODOLOGICAL CONSIDERATIONS

Joanne Vining and Joseph J. Stevens

The authors gratefully acknowledge the contribution made by Dr. Terry Daniel to the development of the conceptual basis of this chapter.
INTRODUCTION

Numerous methodological approaches have been applied to the assessment of the landscape, but they differ in scope, purpose, logic, and utility. The purpose of this chapter is to provide an overview of those methods. Several general categories of assessment methods will be distinguished, defined, and evaluated. The primary emphasis in this chapter is on the assessment of perceived landscape quality via observer-based measures.

Landscapes are managed for two basic purposes: to obtain certain tangible commodities or products from the land, such as water, timber, or forage, and for production of intangible commodities. These “intangibles” may be divided into two classes—symbolic and aesthetic. The landscape, or some particular landscape feature, can serve as a symbol or representation of some entity that has value. For example, a mountainous landscape might represent a challenge for one person or a mystical experience for another. Other symbolic goals for landscape management are the so-called “existence values.” That is, even though a wilderness area, wildlife preserve, or urban forest might rarely be contacted directly, the knowledge that it exists is an important symbolic value for many.

The aesthetic value of a landscape is embodied in its visual merit. Of course, there is considerable overlap between symbolic and aesthetic values. Symbolic values may interact with the perception of the visual quality of a scene (Anderson 1981), or, conversely, aesthetic quality may enhance the symbolic image. However, symbolic values are not dependent on the features of a specific landscape; symbolism transcends the immediate experience of the landscape. On the other hand, aesthetic values are linked to the experience of a particular landscape at a specific time and therefore involve a more direct interaction between a person and particular landscape features.

Traditionally, two approaches have been taken in assessing landscape aesthetics. Arthur, Daniel, and Boster (1977) termed these two approaches the descriptive inventory and public preference models. The descriptive model, also called the formal aesthetic model by Daniel and Vining (1983), assumes that aesthetics are inherent in the landscape so that a description of landscape characteristics can presumably provide an evaluation of its aesthetic quality. Descriptive approaches typically rely on the standards of trained experts (Smardon Chapter 9 this volume) and do not assess the perceptions of the public (Palmer, this volume; Daniel and Vining 1983; Arthur et al. 1977; Zube, Sell, and Taylor 1982).

Arthur et al. (1977) questioned the validity of the assumptions of descriptive models and suggested that an assessment approach based on the preferences of the general public is more practical and appropriate. Daniel and Vining (1983) echoed this concern and further stated that the concept landscape quality requires human perceptual and judgmental processes, preferably those of the general public. The conceptual basis for the public preference approach to landscape quality assessment is diagrammed in Figure 10.1. This model views design or the assessment processes as a function of the characteristics of the environment, the actions of designers and managers, plus the needs, demands, and perceptions of the observing public or user. This is in contrast to the descriptive approach in which only the environmental characteristics and activities of designers or managers are explicitly considered.

The user-based or public preference approach explicitly incorporates the viewpoint of the public. More importantly, it provides a double feedback loop represented by the clockwise and counterclockwise circles of arrows in Figure 10.1. In the inner counterclockwise loop, management actions produce changes in the environment which elicit perceptions and judgments of the landscape from the observing public. These perceptions and judgments then constitute feedback to the manager regarding the landscape, thus affecting future designs and plans. The outer clockwise loop represents a similar feedback system operating in a different direction. Here, the actions of the public produce certain environmental effects which then influence the manager’s perception of environmental capacities and constraints. The manager then completes the feedback loop to the public through education and communication regarding the environmental consequences of users’ behavior, thereby affecting their behavior. This model, which explicitly and overtly incorporates the actions, perceptions, and judgments of the public in the design or assessment process, forms the conceptual basis for the landscape quality assessment methods presented in this chapter.
There are several reasons why a conceptual link is needed between the user or observer of the landscape, designers and managers, and the landscape itself. First, understanding the interaction between people and environments is important, regardless of whether immediate applications or requirements exist. As with other areas of basic science, the study of this interaction may yield useful information or lead to unexpected gains in more utilitarian applications.

A second reason is legal. The National Environmental Policy Act of 1969 and other legislation addressing specific actions or agencies require the evaluation of the aesthetic effects of most land management decisions (see Smardon et al. Chapter 2 this volume). Evaluation of scenic consequences is often done by landscape professionals, but the addition of user-based landscape quality assessment may be advisable. Legal challenges of management decisions are commonplace. Systematic assessment of the public's perception of scenic effects of landscape management and design enables more informed planning decisions, provides important communication and educational messages for the public, and may help to circumvent costly legal battles.

Although fulfilling legal requirements and decreasing the number of angry, court-bound citizens are worthy reasons for the assessment of perceived landscape quality, a more compelling ethical reason also exists. Public lands are held in trust for all citizens and should be managed with their best interests in mind. User-based assessment of perceived landscape quality is therefore one way of obtaining public opinions that may be used in the formation of policy or in environmental planning decisions. Participation of the public has been mandated legally; for the responsible planner it is also an ethical mandate.

A fourth reason for obtaining systematic assessments of public landscape preference is to improve the design or planning process. Like the postoccupancy evaluation in architecture, a post-management evaluation of landscape quality can provide useful feedback to planners. This information may then be used to avoid past errors, imitate past successes, and plan new projects more intelligently, responsibly, and responsively. The public preference model of landscape quality assessment provides a conceptual basis from which assessment methods may be generated, utilized, and evaluated.
It is important for landscape designers and planners to understand these methods in order to (1) conduct their own assessment for planning and evaluation purposes, and (2) properly interpret the results of assessments conducted by others. The remainder of this chapter will review major procedural and methodological issues that should be considered when attempting to assess perceived landscape quality or to evaluate the assessment efforts of others.

**ASSESSMENT METHODS**

Assessment of landscape quality is a three-step process. The first phase entails the design and planning of the assessment. The second phase consists of the actual measurement procedures: characteristics of interest, whether they be attitudes, judgments, or perceptions, are systematically assessed. The third stage of landscape quality assessment involves drawing inferences about measurement data collected in the second phase. For example, one might measure attitudes or perceptions of a specific group in a particular environment and then attempt to make inferences to other groups or environments. In this section, three measurement methods will be described: the survey or questionnaire, perceptual preference assessment, and behavioral measures. The following section will describe four assessment designs which allow varying degrees of inferential power.

To unify this discussion, the problem of determining a site for a nuclear power plant will be used as an illustrative example. Assessment of public preference for such a plant would obviously enhance the planning process and might also defuse opposition to the project. When appropriate, Petrich’s (1979) study of nuclear power plant sites will be used as a case study. Otherwise, a similar hypothetical nuclear plant project will serve as an example.

**Surveys and Questionnaires**

Surveys and questionnaires have been a very popular means of assessing the opinions, attitudes, and perceptions of the general public and can be very useful for probing complex management options or issues. Generally, a series of questions is submitted in written or oral form to a sample of people representative of those likely to experience the impact of the proposed management action. Surveys and questionnaires are deceptively easy to compose and distribute, and are often used when other methods might be more appropriate. The correct use of surveys and questionnaires, however, demands consideration of many issues, only a few of which are described here. Dillman (1978) provides a thorough discussion of the issues involved in the use of surveys or questionnaires.

Sampling is a critical issue for users of surveys and questionnaires. It must be determined in advance which people will be the appropriate respondents. For our nuclear plant example one might survey all those who might benefit from lower rates or better service, or residents of a geographic area surrounding the proposed development. Once the appropriate population is identified, a sampling procedure may be used to select the respondents. A basic rule is that results may be generalized only to those individuals who in principle had an equal chance of being included in the survey sample.

A related issue critical to survey administration is the proportion of responses received from sampled individuals. Phone surveys generally result in high response rates, but may discriminate against people who do not own phones. Surveys distributed through the mail or on site may, for various reasons unknown to the surveyor, not be returned. For example, a questionnaire might be lost and therefore not returned by an interested individual, or not returned because an individual is disinterested. Those surveys that are returned may then constitute a biased sample and the surveyor will have no idea of the reason for or direction of the bias.

Interviews are perhaps the most effective means of administering surveys primarily because personal contact usually results in a higher response rate. Furthermore, the interviewer may probe for explanations of ambiguous or intriguing responses. However, because of this personal contact, and the unique interactions between particular respondents and surveys, the results of interviews may vary and often differ substantially from those obtained from identical written surveys.

Survey questions may be phrased in many different ways, any of which may introduce bias. For example, the management option preferred by the surveyor may be presented in a favorable light, or the wording of the questions may constrain or
preclude some responses. Relatively subtle differences in labeling of stimulus scenes (Anderson 1981) and in emphasis of problem issues (Vining 1983) have been shown to influence responses to judgmental or problem-solving tasks similar to those in the public participation process. For instance, asking for a response to neighborhood noise could produce a different answer than a question about neighborhood sound quality.

**EXAMPLE 1: Methods and Procedures of Response Measurement**

This section will briefly describe several of the most common procedures for quantifying the response of the observer. Each of the methods is characterized by certain advantages and disadvantages. The actual performance of each method, however, is largely dependent on the specific procedures used to implement the method in the particular assessment application. The utility and interpretability of the method is also dependent on the manner in which the resulting observer responses are analyzed (Example 4 presents one such analysis method). A more complete discussion and comparison of different response methods can be found in Baird and Noma (1978), Engen (1971), Hake and Rodman (1966), and Torgerson (1958).

The central purpose of the different response methods is to provide an observable indicator or response from which a scale of measurement may be derived. It is possible to use “direct” response scales similar to the measurements that might be performed in physics. For example, the observer might judge the length of an object by reporting its length proportional to a second object used as a standard. The observer’s judgment of length would represent a “direct” psychological scale of length. Often, however, direct scales of measurement are neither feasible nor attainable. This often occurs when the property to be measured has no immediately definable physical scale. For example, a concept such as beauty does not automatically suggest a physical measure of beauty. Many response methods are an attempt to provide an indirect scale of measurement that allows the quantitative representation of the observer’s response where no other scale is available.

**The Method of Paired Comparisons.** This response method consists of the systematic pairing of objects or stimuli. As each pair is presented, the observer makes a judgment indicating which member of the pair has a greater value of some attribute. For example, scenic beauty of several landscapes might be assessed by presenting pairs of landscape scenes to the observer. On each presentation, the observer would indicate which scene is perceived to have greater scenic beauty.

The method of paired comparisons is a relatively simple task for the observer. Only two environmental stimuli or objects must be judged on a given occasion and the simultaneity of judgment often allows the observer to make fine discriminations among objects. Usually, all possible pairs of the objects (N[ N – 1]/2 pairs) are presented to the observer in a random order. The number of pairs increases rapidly as more objects are used in the task and, therefore, the method is cumbersome when many objects are judged. Analysis of observers’ responses typically involves the calculation of the proportion of times each object was judged as “greater” than every other object. These proportions are then transformed to standardized values that represent the scaled judgments of the objects (Torgerson 1958).

**Categorical Rating Scales.** The most common response procedure is the rating method. This method requires the observer to choose which of a number of categories best reflects his or her perception of objects or stimuli presented one at a time. The observer’s rating of the object may be expressed by numerical or nominal categories. In practice many rating scales combine a numerical scale with nominal adjectives that describe some or all of the numerical categories. For example, a landscape view might be presented to the observer whose task would be to rate the scenic beauty of the landscape on one of the following scales.

<table>
<thead>
<tr>
<th>Low Beauty</th>
<th>High Beauty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Very Low</th>
<th>Low</th>
<th>Neutral</th>
<th>High</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenic Beauty</td>
<td>1 2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Rating methods are also commonly used for the assessment of attitudes, beliefs, or statements. For example, a survey might incorporate the rating method by asking the respondent to rate his or her agreement with a particular statement:

Preservation of the natural landscape is very important to me.

<table>
<thead>
<tr>
<th>Agree Strongly</th>
<th>Agree</th>
<th>Disagree</th>
<th>Disagree Strongly</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

The number of rating categories should probably be limited to 10 or fewer. There is some disagreement as to whether a neutral middle point should be included in the scale. It has been argued that the omission of a neutral center category forces the observer to make a more definite choice, resulting in a more reliable measurement. Others have argued that omission of the neutral point results in the use of two interior categories of the scale as a neutral point, thereby reducing the true number of categories (Tzeng 1983). Whether a neutral category is used depends on the nature of the observers, the task, and the stimuli to be rated. If, for example, very fine distinctions must be made, a neutral category might be overused, resulting in decreased reliability. The type of labels associated with the rating categories can also influence the observer’s response (Lam and Klockars 1982). Caution should therefore be exercised in the construction and design of rating scales. It is also important to use a method of analysis that incorporates some standard of comparison or scaling of the obtained ratings. Without such a procedure, it is difficult to interpret how high a rating is “high” (Daniel and Boster 1976; Torgerson 1958). Overall, however, the method of rating is relatively easy to administer and is quite tractable for inclusion in a wide range of assessment procedures and designs.

 Sorting Methods. A number of procedures have been used in which the observer arranges several objects or stimuli into piles or categories. These procedures are quite similar to the method of rating in that the observer assigns each object to a particular category. The sorting procedures differ from rating methods in that the procedure usually entails the direct physical manipulation of placing objects into the piles. The piles or categories may be labeled as in the rating methods. For example, several landscape scenes might be sorted by the observer into five piles, ranging from pile 1, low aesthetic value, to pile 5, high aesthetic value. Or one might ask the observer to sort scenes based on his or her own idea of intrinsic differences between them. The sorting procedures usually require that all objects are available throughout the sorting process. In some sense, the observer’s task involves the simultaneous comparison of all objects. As a result, sorting procedures often become quite difficult for the observer as the number of objects sorted becomes large.

A useful variation of the sorting procedure that facilitates the analysis of observer’s responses has been used by Pitt and Zube (1979). This procedure requires that a certain number of objects are sorted into each pile or category. This standardizes the sorting process so that every observer creates categories of equal size, thereby facilitating the analysis of responses.

The Method of Magnitude Estimation. This method represents the most direct approach to the scaling of the observer’s perceptions. The observer examines an object or stimulus and assigns a number to his or her perception of the object on a stated attribute. For example, the observer might be instructed to assess the scenic beauty of a series of landscape scenes and to provide a numeric estimation of the scenic quality of each scene. No constraints are placed on the range of numbers the observer may assign. This procedure is sometimes altered by first presenting the observer with a standard object that is assigned a predetermined value or anchor. For example, a standard landscape might be presented to the observer with the instruction that the scene represented a “10.” The observer would then be instructed to judge the magnitude of all other objects given that the standard was a “10.” Use of such an anchor, however, usually influences the rest of the magnitudes obtained from the observer.

With the method of magnitude estimation, a wide range of objects are presented to the observer, one at a time, in random order. The method is also used most commonly when the observer’s judgments can be compared with a corresponding physical continuum, such as particulate air pollution. Since different observers are certain to assign different values and ranges of values to the objects, analysis of observers’ responses should include some form of standardization of the estimates.

The response given by a survey participant depends on the way in which the question is worded. There are two basic formats for survey ques-
tions: open-ended and closed-ended. An open-ended question elicits a more variable response from the participant. For example, a participant at a public forum might be asked what he thinks of a proposed nuclear power plant development near his home, and why he feels as he does. A closed-ended question narrows the response ahead of time. A participant might be asked to rank several alternatives in order of attractiveness or suitability, or to specify which of several alternatives she favors. Open-ended questions provide rich information at the cost of coding, analyzing, and interpreting those data. Because responses are more variable, they may be difficult to interpret. Closed-ended questions provide more precise and analyzable information, but at the expense of constraining the respondents' answers. For example, if a respondent is asked only to choose one of three sites for a nuclear power plant, the surveyor will not know why or with what level of conviction the response was made. Open-ended questions are optimal in pilot studies where issues must be identified. After this point, closed-ended questions should primarily be used, but questions must be written carefully and tested to ensure they measure the intended information.

A final issue with survey and questionnaire use is the distinction between attitudes and behavior. Briefly, the nature of this distinction is that people occasionally will behave contrary to, or at least not strictly in accordance with, their expressed attitudes and beliefs. For example, a respondent may indicate that conservation is preferable to any new power plant, but fail to use conservation or alternative energy systems at home. The nature of this distinction is complex, and research on the issue is equivocal at this point (Bell, Fisher, and Loomis 1978). The best approach is to employ more than one method for collecting data, and to interpret data cautiously. That is, an attitudinal survey is not necessarily a direct indication of the behavior of respondents.

EXAMPLE 2: Psychophysics and the Representation of the Observer's Response

Several of the methods of landscape assessment can be traced to the methods of psychophysics. Psychophysics is the study of the relationship between the occurrence of environmental events (the physical world) and the resultant perceptual response (the psychological world) of the observer. To the extent that we rely on the perceptions of the observer in landscape assessment, the methods of psychophysics are particularly relevant.

In analyzing and interpreting the observer's perceptual response, it is often assumed that the response is more than a simple reaction to environmental events. For example, suppose an observer has been presented with a photograph of a landscape view and asked to rate the "scenic value" of the landscape using a five-point scale ranging from 1 (very low scenic value) to 5 (very high scenic value). Most would agree that the representation below, Figure 10.2, is an oversimplified diagram of the observer's perceptual/judgmental process. In essence, the observer's rating (r = 4) arises directly from the sensory process of observing the scene. This representation is quite mechanical and does not consider the observer's cognitive or mental processes that mediate the response to a stimulus. Although these intervening processes are not directly observable, a more satisfying account of landscape perception would include at least several such processes. The purpose of theoretical psychophysics is to study, albeit indirectly, these hypothesized processes and their relation to the observer's response.

Therefore, there is a useful link between the assessment of perceived landscape quality and psychophysics. In order for the landscape professional to benefit from psychophysical methods, this link must be explicit and avoid the oversimplification of Figure 10.2. The observer's response is a function of several complex and interactive processes. The interpretation of that response should acknowledge the operation of cognitive processes such as judg-

![Figure 10.2](image-url)
ment or decision making, as well as the observer’s particular experiences and frame of reference. Therefore, the Figure 10.2 representation of perception must be modified to include, at the least, several intervening processes as diagrammed below. The diagram in Figure 10.3 emphasizes two intervening processes: cognition (for example, judgment, memory) and perception. These processes are mutually interactive and may influence or alter the interpretation of the observer’s response (as indicated by the $i$th and $j$th influences on the rating $r$) from that obtained in a simple, isomorphic sensory system (for example, Figure 10.2).

A complete discussion of theoretical psychophysics is far beyond the scope of this chapter (see Baird and Noma 1979; Torgerson 1958). However, it is important to understand the difficulties of simplistic or direct interpretations of assessment information. In Model 1 (Fig. 10.2) the observer has rated the scenic value of a landscape by marking a “4” on the rating scale. Many would interpret this response as a direct indication that the observer’s perceived value of the landscape was quite high (a “4” out of 5). However, if the observer never used scale values below 4, the rating might actually be low. In this case, memory or experience with particular landscapes might intervene in the perceptual/judgmental process. A direct interpretation of the rating of “4,” common in many assessment studies, is tantamount to accepting the representation of perception given by Figure 10.2. If we accept the richer conceptualization of perception afforded by Figure 10.3 (or any other complex model), the processes that contribute to the observer’s response may be considered. Through such an analysis, the interpretation of the observer’s perception becomes more meaningful and possibly more accurate. One example of this type of analysis is the Theory of Signal Detection (TSD) which has documented the interaction between perceptual and judgmental processes (see Green and Swets 1966; Swets 1973).

Assessment and the interpretation of assessment can be enhanced through the use of methods that are designed to accommodate the complexities of the perceptual response. In Model 1, (Fig. 10.2) if we accept the more complex view, the rating of “4” may be directly due to the landscape presented or due to intervening processes. To best understand the observer’s response both alternatives should be considered.

### Perceptual Preference Assessment

The goal of perceptual preference assessment is to measure environmental quality judgments more directly. Typically, participants are shown a landscape, either in person or simulated with a photograph or videotape, and asked to indicate the extent of their preference for that landscape. The landscape quality assessment may thus be more directly related to the actual features of the landscape than when only verbal surveys are used. For example, Petrich (1979) asked his subjects to provide landscape quality judgments for slides of proposed power plant sites with and without superimposed power and cement plants and air pollution. Subjects’ ratings could therefore be meaningfully compared across three types of land use (cement plant, power plant, undeveloped) all with or without air pollution that might result from development (see Figure 10.4). Rather than relying on verbal descriptions or memory for a site, the perceptual preference method measures
the perceptual response to the environment and relates it to a specific environmental feature.

One critical issue in perceptual preference assessment is the manner in which the landscape is presented. Ideally, perceptions are assessed while the observer is in the environment in question. However, this is not always practical, and, in the case of proposed future development, not always possible. Thus the environment is often represented with photographs, drawings, or models. Of these, the color slide is most commonly used. Several studies have examined the validity of the color slide representation of landscapes (Daniel and Boster 1976; Seaton and Collins 1972), and it appears that color slides provide good surrogates for landscapes, especially if the landscapes are relatively homogeneous. For example, Petrich (1979) successfully assessed two reasonably similar landscapes in the same geographic region with forty photographs. Responses to radically different sites can prove difficult to interpret, particularly if those sites are not well represented photographically. Daniel and Ittelson (1981) point out that responses to very diverse environments reflect stereotypical reactions to the symbolic environment rather than a perceptual response based on specific landscape characteristics. If perceptual responses are desired, environmental representations should be restricted to a reasonable range of environments and those environments should be sampled well. The validity of pictorial representation generally increases with the number of photographs or drawings used to represent an environment.

**EXAMPLE 3: The Measurement of the Observer’s Response**

Example 2 discussed some aspects of theoretical psychophysics and the implications of simple versus more complex representations of the observer’s response. If a complex representation of perception is accepted, analytic procedures and methods that are sensitive to these complexities should be employed. Methodological psychophysics is devoted to the development and refinement of such procedures, usually called scaling methods. Methodological psychophysics focuses on three distinct tasks: scaling of observers, scaling of objects or stimuli, and the scaling of both observers and stimuli. In the first case, interest lies in differences among people. The second case is concerned with the measurement of the physical world through the judgments of observers. This case is most relevant to landscape assessment where the interest is most often in the measurement of perceptions of landscape qualities. Thus, methods that separate assessment differences into observer differences and landscape quality differences are needed. Once this separation has been performed, observer differences may be held constant to obtain a “pure” measure of differences in landscape qualities.

This process of separation commences with the statistical standardization of each observer’s internal measurement system. When the attitudes, beliefs, judgments, perceptions, or aesthetic preferences of the observer are assessed, an implicit assumption is that some continuum exists within the observers. Thus the human observer is viewed as a measuring instrument with some internal “yardstick” of measurement. Unfortunately, we do not have direct ac-
cess to the observer’s “yardstick” and, therefore, must attempt to infer or scale each observer’s measurement. This is especially important in that other cognitive processes (for example, attention to the perceptual task or judgmental criteria) may influence the observer’s use of his or her “yardstick.” A simple scheme of the observer as a measurement instrument might be illustrated as shown in Figure 10.5.

This scheme conforms to the example used earlier in which an observer has rated the scenic value of a landscape view. In this example, Observer A had responded to the slide presentation by marking a rating of “4.” Observer B has rated the same landscape view as “2.” A direct interpretation of this example would be that Observer A perceived the slide as possessing greater scenic value than Observer B (actually, a 4 on Observer A’s scale is quite similar to a 2 on Observer B’s scale). But, as discussed earlier, neither rating can be interpreted as a direct representation of the perception of the observers (as in Figure 10.2) because the ratings are influenced by cognitive processes which are likely to differ from one observer to the next. The task of both observers, however, is to translate their own “yardsticks” into ratings on the one to five scale. Although both observers use the same rating scale, each may use it in a different way. Thus the rating obtained from the observer is a representation of both perception of the landscape and the observer’s translation of his or her own internal “yardstick” to the response method.

The purpose of stimulus-centered (that is, landscape-centered) scaling methods is to equate or standardize the “yardsticks” of several observers. Following this process of standardization, remaining response differences among landscapes are interpreted as differences in the perception of landscape qualities. Example 4 will present an example of this approach to the analysis of the observer’s response, the Scenic Beauty Estimation Method (Daniel and Boster 1976).

A second major issue in perception preference assessment is the measurement of the response of the participant. Many different methods have been used. Daniel and Boster (1976) used a categorical rating scale for their Scenic Beauty Estimation Method. Participants were asked to rate landscapes represented by color slides on a scale from 1 to 10, where 1 indicated low scenic quality and 10 high scenic quality. Others have used rank order (Buhoyff, Wellman, Harvey, and Fraser 1978), paired comparison (Buhoyff and Wellman 1979), magnitude estimation (Buhoyff, Wellman, and Daniel 1982), and Q-sort procedures (Pitt and Zube 1979). Each of these methods has characteristic advantages and disadvantages (Daniel and Boster 1976; Torgerson 1958; also see Example 1). Consideration of such issues should guide method selection since empirical comparison of several of these techniques has shown them to be comparable (Buhoyff et al. 1982).

Treatment of preference data can be complex and often involves fairly sophisticated statistical treatment (see Example 4). However, perceptual preference data are relatively easy to collect, and, if properly gathered, can provide a link between the perceiver and the environment that is often more direct and interpretable than that obtained by either survey or behavioral observation methods.
EXAMPLE 4: Analyzing the Observer's Response; An Example of Scaling Using the SBE Model

The observer’s response is not necessarily a simple or direct function of visible landscape properties. Rather, the response of the observer is determined by at least several cognitive factors that are involved in the perceptual process, and direct interpretation of the observer's response can be quite misleading. This problem can be minimized through the use of several methods for the scaling and analysis of observer responses (see Baird and Noma 1978; Bock and Jones 1968; Torgerson 1958). One method that has been developed specifically for the analysis of perceptual preference of landscapes is the Scenic Beauty Estimation (SBE) model (Daniel and Boster 1976).

The SBE model assumes that categorical ratings assigned to a landscape scene by the observer are determined by two factors: the "true" perceived scenic beauty of the landscape and the judgmental

**TABLE 10.1. SBE Method Frequencies (f), Cumulative Frequencies (cf), Cumulative Probabilities (cp), z Scores (z), and the Mean z Score (z) for Each of Three Observers Rating Three Landscapes**

<table>
<thead>
<tr>
<th>LANDSCAPE I</th>
<th>LANDSCAPE II</th>
<th>LANDSCAPE III</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OBSERVER A</strong></td>
<td><strong>OBSERVER B</strong></td>
<td><strong>OBSERVER C</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rating</th>
<th>f</th>
<th>cf</th>
<th>cp</th>
<th>z</th>
<th>Rating</th>
<th>f</th>
<th>cf</th>
<th>cp</th>
<th>z</th>
<th>Rating</th>
<th>f</th>
<th>cf</th>
<th>cp</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>20</td>
<td>1.00</td>
<td>--</td>
<td>6</td>
<td>20</td>
<td>1.00</td>
<td>--</td>
<td>6</td>
<td>20</td>
<td>1.00</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>14</td>
<td>.70</td>
<td>.52</td>
<td>8</td>
<td>14</td>
<td>.70</td>
<td>.52</td>
<td>8</td>
<td>14</td>
<td>.70</td>
<td>.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
<td>.30</td>
<td>.52</td>
<td>3</td>
<td>6</td>
<td>.30</td>
<td>.52</td>
<td>3</td>
<td>6</td>
<td>.30</td>
<td>.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>3</td>
<td>.15</td>
<td>-1.04</td>
<td>2</td>
<td>3</td>
<td>.15</td>
<td>-1.04</td>
<td>2</td>
<td>3</td>
<td>.15</td>
<td>-1.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>.05</td>
<td>-1.64</td>
<td>1</td>
<td>1</td>
<td>.05</td>
<td>-1.64</td>
<td>1</td>
<td>1</td>
<td>.05</td>
<td>-1.64</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ Z = -2.68 \]  
\[ Z = -2.68 \]  
\[ Z = -2.68 \]  
\[ Z = -1.67 \]  
\[ Z = -1.67 \]  
\[ Z = -1.67 \]  

**SBE = [\[-.67,-(.67)]**  
\[ \times 100 = 0 \]  

<table>
<thead>
<tr>
<th>Rating</th>
<th>f</th>
<th>cf</th>
<th>cp</th>
<th>z</th>
<th>Rating</th>
<th>f</th>
<th>cf</th>
<th>cp</th>
<th>z</th>
<th>Rating</th>
<th>f</th>
<th>cf</th>
<th>cp</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>20</td>
<td>1.00</td>
<td>--</td>
<td>10</td>
<td>20</td>
<td>1.00</td>
<td>--</td>
<td>9</td>
<td>20</td>
<td>1.00</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>8</td>
<td>.40</td>
<td>- .25</td>
<td>8</td>
<td>10</td>
<td>.50</td>
<td>0.00</td>
<td>10</td>
<td>11</td>
<td>.55</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>.10</td>
<td>-1.28</td>
<td>1</td>
<td>2</td>
<td>.10</td>
<td>-1.28</td>
<td>1</td>
<td>1</td>
<td>.05</td>
<td>-1.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>.05</td>
<td>-1.64</td>
<td>1</td>
<td>1</td>
<td>.05</td>
<td>-1.64</td>
<td>0</td>
<td>0</td>
<td>0.00</td>
<td>-1.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.00</td>
<td>-1.96</td>
<td>0</td>
<td>0</td>
<td>0.00</td>
<td>-1.96</td>
<td>0</td>
<td>0</td>
<td>0.00</td>
<td>-1.96</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ Z = -5.13 \]  
\[ Z = -4.88 \]  
\[ Z = -5.43 \]  
\[ Z = -1.29 \]  
\[ Z = -1.22 \]  
\[ Z = -1.36 \]  

**SBE = [-1.28,(-1.28)]**  
\[ \times 100 = 0 \]  

<table>
<thead>
<tr>
<th>Rating</th>
<th>f</th>
<th>cf</th>
<th>cp</th>
<th>z</th>
<th>Rating</th>
<th>f</th>
<th>cf</th>
<th>cp</th>
<th>z</th>
<th>Rating</th>
<th>f</th>
<th>cf</th>
<th>cp</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>20</td>
<td>1.00</td>
<td>--</td>
<td>0</td>
<td>20</td>
<td>1.00</td>
<td>--</td>
<td>0</td>
<td>20</td>
<td>1.00</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>20</td>
<td>1.00</td>
<td>1.96</td>
<td>1</td>
<td>20</td>
<td>1.00</td>
<td>1.96</td>
<td>0</td>
<td>20</td>
<td>1.00</td>
<td>1.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>19</td>
<td>.95</td>
<td>1.64</td>
<td>1</td>
<td>19</td>
<td>.95</td>
<td>1.64</td>
<td>0</td>
<td>20</td>
<td>1.00</td>
<td>1.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>18</td>
<td>.90</td>
<td>1.28</td>
<td>8</td>
<td>18</td>
<td>.90</td>
<td>1.28</td>
<td>9</td>
<td>20</td>
<td>1.00</td>
<td>1.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
<td>.30</td>
<td>- .52</td>
<td>10</td>
<td>10</td>
<td>.50</td>
<td>0.00</td>
<td>11</td>
<td>11</td>
<td>.55</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ Z = 4.36 \]  
\[ Z = 4.88 \]  
\[ Z = 6.01 \]  
\[ Z = 1.09 \]  
\[ Z = 1.22 \]  
\[ Z = 1.50 \]  

**SBE = [1.09,(1.09)]**  
\[ \times 100 = 0 \]  

**NOTE:** There will always be a cp=1.00 at the lowest rating category (1), so no information is added by a z-score entry identical for all observers. The Z is therefore based on one less than the total number of rating categories (i.e., 5-1=4). In all other cases where cp=1.00 or 0.00 (z=0), the convention is followed that cp=1/2(n) or cp=1/2(n), respectively (see Bock and Jones, 1968). In the example above, N=20, therefore cp=1/40 or 1/40, yielding z scores of 1.96 and -1.96, respectively.
criteria applied by the observer. Further, it assumes that the scenic beauty of the landscape is not a single value, but is best represented by a distribution of values such as that which might result from the rating of several scenes from a particular landscape. Simple inspection of a distribution of ratings will not reveal which of the two factors ("true" versus judgment) are responsible for differences in observer ratings. Comparison of the distribution of ratings for several different landscapes, however, allows the distinction of actual differences in perceived scenic beauty from differences in the judgmental criteria of the observer. True perceptual differences will be manifested by consistent differences in the ratings assigned to the several landscapes by the same observer. Differences in ratings that are due to the observer's judgmental criteria will result in distributions of ratings that are overlapping or equal for the landscapes. The purpose of the SBE model is to eliminate the ambiguity introduced by differences in the observer's judgmental criteria. The resulting Scenic Beauty Estimates (SBEs) provide a quantitative index of the perceived scenic beauty of the landscapes.

Table 10.1 illustrates the use of the SBE method in the analysis of the hypothetical perceptual preference ratings of three observers for three different landscapes. Each observer has rated a number of scenes of each landscape on a five-point categorical rating scale. The frequency (f), cumulative frequency (cf), and cumulative probability (cp) values are shown for each observer and landscape. The z values are the standard normal deviates associated with each of the cumulative probability values. Mean z values are also shown.

The SBE analysis compares the observer's ratings distribution for one landscape with the distributions for several other landscapes. The resulting SBEs represent a more interpretable index of the perceived scenic beauty of the landscapes than the original ratings themselves. It can be seen in Figure 10.6, in fact, that direct interpretation of the "un-scaled" ratings (in this case, mean average ratings for each landscape) might result in substantially different conclusions regarding perceived scenic beauty than interpretation of the SBEs.

**Behavioral Measures**

A problem with both survey and perceptual preference assessments is that the link between preference, or attitudes and opinions, and actual behavior in the landscape may not be explicitly evaluated. Observing or measuring behavior in the target setting, although not feasible in many cases, is a useful means for obtaining this information. Behavioral measures are a broad class of methods for directly observing and measuring human activities which have many advantages, the most obvious of which is their face validity. Recording people's behavior provides a direct indicator of human activities in a particular environment.

A missing element in behavioral approaches, however, is the reason for the behavior. For example, in investigating the suitability of various sites for nuclear power plants one might assume, as did Petrich (1979), that the more scenic a site is, the less appropriate it might be for industrial development. One might then assess the attractiveness of several sites indirectly by measuring and comparing the number of visitors to each site. However, several factors, such as the proximity of a site or its accessibility, might mitigate the decision to visit a particular site. These intervening causes are always a threat to the conclusions that can be reached because measures of behavior, whether a rating on a 10-point scale, or the number of visits to a site each year, are always indirect indices of perceived landscape quality. One way to compensate for this disadvantage would be to combine behavioral observation with preference or survey data.

Another disadvantage of behavioral observation methods is that the behavior of nonusers or nonparticipants is not recorded. Although this may seem a trivial point, consideration of an example will clarify its importance. If visitation rates are used to assess visitors' perception of scenic quality, no information will be gained about those who opted not to visit a particular site. The reasons for not visiting an area are as important as reasons...
for visiting and could be critical in planning or managing those areas. Finally, behavior may not be observed at projects which have not yet been constructed.

**ASSESSMENT DESIGNS**

The assessment methods discussed above represent alternative approaches to the collection of information about landscape quality. These methods differ both procedurally and in their assessment focus (that is, what is measured). We can also contrast different approaches to landscape assessment on the basis of the design of the assessment process. The assessment design is the plan that specifies what is to be done and how it will be accomplished. Some assessment designs are characterized by formal and explicit “blueprints” that embody generally accepted requirements for assessment. Others are less formal. With varying degrees of success, any may serve as the basis from which to organize and draw inferences from data collected via the three assessment methods presented in the previous section.

We will discuss four general types of assessment designs: the case study, the experiment, the correlational design, and the quasi-experiment. These four categories of assessment design reflect substantive differences in the philosophy, logic, methodology, and purpose of assessment. However, clear distinctions among the categories may not always be present or readily apparent in practice. The categories of designs will be contrasted primarily on the basis of three criteria. First, the goal of a particular design may be either the description, prediction, or explanation of the assessment outcome. Second, the categories of assessment designs will be examined in terms of the degree of control that is exerted over the assessment process. Third, the utility and difficulty of interpretation of the assessment outcome will be considered for each type of design.

The choice of an appropriate assessment design is complicated by numerous competing considerations. A complete discussion of the advantages and disadvantages of particular designs is far beyond the scope of this chapter. The reader is therefore referred to excellent source books on assessment and design by Babbie (1979), Cook and Campbell (1979), and Kerlinger (1973). It should also be emphasized that no design fulfills the needs of all assessments. The ultimate evaluation of the effectiveness of an assessment design is its concordance with the constraints, needs, and purposes of assessment within the unique context in which the particular assessment occurs.

**Case Study**

A case study involves the assessment of a single environment or environmental problem, and often emphasizes the uniqueness of a site rather than features it may have in common with other sites. The goal of this type of assessment is description, and the optimal case study will combine public preference data with assessments of various physical, cultural, or cognitive characteristics of the landscape. For example, Petrich (1979) examined not only preference data, but also historical and cultural values of the landscape, and scenic features such as visibility and land-use diversity. Petrich examined two sites, however, and therefore his study is most correctly viewed as two related case studies.

Of the four designs presented in this chapter, the case study is most frequently used for assessments of perceived landscape quality, and examples abound. Daniel, Anderson, Schroeder, and Wheeler (1977) created computer-generated maps of relative perceived scenic quality in an Arizona forest. Schroeder and Daniel (1980) investigated the perceived scenic quality of several road corridors, and Zube et al. (1974) examined the aesthetic quality of the Connecticut River Valley.

Compared with other assessment designs, however, the case study method suffers from several characteristic disadvantages. Because case study data essentially describe a single geographic area, comparisons with and generalizations to other environments are highly restricted. Petrich’s (1979) pair of case studies of nuclear plant sites are comparable but generalizations to other sites, or even to other environmental development problems, would be limited.

The case study also suffers from a lack of control over the assessment process and can essentially be characterized as an anecdotal method wherein only a single site or environment is assessed. Further, the choice of the site and the variables to be measured is not accomplished systematically. Neither the variables of interest nor extraneous variables are subject to control by
the investigator. Thus inferences from or interpretations of case study assessments are threatened by uncontrolled and competing explanations (extraneous variables). The report of the case study may also be somewhat anecdotal and it is often not possible to determine from case study results exactly how information was collected.

Thus inference from case study designs is limited by the inability to control variables or sample sites and/or observers. For example, one section of Petrich's (1979) evaluation of nuclear plant sites provided only two site options, both of which received high scenic quality ratings and very low ratings of suitability as sites for any power plant development. Petrich concluded that neither site should be used. It is uncertain, however, whether Petrich's conclusion was warranted. Since there was no standard of comparison to help interpret survey results, high or low ratings might have been due to attitudes toward power plants, wording of questions, or other unknown factors. It is also uncertain how these results can be generalized to other settings.

It is also difficult to evaluate the effectiveness (for example, reliability, validity, and so forth) of the assessment, since the case study typically makes only singular measures of a particular site. For example, a scenic beauty map of a forest (for example, Daniel et al. 1977) will indicate which parts of an area are attractive (description), but not why (explanation) or whether they will remain so in the future (prediction).

The case study method does have some advantages, however. It is often less expensive of time and resources than other methods. Case studies are appropriate when a single site or problem is of consequence; they are also very useful for exploratory or pilot studies. Finally, the usefulness of the case study method can be vastly improved through the standardization of measures, sampling, and definition of variables so that comparisons can be made across separate sites and studies.

**Experimental Designs**

The general logic of the classical approach to experimentation is well known. Preferences for a site, for example, might be compared before and after some management action or treatment had taken place. In this example, physical change in the environment is controlled by the experimenter. This treatment is therefore called an independent variable. Some measure of the observer's perceptions or judgments is used as an indicator of the effects of that treatment and is called the dependent variable. The dependent variable is not manipulated by the experimenter; it is only measured. It might also be useful to consider or identify other variables that might influence the experiment, such as environmental attitudes, differences in the observers' cultural or geographic backgrounds, and so forth. Such variables are termed extraneous variables.

Experimental designs have three identifying characteristics. First, the emphasis is on explanation rather than description. The purpose of an experiment is to provide evidence that changes in the dependent variables can be explained as a function of the changes induced in the independent variable by the experimenter. This kind of explanatory evidence supports inferences about the cause of events.

A second identifying characteristic of the experiment is the ability to exert control over the assessment process. Experimental control is manifested in three ways. First, the dependent variables to be measured are explicitly chosen and defined by the experimenter. The level or occurrence of the independent variable is controlled by the experimenter, and extraneous variables are often controlled to ensure that their effects are not confounded with those of the independent variable.

Experimental control is also manifested in the assignment of sampling units to groups or conditions. For example, in the traditional experiment consisting of an experimental group and a control group, the assignment of experimental subjects to each group is completely controlled by the experimenter. Two types of control of assignment are most common: matching and random assignment. Matching of experimental groups involves balancing the composition of the groups such that, prior to the experiment, the groups are equal or comparable. Though this procedure seems logically satisfactory, in practice it is somewhat costly. Furthermore, it is seldom certain that all relevant characteristics of the group members have been identified and matched. A more acceptable procedure is random assignment, which results in groups whose composition is solely due to chance. This procedure supports much of the rationale of experimentation.
and greatly enhances the interpretation of experimental results.

The third type of control manifested in the experiment involves the ability to control the experimental setting or the surrounding conditions of the experiment. At one extreme, this type of control is exemplified by the laboratory experiment, where all conditions surrounding the experiment are held constant. Controlled conditions minimize the risk that an unknown extraneous variable may influence experimental results.

A third identifying characteristic of the experiment is complete specification of methods and procedures before observation has begun. Specification entails the definition of the experimental variables, statement of hypotheses predicting the outcomes of the experiment, and precise description of the exact methods and procedures to be used in the experiment. This process, known as operationalization, must fulfill the requirement that any independent experimenter could duplicate the experiment in all relevant details.

In our example of the assessment of sites for a nuclear power plant, Petrich (1979) manipulated the type of site development and its effects by superimposing cement plants, nuclear plants, and air pollution on photographs of the two geographic areas under consideration (see Figure 10.7). Observers examined scenes of the same geographic areas with and without development and air pollution. Because of this control over the independent variable (kind of development superimposed), Petrich was able to interpret observers’ perceptions of the two sites by comparing reactions to the various developments. More importantly, he could conclude that superimposed air pollution or development provided causal evidence of a decline in scenic preference because those were the only variables manipulated; extraneous variation was eliminated through experimental control.

Thus the experiment is characterized by formal requirements for operationization, a great degree of control over the assessment process, and a focus on explanation and causal inference. The outcome of the well-designed experiment is particularly amenable to interpretation because competing interpretations are eliminated by experimental control. A potential difficulty of the experiment is its “artificial” nature. It may be questionable whether the results of a particular experiment can be generalized to other samples of people, environments, or variables. Also, formal experiments are often costly or impossible to conduct in the natural setting or field environment, and the requirements for experimental control are often

quite intrusive in the field setting. Therefore, although the experiment may yield the most precise and interpretable assessment information, it may be very difficult to generalize from the artificiality of the experiment to the setting of interest.

**Correlational Methods**

The goals of correlation methods are description, explanation, and prediction. In contrast to experimental designs, correlational methods institute statistical, rather than experimental, control over the variables of interest. Typically a set of candidate variables are identified *a priori* and measured as they vary naturally. These independent variables are then systematically associated with a dependent, or criterion variable, typically preference, via simple or multiple regression procedures. The result of this process is an equation predicting preference from levels of a number of environmental characteristics. Additionally, regression procedures provide a measure of the strength of the relationship between preference and the independent variables. Tabachnick and Fidell (1983) provide a readable and useful description of regression techniques and available computerized statistical packages for their calculation. A more rigorous treatment of multivariate techniques may be found in Pedhaezur (1982).

The human or psychological half of the equation is usually measured with perceptual preference techniques. Generally, a numeric value is assigned to a landscape via a rating scale, rank order, paired comparison, magnitude estimation, or Q-sort assessment method (see Example 1). Behavioral or survey measures could also serve as dependent variables as long as they constituted a response to a specific environment or scene in an environment and may be quantified in some way.

The environmental half of the equation is composed of any number of characteristics which range from objective to subjective, depending on the goals of the assessment. Objective measures are generally made either onsite or from photographs. For example, Latimer, Hego, and Daniel (1981) and Malm, Kelley, Molnar, and Daniel (1981) associated high levels of particulate air pollution with low preference for western canyons. Arthur (1977) and Daniel and Schroeder (1979) measured physical characteristics of forests onsite and found that large trees, large amounts of ground cover and shrubs, and small amounts of downed wood and small trees were associated with high perceived scenic quality. Bulloch and Wellman (1980) measured characteristics of scenic vistas from photographs and developed the following prediction model:

\[
\text{Landscape preference} = \\
10.83 \text{(area in sharp mountains)} \\
-0.59 \text{(area in sharp mountains)}^2 \\
1.57 \text{(area in distant forest)} \\
-8.60 \text{(middle ground area of insect-damaged trees)} \\
-64.59 \text{(proportion of forested area)} \\
0.97 \text{(area in flat topography)}
\]

The numbers preceding each variable in parentheses are beta coefficients for the regression equation and represent the relative contribution of each variable in predicting the criterion variable, landscape preference. Thus, although sharp mountains and distant forests are both positively associated with landscape preference, sharp mountains are roughly eleven times more important as predictors. In this model, sharp mountains were associated with landscape preference in a curvilinear function. That is, sharp mountains are preferred only up to a certain point: if too much of the photograph is taken up with sharp mountain peaks (the squared second term of the equation), their contribution is negative.

Subjective measures of environmental characteristics are generally obtained from judges' ratings of photographs. For example, Vining, Schroeder, and Daniel (1984) used ratings of subjective characteristics of homes and homestites (for example, design congruity, fitfulness, obtrusiveness) to develop models of preference for forested residential sites. R. Kaplan (1975), S. Kaplan (1975), and Ulrich (1977) found that mystery, a subjective or psychological property of the environment, predicted preference well. Other examples of subjective environmental characteristics are offered by Daniel and Vining (1983) and Zube, Sell, and Taylor (1982).

Subjective measures of landscape characteristics suffer two important disadvantages. First, since they are evaluated by independent judges, the agreement among those judges, or reliability, is a critical element of their usefulness. If judges
do not agree in their assessments, the variables assessed are of questionable validity (Cook and Campbell 1979). Second, in project-level assessment, the applications of the analysis are more important than its theoretical value. The relationship between subjective properties of the environment such as mystery or fittingness and actual physical or biological features of the environment has not yet been determined. Thus specifying that an increase in mystery will increase preference provides little guidance to an environmental manager. Furthermore, it may be much more difficult to measure changes in these subjective determinants of environmental preference than to measure the preferences directly.

Correlational approaches have important advantages over the case study method and may, in some cases, be more useful or more practical than experimental or quasi-experimental designs. Although some correlational assessments may seem like case studies because they may be done in single environments, an important distinction must be made. In the correlational approach, preference in one environment is systematically and mathematically associated with the characteristics of the environment. For example, a correlational approach to Petrich’s (1979) nuclear plant site might have associated measures of vegetative characteristics, air pollution, or visual obstruction (objective), or of cultural and historical values (subjective) with perceived suitability of various sites. Assuming that drab vegetation, short site lines, air pollution, and low cultural values would be associated with the suitability of a site for nuclear plant development (or with low scenic preference), one might attempt to generalize these findings to other similar sites. Specifically, one would use the multiple-$R^2$ statistic, which indicates the amount of variance in preference ratings accounted for by the predictor variables, to determine the extent of the same relationship for other sites. Obviously, even a correlational approach would benefit by comparison across several sites, and this approach may be combined with experimental or quasi-experimental methods to generate even more predictive power.

The greatest difference between correlational and experimental methods is in the degree to which causal relationships (that is, prediction and explanation) may be specified. The explicit manipulation of predictor variables is absent in correlational approaches. In the above example, there is no way to know whether it is vegetation, or some correlate of vegetation, such as topography, that causes low scenic preference. Thus to a certain extent correlational designs are restricted to less powerful inference and interpretation than experimental designs. They are often much more practical, however, and are based on a different, but not necessarily inferior, tradition and logic (Brogden 1972; Cronbach 1957).

**Quasi-Experimental Designs**

The quasi-experiment, characterized by incomplete control of the assessment process, is often a good compromise between the case study and the experiment. As with the experiment, the foci are explanation and prediction. Quasi-experimental designs often attempt to apply experimental methods within the existing constraints of a natural setting, where random assignment to groups may not be possible. This inability to control the composition of experimental groups poses a serious difficulty in the interpretation of the experiment because results may be attributable either to the effects of the independent variable or to differences in the composition of the groups. A number of tactics of design and statistical control are available to minimize this difficulty and the reader is referred to Cook and Campbell (1979) for further details.

In addition to a lack of control over group assignment, the quasi-experiment seldom attains the precision of control over the experimental setting that is achieved in the experiment. As a result, interpretation of assessment outcomes is usually less precise. This problem, however, may be tempered by the utility of assessment information from the field setting and by the ability to circumvent problems of experimental control through statistical procedures of control. Furthermore, the choice of true versus quasi-experiment may not be at issue. It is often the case that the quasi-experiment is the only viable design available. As Weisberg (1979) has pointed out, the use of quasi-experimental methods may be a necessity, especially given flaws that may intrude on assessments initially planned as true experiments.

The quasi-experiment may be illustrated by considering a situation that might have arisen in Petrich’s study of the nuclear power plant sites had he chosen to study visitors to the two pro-
posed sites. This procedure would constitute a nonrandom sampling technique, but might be used because it better represents the most concerned groups, that is, actual visitors to the site. Although this group might be more likely to suffer the impact of a nuclear plant development, and thus might be the most appropriate group to sample, it might also be true that visitors to each site differ in some systematic manner. For example, visitors to a site farther from the city might be more affluent or more prone to antidevelopment attitudes than visitors to the second, closer site. Control of these extraneous variables is possible (as through covariance analysis), if they are recognized and measured. Thus a major issue in the use of quasi-experimental designs is the identification and measurement of extraneous, or uncontrolled, sources of variation so that statistical control is possible.

**CRITERIA FOR EVALUATING ASSESSMENT METHODS**

Extensive criteria are available for evaluating the effectiveness of assessment. Three primary criteria will be discussed here: reliability, validity, and sampling. The use of these criteria aid in the determination of the adequacy of assessment.

**Reliability**

Reliability is often termed the consistency or stability of measurement. When a particular event is measured, we want to be sure that the measuring instrument is consistent. If we obtain a different measurement each time we measure the same event, our instrument is not reliable. The measurement instrument must be sensitive enough to reflect changes in the event of interest, but insensitive to changes in irrelevant events.

Depending on the design of the particular assessment, we may wish to assess different aspects of reliability. For example, if the assessment depends on the judgments of two or more observers, we would want to establish the consistency among judges. This type of reliability is usually termed inter-observer reliability or between-observer agreement. Agreement can be calculated by determining the number of times the observers agreed divided by the total number of observer judgments. A second method of estimating inter-observer reliability is useful when observers are using a rating technique. In this case, the correlations of one observer's judgments or ratings to the second observer's ratings is calculated.

A second aspect of reliability involves the consistency of measurement over time. Usually, we want to ensure that the assessment was not influenced by the particular time of measurement. A second assessment, perhaps at a slightly different time of day or year, can be compared or correlated with the first assessment. A high consistency or correlation between measurements at the two different times demonstrates the temporal consistency of measurement. This type of reliability is traditionally termed test-retest reliability.

Thus the calculation of reliability allows us to check our measurements to ensure that unwanted or extraneous events have not influenced the obtained measurement. The demonstration of reliability enhances our confidence that the measurements obtained reflect actual changes in the event of interest.

**Validity**

This criterion denotes the degree to which the assessment method measures what was intended. Validity is the accuracy of measurement and can be evaluated in several ways. First, we would want general agreement that the method is appropriate. This type of evidence is called face validity: on the face of things the method appears accurate. Face validity may be determined by judges, but most often is determined by the general acceptance of the assessment method.

A second type of validity is termed content validity or the degree to which the substance, topic, or items of the assessment instrument reflect the purpose of measurement. For example, in the use of a survey, each survey item should reflect the purpose of the survey. As with face validity, content validity can be established by judges. Further evidence of content validity can be established through statistical procedures such as factor analysis.

A third demonstration of validity is obtained through the comparison of the assessment method with previously established methods or results. This form of cross-validation or conjoint validity is usually obtained by correlation. Agreement between the assessment method and another accepted independent measure of the event of in-
terest demonstrates that the assessment method also measures the event of interest.

Two important aspects of validity are the sensitivity and utility of the assessment method. Essentially, these aspects modify our previous definition of validity: the degree to which what is intended is measured for what purpose. The utility of an assessment method refers to the requirement that measurement produces useful information that can be applied. The sensitivity of the assessment method is the ability of measurement to detect important changes in the event measured. For example, a measurement system which designated landscapes as either pretty or ugly would be neither as sensitive nor as useful as one that provided finer distinctions or gradations between “prettiest and ugliest.”

An interplay exists between these different facets of validity and between validity and reliability. It is often the case that different assessment methods display different degrees of the types of validity. Examination of the costs and benefits of each potential assessment method is necessary for an informed choice of methods. For example, the method with highest overall validity may be rejected in favor of a method with lower overall validity but particularly high utility. Similarly, sacrifices in reliability may be necessary in order to use a method with high validity. It must be stressed that reliability is prerequisite to effective measurement; however, methods must ultimately be evaluated within the context of the entire milieu of measurement and its intended purpose.

Sampling

The third criterion for evaluation of assessment regards the way in which units (for example, people, landscapes, and so forth) are chosen for measurement. Because it is seldom feasible to measure every unit, some procedure must be used to choose measurement units. If human perception is the subject of assessment, then people as well as landscapes are likely to be chosen.

Two major kinds of procedures are available for the sampling of units: intuitive procedures and probability procedures. Intuitive procedures do not incorporate explicit rules for the selection of measurement units. Units are chosen on the basis of perceived appropriateness, informal logic or reasoning, or, quite commonly, on the basis of convenience. Such procedures are prone to error and can greatly inhibit the interpretability of assessment and the ability to generalize from results.

Probability procedures specify the likelihood or chance of any unit being chosen. The first step in these procedures is determining the domain or population of all units that might be chosen. Units are then selected from the domain by one of two major methods—random or stratified sampling. Random sampling operates on the principle that every unit has an equal chance of being chosen. For example, all units in the domain may be assigned an arbitrary number. Using a table of random numbers, a previously specified number of units are chosen from the domain. This procedure guards against unwanted influences on unit choice (for example, experimenter bias). Units are chosen by random chance and the choice of each unit is equally probable.

Stratified sampling adjusts the units chosen to match the domain on some characteristic of interest. For example, if the assessment measures attitudes about the environment, and if political party affiliation is known to be related to attitude, then the sample could be stratified on party affiliation. If there are 30 percent Republicans and 70 percent Democrats in the domain of interest, the units can be chosen so that the resulting sample contains a comparable percentage of Democrats and Republicans. Within this constraint, however, units are chosen randomly. The resulting sample would be more representative of the domain with respect to party affiliation. Whenever knowledge of a relationship of such a characteristic with the measure of interest is available, stratification provides an improvement over simple random sampling.

Often, assessment methods are only concerned with the sampling of observers or judges. It may also be necessary, however, to consider the sampling of other units such as environments, times of measurement, orientations of landscape views, and so forth. The use of precise sampling methods is directly related to the ability to interpret results and generalize results to other hypothetical assessment situations. Precise sampling requires the specification of the unit domain and the objective choice of units through explicit rules. It is often most cost efficient to restrict the size of the domain and accept the resulting limits on generalization. Use of these procedures usually allows
TABLE 10.2. Assessment Methods and Designs

<table>
<thead>
<tr>
<th></th>
<th>Case Study</th>
<th>Experiment</th>
<th>Correlational Design</th>
<th>Quasi-Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveys &amp; Questionnaires</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Perceptual Preference</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Behavioral Measures</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

the interpretation of results free from errors of selection and allows generalization of results to the entire domain.

SUMMARY AND CONCLUSIONS

This chapter has provided a brief overview of methods for landscape quality assessment. Many different approaches are used to gather landscape assessment information and it is beyond the scope of this chapter to discuss all these approaches in detail. Rather, we have considered three assessment methods and designs based on several distinctions. In practice, these distinctions may not be clear or obvious, and a given project may incorporate features of any of several different types of designs. For example, an experimental design would be poorly suited to a project with the goal of obtaining initial descriptive information characterizing perceptions of a particular landscape. Conversely, a project whose goal was to establish the causal relationship between the degree of air pollution and perceived air quality is not amenable to the case study approach. In either case, the ultimate utility of assessment information is closely related to the appropriateness and adequacy of the application of methods to the goals of the particular project.

As can be seen in Table 10.2, each assessment method can be incorporated into any of the assessment designs. In fact, considerable inferential power and validity may be gained if more than one assessment method is used in a particular design. The quality of an assessment is less a function of which cell of Table 10.2 is used than of the care with which the assessment is executed. From consideration of cost-effectiveness, the case study and experimental methods are least desirable. Though the case study may be inexpensive in several ways, the outcomes of assessment are difficult to interpret and generalize; thus little time is spent, but little is gained. On the other hand, the experiment is often quite difficult to implement in the field setting and may be intrusive. As a result, in most landscape quality assessment projects the most practical choice of assessment design is likely to be the quasi-experiment or the correlational study. These approaches can be matched to any of the three assessment methods to fulfill the particular requirements of the project.

Regardless of the assessment method or design used, the consumer of assessment information must be critical of the quality of that information. To this end, several evaluative criteria were described. Foremost among these criteria is the establishment of the reliability of measurement. If measurement is capricious, no confidence can be placed in an assessment. Given reliability, it must also be determined whether the particular assessment displays adequate validity for the purpose of the project. It is also crucial to determine the adequacy of sampling. How project results are generalized to other observers and settings depends on the sampling methods used. Few assessment projects entirely satisfy all evaluative criteria. Rather, it is usually the case that precise methodology is tempered by needs for project relevance, utility, and cost-effectiveness.

Finally, we have presented a conceptual basis for landscape quality assessment emphasizing the need for a user-based or observer-sensitive approach. In addition to obvious physical and biological characteristics, the landscape manifests the impacts of past human use and the goals and hopes of plans for the future. Explicit consideration of the perceptions, judgments, needs, and demands of the observing public can contribute to wiser resource use and more effective and intelligent planning of future landscapes.