POLICY CAPTURING WITH THE USE OF VISUAL STIMULI: A METHOD OF QUANTIFYING THE DETERMINANTS OF LANDSCAPE PREFERENCE

by

Dennis B. Propst

Dissertation submitted to the Graduate Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
in
Forestry

APPROVED:

G. J. Buhyoff, Co-chairman

J. W. Hamilton, Co-chairman

J. D. Wellman

J. W. Roggenbuck

W. A. Leuschner

May, 1979

Blacksburg, Virginia

ACKNOWLEDGMENTS

Foremost on my list of "thank-you's" are my parents, . Without their unbending moral and financial assistance, I would have never gotten this far in my educational career.

Equally as important has been the companionship of my wife, , who has stuck with me through some difficult financial and emotional times. Also, her graphical artistry has certainly enhanced this dissertation.

My mother-in-law, , typed nearly all of this manuscript. I cannot thank her enough for the extra hours of her time that she spent on such a laborious task.

Next, I would like to thank my committee, Drs. Gergory J. Buhyoff, John W. Hamilton, J. Douglas Wellman, Joseph W. Roggenbuck, and William A. Leuschner for their professional guidance and patience throughout my stay at VPI & SU.

Not every graduate student is fortunate enough to have committee co-chairmen of the caliber of Dr. Buhyoff and Dr. Hamilton. I greatly appreciate their willingness to take time from their busy schedules to help me with my problems. I will always be grateful to Dr. Buhyoff for his advice and moral support during my graduate student career. Dr. Buhyoff's research abilities make him an outstanding model for any student.

Finally, this study was funded in part by the U.S. Department of Interior through a contract with the USDI National Park Service, Southeastern Region, Atlanta, Georgia. The findings, opinions, and recommendations expressed herein are those of the author and not necessarily those of the U.S. Department of Interior.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

During the past 15 years there has been an upsurge of interest in scenic beauty assessment. In general such assessment deals with evaluating landscape beauty and measuring people's visual perceptions within natural environments. The increased attention given scenic beauty assessment has somewhat paralleled the environmental movement. Rising public concern for preserving the scenic quality of natural environments stems from the conception of landscape beauty as being as important as any of our other natural resources, and the realization that the visual resource can be destroyed. As of 1971 the federal government had funded 50 studies related to man-environment research, an area in which aesthetics is a predominant theme (Bagley et al. 1973). Many other studies related to scenic beauty assessment have been federally funded since then. In addition state, regional, city and private agencies have begun to consider aesthetics in their planning procedures (for examples pertaining to metropolitan and river basin planning see McHarg 1969).

There are important economic considerations in scenic beauty assessment as well. To accomplish management goals, resource managers face the hurdle of allocating scarce resources, such as land, labor, and capital. In order to compare the viability of projects which compete for such scarce resources the manager must evaluate the projects' relative costs and benefits (dollar and nondollar). Since the National Environmental Policy Act (U.S. Congress 1969, P.L. 91 -190) requires federal resource managers and planners to consider the aesthetic impacts of all decisions, they must also consider the costs and benefits associated with landscape manipulation.

Historical Perspective

The term "aesthetics" comes from the Greek verb aesthetikos meaning perceptible by the senses. Webster's Third New International Dictionary defines aesthetics as the philosophy and science of art; specifically, the science whose subject matter is the description and explanation of the arts, artistic phenomena, and aesthetic experience and includes the psychology, sociology, ethnology, and history of the arts and essentially related aspects. Evidence of consideration of the pleasure derived from natural and man-made works of art goes back to the ancient philosophers. For example, Aristotle felt that the fine arts were derived from man's desire to imitate nature and that the end result of the perception of fine arts was pleasure (Mure 1964). This is because

> in recognising the relation of copy to original there is a pleasure over and above the pleasure we take in the copy as simply an object beautiful to look upon. This is shown by the fact that the realistic representation even of an object in itself painful to behold gives us pleasure (Mure 1964, p. 225).

The intention here, however, is not to discuss the history of the philosophy of aesthetics, but to establish the origins and meanings of the word "aesthetics" and its relation to present-day scientific scrutiny.

The 18th Century German philosopher Alexander Baumgarten (1714-62) coined the term aesthetics conceiving of it as the science of sensory cognition. According to Dickie (1971),

> He (Baumgarten) thinks of art as a low-level means of cognition, that is, gaining knowledge. In short, Baumgarten conceives of art as falling under the domain of both the sensory faculty and the intellectual faculty as a mode of inferior cognition (p.10).

Baumgarten thus sought to study the acquisition of knowledge under the rubric of aesthetics versus the acquisition of knowledge through logic and reason. Nearly two centuries later, Santayana (1907) apparently agreeed with Baumgarten's concerns. Santayana states,

> beauty is a species of value...A first approach to a definition of beauty has therefore been made by the exclusion of all intellectual judgments, all judgments of matter of fact or of relation" (p. 20).

To Santayana, aesthetic judgments are judgments of value while intellectual judgments are judgments of fact.

Kant used the term aesthetics to include not only judgments of beauty but also judgments of pleasure in general.

> For Kant, all aesthetic judgments focus on pleasure, which is a property of the experiencing subject rather than of the objective world. Such judgments are <u>subjective</u> because pleasure does not play a role in the cognition of the objective world external to the subject.... But if judgments of beauty are subjective, Kant also thinks that they are stable and universal in a way that other pleasures are not. That is, he seeks a theory which will show that although the pleasure felt in the taste of, say, chocolate or anchovies is merely personal, the pleasure felt with beauty is universal and necessary" (Dickie 1971, p. 26-27).

However, Pickford (1972) emphasizes that

> One important point is that the psychology of aesthetics is not to be identified necessarily or only with the study of liking of pleasure-giving stimuli, objects or situations. Aesthetically valued objects, situations or stimuli may not always be those which give pleasure or most pleasure, even displeasing or ugly objects may be aesthetically good, and there are many pleasure-giving situations and objects which are not of aesthetic value or interest" (pp. 246-247).

Although Pickford is not referring to the beauty of natural landscapes, one may readily generalize and conclude that landscape scenes may be pleasure-giving without necessarily being composed of those elements (e.g., form, texture, composition) which make a work of art aesthetically valuable. Conversely, a landscape may contain several artistic elements and still be judged low in scenic quality.

Bagley et al. (1973) trace the historical development of the Western concepts of aesthetics and list several reasons for the rising emphasis recently placed upon such concerns. After 1950, suburban sprawl and central city development and subsequent decay were recognized as problems by professional planners. Therefore, among other factors (crime, crowding, etc.), urban planners and architects became more concerned with the aesthetic impacts of development and inner-city rehabilitation. In the 1960's, the Keep America Beautiful campaign of Lady Byrd Johnson and the use of the scientific method to study scenic beauty (e.g., Sonnenfeld 1966; Peterson 1967) were but two manifestations of the growing concern for the protection of the visual resource. Finally, Earth Day and the subsequent environmental movement rekindled an appreciation of nature in its raw state, and, consequently, the 1970's will probably be remembered as a time of preserving natural areas and reassessing the use of our natural resources.

Today, citizen pressure groups no longer accept development on cost efficiency arguments alone, but require the consideration of aesthetic issues as well. Federal legislation (discussed below) requires assessment of the aesthetic impacts of all federally funded projects.

Several major themes emerge from this historical discussion of aesthetics. First, aesthetics deals largely with individual perceptions of beauty and the acquisition of knowledge through such perceptions. Second, visual perception is a cognitive activity. Perception is not just a passive reception of data, for the mind requires sensory information in order to function properly. Third, everyone gains some sort of pleasure from the perception of beauty. Finally, the 20th Century conception of aesthetics includes not only human responses to art but also human responses to objects other than art works, including natural landscape scenery.

Aesthetics and Scientific Inquiry

People spend a great deal of time looking. This is not too surprising considering the fact that a significant proportion of all human information processed is acquired visually. There is also evidence that vision dominates the other sense modalities (Rock and Harris 1967). According to Newby (1971, p. 68), "...man is truly a visual animal with respect to his environment. He learns more, reacts more, and appreciates more through his visual system than through any other sense."

The scientific method is increasingly being used to study aesthetics. Bagley et al. (1973) summarize this notion well:

> From the 19th Century on, aesthetics has steadily progressed in the use of the scientific method. It has become more empirical, basing its conclusions on the observation of works of art and related phenomena, instead of deducing them from religious and metaphysical assumptions...Aesthetics makes constant use of terms, concepts, and data from psychology and from the social sciences (pp. 22-23).

Scenic beauty assessment is thus recognized as a legitimate area of scientific inquiry and especially crucial at a time when natural beauty values clash with technological development (Shafer and Tooby 1973). Scenic beauty, however, is one of the most difficult resources to measure objectively. This measurement problem is primarily a result of scenic beauty's being defined not only in terms of physical characteristics, but also in terms of human judgment. Also, the aesthetic experience is usually difficult to separate from other experiences (Arthur et al. 1977). As an example, aesthetic satisfaction is often highly correlated with recreation satisfaction (Arthur et al. 1977). Scenic attractions are often the objective or an important element of many recreation activities, such as driving for pleasure, and photography. In support of this hypothesis, there is some empirical evidence that scenic attractiveness is a component of total site attractiveness (Buhyoff 1979).

Legislative Requirements

Despite all of the conceptual problems related to scenic beauty measurement, legislation requires that the aesthetic impacts of resource planning and management decisions be evaluated. In general the National Environmental Policy Act (U.S. Congress 1969, P.L. 91-190) requires the "Federal Government to use all practicable means... to ...assure for all Americans safe, healthful, productive, and aesthetically and culturally pleasing surroundings..." Specifically, NEPA directs all federal agencies "to identify and develop methods and procedures...which will ensure that presently unquantified environmental amenities and values may be given appropriate consideration in decision-making along with economic and technical considerations." This section of the Act provides the basis for the development of methodologies for measuring and quantifying aesthetics (Bagley et al. 1973). In addition the Multiple Use and Sustained Yield Act (U.S. Congress 1960, P.L. 86-517) requires the U.S. Forest Service to consider the more intangible forest resources (aesthetics, wildlife, and recreation) as well as the more tangible ones (timber and forage) in the management of national forests. Likewise, the Federal Land Policy and Management Act (U.S. Congress 1976, P.L. 94-579) and the Forest and Rangeland Renewable Resources Planning Act (U.S.

Congress 1974, P.L. 93-378) mandate the protection of scenic quality and the quantitative measurement of scenic benefits.

The Forest Service has responded to these legislative initiatives by adding to their policy manual a section entitled "Landscape Management" in which landscape beauty is recognized as a basic resource to be considered equally with other forest resources. In line with this policy, training documents have been developed by the Forest Service for the purpose of illustrating the concepts, elements, principles, and applications of the agency's landscape management program (U.S. Department of Agriculture 1973, 1974).

Finally, the American courts have played a significant role in matters concerning the protection of scenic values. For example, in the Scenic Hudson case of 1966 (Scenic Hudson Conference v. Federal Power Commission), citizen groups attempted to block the construction of a pumped-storage plant on the Hudson River due to potential adverse effects on scenic values and fish populations (Sax 1970, p. 131). The court agreed that the Federal Power Commission had not adequately considered deleterious impacts on all the natural resources of the area (scenic beauty included) and returned the case to the Commission for further research. Few of the court decisions have ultimately rested upon protection of the visual resource; however, the negative aesthetic impacts of certain extractive practices (e.g., clearcutting, stripmining) and developmental procedures (e.g., high-rise construction, dredging of marshes) have drawn enough public attention to force judicial interpretations of the laws governing man's role in manipulating the environment. To illustrate, the recent Monongahela decision (Izaak Walton League v. Butz) began as a public outcry over the aesthetic impacts of clearcutting practices on the Monongahela and Bitterroot National forests. While the issue of clearcutting served only as a focal point for many groups interested in the overall reform of all forest management practices (Fairfax and Achterman 1977), degradation of scenic values brought the clearcutting issue enough national notoriety to result in a legal interpretation of the Forest Service's Organic Act.

The legislation pertaining to visual resource management poses two problems. First, there are no standards defining an acceptable level of aesthetic impact. Second, the acts suggest no methods which managers and planners may use to facilitate their decisionmaking process. Thus, managers and planners are forced to use their own professional judgment, a purely descriptive inventory system, or one of a multitude of "quantitative" procedures. Sole reliance on professional judgment could be ineffective because research has shown that managers and planners do not always know the desires and attitudes of their current and potential clientele (Lucas 1970; Clark et al. 1971; Peterson 1974; Buhyoff et al. 1978). Little is known about which procedures are the most reliable; there have been few attempts at replication.

Objectives

The purpose of the research described herein is primarily of a methodological nature. The aim is to examine the usefulness of a procedure known to industrial psychologists for approximately 18 years. This procedure, called "Policy Capturing" (Hoffman 1960; Bottenberg and Christal 1961; Christal 1968a) has never been employed in the study of scenic beauty. Specifically, the objectives guiding this research are twofold: (1) to determine if Policy Capturing is a an appropriate procedure for modeling the human judgment process regarding scenic beauty preferences, and (2) to identify and determine the relative importance of those landscape features which explain variations in scenic beauty preferences.

Chapter 2

REVIEW OF LITERATURE

Several comprehensive literature reviews pertaining to scenic beauty assessment have already been written (see Arthur et al. 1977; Craik 1971, 1972b; Fabos 1971; Riesenman 1977). It is not the author's intention to summarize these excellent reviews. Instead the purpose of this chapter is to relate past findings to the research methodology employed, Policy Capturing (PC). Since PC is basically a multiple linear regression procedure, discussion will be aimed primarily at those studies which used a similar type of analysis. In addition, a section on PC and its theoretical underpinnings will be presented.

Regression Analysis in Landscape Assessment

Peterson (1967) was interested in constructing models which identified and weighted the dimensions accounting for people's preference for the visual appearance of residential neighborhoods. To measure subjectively his independent and dependent variables, Peterson used a rating scale method based on Thurstone's Law of Comparative Judgment (Thurstone 1959). Peterson obtained the following two-variable model accounting for 79 percent of the variance in preference:

Preference = -0.563(Apparent Age of Neighborhood) +
0.464(Closeness to Nature)

Factor analysis with varimax rotation reduced the original nine independent variables (greenery, open space, age, expensiveness, safety, privacy, beauty, closeness to nature, quality of the photography) to four factors which accounted for 97.3 percent of the variance. When used as independent variables, the model became:

Preference = -0.723(Physical Quality of the
Neighborhood) + 0.587(Harmony with Nature) +
0.264(Noise) + 0.126(Quality of the
Photography).

This relationship yielded an R^2 of 0.95. The fourth factor generated by factor analysis was termed "Noise," not because of the absence of meaningful information, but because it was not directly interpretable in terms of the nine original variables (Peterson 1967).

Although the variables explained a great deal of variance and are intuitively meaningful, Peterson warns that the lack of a theoretical basis for explaining the relationships demonstrates only that the data fit the model. Other meaningful models might fit the data equally well. Finally, Peterson admits that a major weakness of his study was that the method employed failed to consider individual variations among subjects. That is, Peterson did not attempt to identify clusters of individuals sharing similar perceptual preferences.

Two years later Peterson and Neumann (1969) attempted to develop preference functions responsive to both individual differences and to visual characteristics of beach environments. Using a rank ordering of photographs as the dependent measure and semantic differential ratings of seven dimensions as values of the independent variables, Peterson and Neumann identified two groups with distinctive preference functions. The first group preferred scenic natural beaches, low use levels, and natural green vegetation ($R^2 = 0.988$). The second group preferred city swimming beaches, high quality sand and attractive buildings. While these findings suggest direct management implications in terms of how to satisfy different user groups, differences in preference between the two groups could not be adequately explained by the variables used. Peterson and Neumann conclude that the appropriate independent variables were not measured.

It was the work of Shafer and his colleagues (Shafer, Hamilton and Schmidt 1969), however, that has been the source of greatest interest in the literature pertaining to scenic beauty assessment. These researchers used grid analysis as a more objective measure of various landscape features than rating scales. Grid analysis simply involves placing a one-quarter inch grid transparency over a photograph and subsequently measuring various parameters of two dimensional space (e.g., number of cells, area, perimeter) accounted for by each pre-defined feature. Using area and perimeter measurements as values of the predictor variables and the rank orderings of random packets of photos as measures of the dependent variable, Shafer found six characteristics which accounted for 66 percent of the variation in landscape preference. His final model was:

$$Y = 184.8 - 0.5436X1 - 0.09298X2 + 0.002069(X1 \cdot X3) + 0.0005538(X1 \cdot X4) - 0.002596(X3 \cdot X5) + 0.001634(X2 \cdot X6) - 0.008441(X4 \cdot X6) - 0.0004431(X4 \cdot X5) + 0.0006666X1^2 + 0.0001327X5^2,$$

where Y = Preference
X1 = perimeter of immediate vegetation
X2 = perimeter of intermediate nonvegetation
X3 = perimeter of distant vegetation
X4 = area of intermediate vegetation
X5 = area of any kind of water
X6 = area of distant nonvegetation

In five of six field tests using Adirondack campers as subjects, the authors found high agreement between observed and predicted photo rankings, suggesting that the model may be applicable to a wide range of recreationists in the Northeast.

Several authors have criticized Shafer's work on methodological and analytical grounds. These criticisms can be summarized as:

1. The dependent variable was only an ordinal measure (Buhyoff and Leuschner 1978). Thus, one cannot tell by how much one landscape is preferred over another.

2. Some of the independent variables, such as perimeter of distant vegetation times area of water, lack practical meaning or intuitive appeal (Buhyoff and Leuschner 1978; Weinstein 1976; Kaplan 1975). That is, these variables have no theoretical foundation, or underlying reason, for being combined.

3. Shafer did not show that the photographs he used were representative of the landscapes in which they were taken (Arthur et al. 1977; Weinstein 1976).

4. With enough regressor variables (Shafer began with 44) a regression equation that will correlate perfectly with any dependent variable can be found (Weinstein 1976).

5. The variables in the prediction model were based on group averages. "Successful prediction of group averages does not necessarily imply successful prediction of individual reactions" (Weinstein 1976, p. 619). When preferences are averaged, measurement errors tend to cancel out and the predictive ability of the model increases.

6. Photographs are only abstractions of the real world and, hence, may not be valid indicators of actual landscapes (Weinstein 1976; Kreimer 1977).

7. Certain photographic quality variables (e.g., time of day, composition, scale of elements, type of lens, etc.) that Shafer did not consider may account for a substantial amount of variation in preference (Kreimer 1977).

8. Due to cognitive dissonance, campers sampled on-site will naturally prefer those landscapes nearest in type to the ones in which they are camping (Kreimer 1977).

Debate over some of these issues continues in the literature (e.g., see Carlson 1977; Shafer and Brush 1977; Buhyoff and Wellman, 1979a, 1979b). A partial explanation for this controversy is that the whole field of visual resource quantification is still in its exploratory stages and new methodologies are constantly being tested and evaluated. Since most of these issues pertain directly to the research described herein, they will be addressed separately in the chapters to follow.

For now, it is sufficient to say that one accepted way of assessing the reliability of a model is through replication (Weinstein 1976). In this respect it is important to note that Shafer's model predicted with equal success the overall preference patterns of recreationists in Utah (Shafer and Mietz 1970) and in Scotland (Shafer and Tooby 1973). While Weinstein and others pose valid criticisms, replication or near-replication has certainly strengthened the basis of support for Shafer's model being a reliable one.

Carls (1974), using a method very similar to Shafer's except for the addition of a few new variables, interviewed over 400 recreationists and members of the general population and obtained their landscape preference scores for 20 Illinois scenes. Factor analysis reduced Carls' 27 landscape variables to 9 orthogonal factors. The variable with the highest loading in each factor was selected to represent that factor as an independent variable in the multiple regression equation. Carls derived the following equation accounting for 48.1 percent of the variance in preference:

$$Y = 205.92 - 0.067X1 - 2.08X2 - 0.081X3 + 0.288X4 + 0.071X5,$$

where Y = estimated preference score
X1 = area of stream
X2 = area of waterfall
X3 = area of lake
X4 = area of people
X5 = area of high development

The area of water variables had a positive relationship with preference and accounted for about 18 percent of the variation; the effects of people and development had a negative relationship and accounted for a little over 20 percent of the variation (Carls' preference measure was scored so that a low Y value meant high preference).

Zube, Pitt, and Anderson (1974) were able to derive a 14-dimension regression model explaining 80.2 percent of the variance in the scenic quality of the Connecticut River Valley. Their best six dimensions[1] in terms of variance explained were:

---

[1] Grain Contrast = the difference in the size of the individual elements of adjacent land uses; Mean Slope = steepness of landform; Length of View = the maximum distance of view; Spatial Definition = the amount of enclosure created by landform; Area of View = size of the view area; Absolute Relative Relief = the range of vertical elevations, based on sample points, per unit area. These definitions are those of Zube et al. (1974).

	Per Cent Variance Explained
Grain Contrast	23.7
Mean Slope	16.9
Length of View	9.7
Spatial Definition	7.4
Area of View	6.5
Absolute Relative Relief	4.8
Total	69.0

Wohlwill (1968) was interested in the relationship between stimulus complexity and preference for landscape scenes. To obtain complexity values Wohlwill had subjects rate a set of slides for amount of variation along five attributes: color, shape, direction of dominant lines, texture, and artificial versus natural influences. The summation across these five attributes averaged over all judges served as values of complexity. For the dependent measure, subjects were asked to rate how much each slide was liked. Graphic (visual inspection of plotted data) and trend analyses (i.e., tests for linear, quadratic, and cubic trends possible in ANOVA designs) revealed a curvilinear relationship between complexity and preference. In other words, preference increased with increasing amount of complexity up to a point and then declined once complexity surpassed some optimal level.

Applying the theory of Signal Detection (Green and Swets 1966) and Thurstone (1927) scaling procedures, Daniel and Boster (1976) developed the Scenic Beauty Estimation

(SBE) method which produces interval level measures. These interval measures are called "scenic beauty estimates" and are abbreviated: SBE's. Using Daniel and Boster's method, Arthur (1977) obtained SBE's for six sites on the Coconino National Forest in north-central Arizona. Arthur then performed three regression analyses with three separate sets of independent variables. She regressed either physical landscape features, timber cruise variables, or design inventory parameters against the SBE's and found the following three models which explained 76, 80, and 97 percent of the variation in scenic beauty, respectively:

1) Physical Features: $SBE = -27.71 - 32.81P1 - 13.68P2 + 7.74P3 + 7.45P4 + 17.60P5 - 25.73P6 + 12.71P7$,

where P1 = amount of downed wood
P2 = distribution of downed wood
P3 = amount of vegetative ground cover
P4 = density of trees
P5 = size of trees
P6 = Variety of tree sizes
P7 = distribution of trees

2) Timber Cruise: $SBE = -21.92 - 18.78T1 + 8.60T2 -38.10T3 -45.51T4$,

where T1 = trees/acre, dbh (diameter breast height) < 13 centimeters
T2 = total basal area/ha, in square meters
T3 = cu. m/ha (based on ponderosa pine 14-29 cm dbh)
T4 = amount and distribution of slash

3) Design Inventory: $SBE = -20.08 -86.94D1 + 79.92D2 + 138.24D3 + 65.33D4 + 229.60D5 - 103.50D6 - 98.84D7$,

where D1 = lighting direction
D2 = clouds (none--billowing white)
D3 = contrast among discrete landscape elements (little--much)
D4 = surface variation (none--much)
D5 = crown canopy (little--prominent)
D6 = presence of detail (little--much)
D7 = visual vividness (little--very vivid)

The design inventory variables were rated on a 5-point scale by two independent groups of landscape architects.

Daniel et al. (1977) were able to inventory these key predictors and generate scenic beauty maps that have the interesting potential of being overlain with other resource (e.g., timber) maps.

Buhyoff and Riesenman (1979) obtained reliable interval scales of landscape preference using Thurstone's Method of Paired Comparisons (Guilford 1954), another psychophysical scaling routine. Buhyoff and Leuschner (1978) further analyzed these data to derive the "psychological disutility" functions of people viewing landscapes containing various amounts of insect (Southern Pine Beetle) damage. Using amount of insect damage as the sole predictor, they were able to derive the following nonlinear regression model which explained 84 percent of the variation in landscape preference:

$$\text{Preference} = 1.03 - 0.28 \ln(\text{Amount of Insect Damage})$$

The above model was found when subjects were told in the scaling instructions that they would be viewing scenes containing varying amounts of insect damage (experimental group). The control group was informed only that their task was to judge which of each pair of photographs they preferred. The corresponding model for the control group explained 33 percent of the variation in preference. Buhyoff and Leuschner interpreted the difference in R^2 between the experimental and control group models as evidence that the experimental group was cueing primarily on one dimension (amount of insect damage) whereas the control group was using other variables in a multidimensional fashion to determine their preferences.

Policy Capturing

Recall that two of the criticisms of the Shafer approach were the lack of a theoretical foundation and failure to account for individual and group differences. In fact, the absence of a theoretical foundation for using regression analysis is noticeable in all of the landscape preference studies just reviewed. The procedure outlined in this section is not necessarily intended to provide a landscape preference theory. However, this procedure does present a conceptual framework and justification for using multiple linear regression to study the relationships involved in landscape preference judgments. That is, PC yields a regression model, or strategy, representing the manner in which an individual or group weights separate bits of information to arrive at an overall decision; such an interpretation of the linear regression equation is based on empirically-supported psychological theory.

Emphasis on individual differences is another strong point of PC. In the landscape studies just reviewed, individual responses were aggregated and a composite model derived. The PC technique, however, involves computing a linear model for each individual and the combining individual models in a logical fashion (to be described later). In this manner, individual rater consistency and differences between individual rater strategies (weights) can be explored. Also, homogeneous clusters of raters with similar policies can be identified.

Theoretical Background

The technique of policy capturing has existed in industrial psychology for approximately 18 years. Hoffman (1960) and Bottenberg and Christal (1961) are usually cited as the pioneers of this analytic tool. PC has never been employed in the study of the scenic beauty of pictures but seems to hold much promise (Christal 1968a). Thus, the research reported herein is mainly a methodological study. As previously stated, the major aim is to see if PC is a useful and appropriate tool for developing predictive models and understanding the internal judgment processes related to scenic beauty preference.

Brunswik (1940) first proposed the use of multiple regression to model human judgment processes (Beach 1967). However, before 1960 there was little research on human information processing (Slovic and Lichtenstein 1971). Since 1960, there have been hundreds of studies dealing with human use of information in judgment and decision making (Slovic and Lichtenstein 1971). The judgmental paradigm relating individuals and stimuli is sometimes referred to as cue-utilization or policy capturing and is embedded in Brunswik's Lens Model (Wiggins 1973).

Figure 1 diagrammatically represents Slovic and Lichtenstein's (1971) explanation of the Lens Model. According to Slovic and Lichtenstein (p. 655) "Brunswik's main emphasis was not on the organism itself, but on the adaptive interrelationship between the organism and its environment." Brunswik thus studied the way in which humans learned the characteristics of their environment, not merely the degree to which humans used such cues. The Lens Model depicts the probabilistic relationships among organismic and environmental components of the judgment task.

The variables X_1, X_2, ..., X_k are cues or independent variables that define each stimulus object. Each of these variables has a specific amount of relevance to the true state (criterion value) of the world, Y_e. The relevance of the ith cue is shown by the correlation, r_{ie} , between each X_i

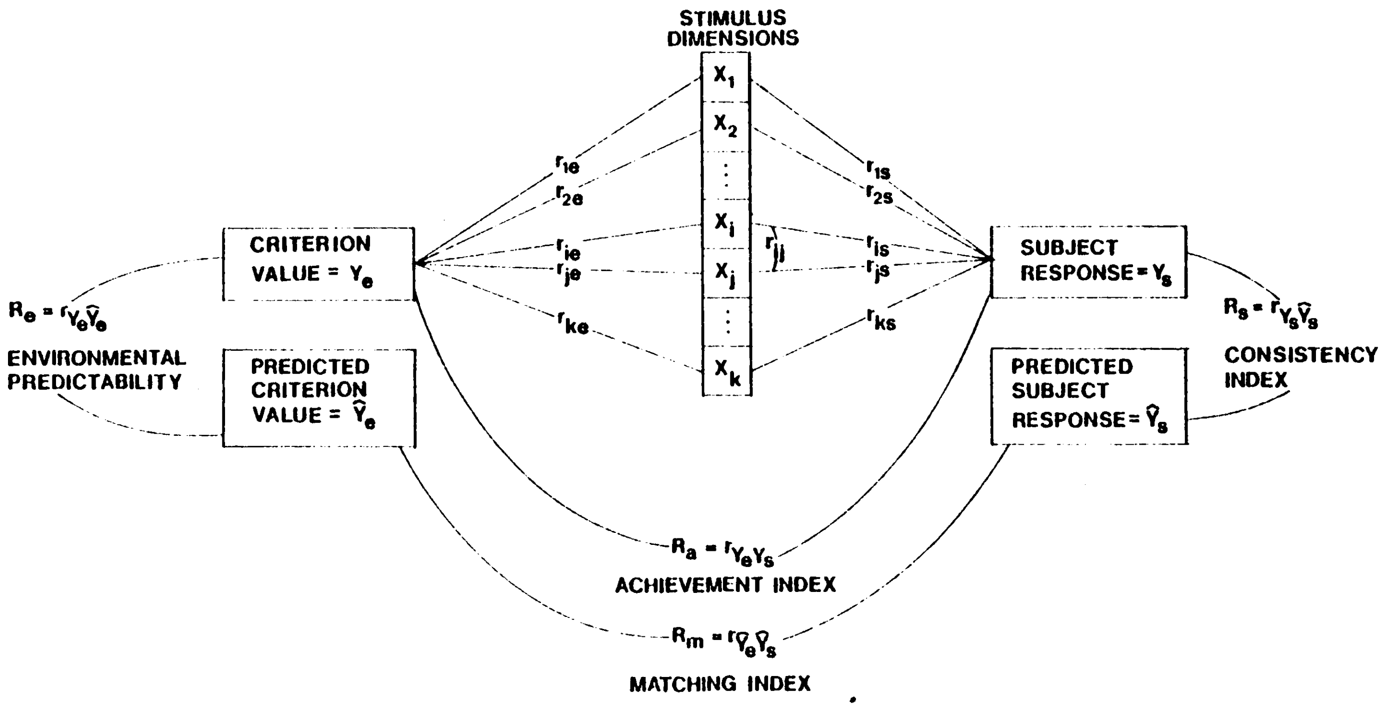

Figure 1. Diagram of Lens Model showing the relationships among the cues, criteria, and subjects' responses (from Dudycha and Naylor 1966 b).

and Y_e. The r_{ij} values represent the intercorrelations among dimensions. The subject's judgment or response is Y_s and the correlation between each subject's judgments and the X_i's is denoted by r_{is}.

Both the criterion value and the subject's response can be predicted from the following linear combination of dimensions:

$$\hat{Y}_e = \sum_{i=1}^{k} b_{ie} X_i \text{ ,} \qquad (1)$$

$$\hat{Y}_e = \sum_{i=1}^{k} b_{is} X_i . \qquad (2)$$

The optimal strategy is represented by Equation 1. In other words, the optimal means of predicting Y_e is to weight the cues as shown in (1). The multiple correlation coefficient, R_e , reveals the extent to which the weighted linear combination of cues is able to predict the criterion, Y_e.

Equation 2 depicts the subject's decisionmaking strategy or <u>policy</u>. The correlation coefficient, R_s , is a consistency index indicating the degree to which the subject reliably uses his or her mathematically weighted policy to reach a decision. The weights (b_{ie} and b_{is} values) are measures of the importance of of each cue in the environment and for the subject.

The values, R_a and R_m, are measures of the subject's performance. The correlation, R_a, denotes the extent of agreement between the criterion and the subject's responses over n observations; R_m represents the degree to which the policy equation of the subject (2) "matches" the environmental model (1).

The major difference between the Lens Model (Brunswik 1952) and the cue-utilization paradigm (Hoffman 1960; Naylor and Wherry 1965) is that the former places equal emphasis on the environmental factors influencing a judge's decision and his or her strategy. Policy capturing, however, is more interested in the judge's weighting process or policy (Equation 2; the right-hand side of Figure 1). In policy capturing, therefore, less emphasis is placed on modeling the environment and more emphasis is placed on controlling the environmental condition (Slovic and Lichtenstein 1971). Environmental control is achieved by making the stimulus dimensions apparent and varying their levels systematically, sometimes at the expense of realism.

Brunswik's Lens Model also serves as the basis for Social Judgment Theory (Hammond et al. 1975). According to Social Judgment Theory (SJT), humans constantly try to manipulate variables (experimentation is one example) in order to reduce ambiguity and thereby gain control of their environment. This notion pervades the social psychology literature as well (Langer 1975; Harvey and Smith 1977). For example, Kelley's (1967) attribution theory is based on the contention that the need for environmental mastery is the primary reason why individuals make causal attributions. A basic tenet of Festinger's (1957) cognitive dissonance theory is that persons attempt to control their world by engaging in rationalization, a tension-reducing process. Sometimes they succeed in gaining control, but when the appropriate variables cannot be manipulated, humans must use their own cognitive ability. In other words,

> They must do the best they can by passive rather than active means to arrive at a conclusion regarding a state of affairs clouded by causal ambiguity. They must, in short, exercise their judgment. (Hammond et al. 1975, p. 272).

Furthermore, this causal ambiguity must be represented in a theory of human judgment (Brunswik 1952, 1956; Hammond 1955). That is, for the sake of realism, the ambiguity of the relationships among the variables must be represented in the task used to study human judgment.

As noted above, Brunswik argued for the representation of the organismic and environmental systems in a symmetric manner. SJT is similar to the lens model in that SJT also includes a set of parallel concepts which apply to task (environmental) systems as well as human (organismic) systems (Hammond et al. 1975).

Hammond and his associates (1975) propose the following three concepts or objectives of SJT:

1. SJT is meant to be relevant to real-life situations.
2. The purpose of SJT is not to find the laws of human judgment, but to describe the judgment process.
3. Another purpose of SJT is to provide "cognitive aids for human judgment." This is especially important for those responsible for formulating any type of social policy from which there is apt to arise much dispute. Thus, an objective of SJT is not only to understand human judgment but also to develop ways of making it better.

Policy capturing shares a common heritage with SJT: the Brunswik Lens Model. SJT relies on multiple linear regression as an analytical tool for capturing individual and group strategies and for describing individual rater differences. The major difference between the two (PC and SJT) is equivalent to the difference between the lens model and policy capturing just mentioned: policy capturing represents only one-half of the lens model and SJT, the half dealing with the human system. That is, policy capturing is basically concerned with the internal judgment process and not the judgment task (environment) itself; SJT is concerned with both.

How Policy Capturing Works

Hoffman (1960) postulated the use of linear models to represent human judgment. Hoffman referred to such linear models as "paramorphic representations" (mathematical descriptions) of human decisionmaking processes. The term "paramorphic representation" is not intended to imply that the actual psycholcgical process involved in decision-making is one of differentially weighting various cues, but rather that the process could be represented by such weighting (Dawes and Corrigan 1974). Hence, "the mathematical description of judgment is inevitably incomplete..., and it is not known how completely or how accurately the underlying process has been represented" (Hoffman 1960, p. 125).

In essence, PC analysis requires that a group of judges study the information relevant to a given situation and then make an overall judgment or decision based upon this information. For example, Hamilton and Dickinson (1978) presented judges with profiles of hypothetical employees (see Figure 2). The judges were asked to review the information on the profile and then make an overall rating of that person's job performance.

A multiple linear regression model is next used to identify the informational cues considered by each judge and to indicate how such cues must be weighted to simulate the judge's decision. In the above example overall job performance ratings were regressed onto the ratings given the various elements of job performance. The result was a statistical model which theoretically approximates how judges weight information cues (Beach 1967; Bottenberg and

Job Elements for Machinists

Hypothetical Employee #37
Policy Capturing Profile

See attached scale for definitions of elements and meaning of points.

Elements	
Craftsmanship	 4
Practical Intelligence and Aptitude	 3
Alertness	 3
Planning and Organization	. . 2
Safety	 3
Positive Work Attitude	. . 2
Initiative	 4

Based on the profile of the hypothetical employee above, please circle your rating of the OVERALL JOB PERFORMANCE using the following scale:

1 2 3 4 5 6 7 8 9 10

Figure 2. Sample profile of machinist employee (from Hamilton and Dickinson 1978).

Christal 1961; Christal 1968b; Naylor and Wherry 1965; Dawes and Corrigan 1974). In this manner the researcher can generate a regression equation which "captures," or defines, the policy of an individual or group.

Clustering Method

Since policy capturing is aimed at revealing individual differences in judgment policies, a number of methods for grouping or clustering judges on the basis of homogeneity of their equations have been developed. Ward (1961) and Bottenberg and Christal (1961) are generally credited with the initial work in this area. The clustering technique most commonly employed in the literature and the one used in this research is called the Job Analysis (JAN) technique (Christal 1963; Naylor and Wherry 1965).

In actuality, JAN is a two-step procedure. During the first step (JAN I) the input data are the values of the predictor and criterion variables. JAN I then yields as output the raw score (unstandardized) regression weights, first-order validity coefficients (correlations between each predictor and the criterion), and the multiple correlation coefficient (R^2) for each judge's policy equation. Each R^2 demonstrates how well a particular judge's policy can be replicated (in a given sample) using a least squares composite (Christal 1968b). If the R^2 is relatively high, one can say that the judge's equation is reliable from judgment to judgment (Madden 1963; Slovic and Lichtenstein 1971). The R^2 also can be viewed as the extent of congruence between the linear equation and the inferential process of the judge (Goldberg 1968b).

JAN II, the second step of the clustering procedure, receives as input the regression weights and validity coefficients generated by JAN I. During JAN II individual equations are combined one at a time and a composite strategy is calculated at each step. The clustering procedure terminates when all individual equations have been grouped into one composite policy. The loss in predictive efficiency (R^2) in going from individual to composite equations is a measure of the reliability of both individual and group policies. If the drop in efficiency is small throughout the clustering process, one can say that the individual rater equations are highly homogeneous (i.e., the judges tend to agree on the relative weighting of the predictors) and that the composite strategy is reliable (Naylor and Wherry 1965; Dudycha and Naylor 1966; Hamilton and Dickinson 1978).

In the original formulation of the clustering process, Bottenberg and Christal (1961) proposed the use of the sequential F test as a criterion for determining the number of different policies which exist among the judges. This statistic tests the hypothesis that the two equations being clustered have the same set of regression weights. However, the F test criterion was found to be neither practical nor effective in indicating how many different groups or policies exist (Naylor and Wherry 1965). Instead, Bottenberg and Christal (1968) proposed the loss in R^2 in going from one group to the next as an alternative criterion. As an example, suppose that the greatest decrease in R^2 occurred when six separate equations were combined into five. In this case, according to Bottenberg and Christal, there are six meaningful clusters or groups and the results should be interpreted in light of the possible reasons for such group differences.

Clinical versus Statistical Prediction

The psychological literature on human decision-making abounds with examples in which linear models produce predictions superior to those of human judges (Goldberg 1970; Dawes and Corrigan 1974). In 1954, Meehl reviewed 20 studies in which statistical methods were compared to the judgments of clinicians (experts). Meehl found that in all 20 studies the statistical method outperformed or at least tied (in only 2 studies) expert judgment. Many other studies addressing the issue of clinical versus statistical prediction were conducted since Meehl's first review. Yet, Meehl (1965) was still forced to conclude that there was only one example showing clinical judgment to be superior over rather simple linear models; Goldberg (1968) argues that even this example does not show clinical superiority. Libby (1976a) reports one empirical example of a case in which judges clearly outperformed their model. Goldberg (1976) notes, however, that when some of Libby's cues are appropriately rescaled, the models outperform the expert. While Libby (1976b) does not agree with Goldberg's conclusion, the evidence still overwhelmingly favors the statistical model.

Goldberg (1970) provides one empirical example of the superiority of a model of man versus man himself. The judgment task in this study was distinguishing between neurotic and psychotic patients on the basis of their personality profiles. The expert judges were 29 clinical psychologists varying in experience and training. Least squares regression weights were computed for each judge in the usual policy capturing manner described above. Goldberg found that linear regression models of expert judges were more accurate predictors than were the humans modeled. Goldberg also discovered that modeling did not increase the predictive ability of the judgmental composite (average judgment of all 29 clinicians to each personality profile). That is, the composite judge was as accurate as its model.

As recently as 1974, Dawes and Corrigan reported that they could find no evidence from numerous studies reviewed for the superiority of expert over statistical prediction. Dawes and Corrigan concluded from their review that linear models were superior to expert judgments because of the following characteristics of the contexts in which such models were used: (1) the independent variables had conditionally monotone (positive) relationships with the dependent variable (or may be rescaled to have such relationships); (2) the dependent and independent variables possessed a certain amount of error; and (3) linear models are robust to changes in weightings of the independent variables (this holds as long as the weights are given an appropriate sign). In fact models whose weights were chosen randomly (except for sign) correlated more highly with dependent variables than did clinical judgment. According to Dawes (1977, p. 356),

> The psychological reason that linear models outperform people is that the models automatically perform the integration process at which people are quite poor (or at least much poorer than they think they are) while leaving the choice and coding of variables, at which people are quite good, to the people.

Linear versus Nonlinear Models

Despite the reported strength of the linear model, the proponents of statistical prediction do not deny the potential influences of nonlinear cue utilization. For instance, Meehl (1965) states that one advantage of clinical judgment may be the judge's ability to use configural (interactive) relationships between regressor variables and a criterion. In addition, both Goldberg (1968) and Slovic and Lichtenstein (1971) report investigations in which judges verbalized their cognitive processes as being complex ones relying on configural and curvilinear utilization of cues. This suggests that the linear model may be inadequate and that it may be necessary to add to the model higher-order (X^2,X^3, etc.) or interactive ($X1 \cdot X2$) terms. However, Goldberg (1968b, p. 488) reports numerous studies in which the introduction of more complex terms into the linear models "rarely served to increase the cross-validity of the new model." In light of this evidence and the results of other research investigating the linear/nonlinear issue, both Goldberg and Slovic and Lichtenstein conclude that humans <u>can</u> process information in a nonlinear fashion, but that the linear additive model is powerful enough to explain most judgments with a small amount of error. However, there is some experimental support to the notion that judges use cues in a configural manner (Slovic 1969). Thus, Slovic and Lichtenstein warn that even though nonlinear models are limited in their ability to outpredict linear models, "nonlinear processes are likely to play an increasing role in our understanding of judgment" (p. 683).

Practical Uses

The PC procedure has become popular mainly because of its ability to identify individual differences in judgment. Also, the method gives insight into many practical considerations, such as the key elements of good employee selection or the critical characteristics of student success in college. For instance, Zedeck and Kafry (1977) had nursing personnel evaluate hypothetical nurses. For registered nurses, Zedeck and Kafry derived an equation in which four predictors--empathy, clinical knowledge and performance, professional growth, and leadership ability--accounted for more than 60 percent of the variance in perceived nurse effectiveness. In the JAN clustering procedure, R^2 ranged from .6640 for 32 individual equations to .3811 for one composite equation. The criterion of largest decrease in R^2 indicated the existence of two separate groups. However, Zedeck and Kafry found attempts at stratifying the two groups according to background and organizational variables to be fruitless. Thus, they concluded that the sample of nursing personnel was homogeneous in their policy. Nevertheless, there were noticeable differences among individual raters in the elements that influenced their overall assessment of subordinates. This finding alone has important implications for identifying organizational goal differences in that it is possible for a rater to receive two different evaluations from two different raters. Zedeck and Kafry (p. 292) argue that "the results can provide reliable information to subordinates with respect to what is expected of them, what is valued by their supervisor (or organization), and what is to be rewarded."

Taylor and Wilsted (1974) were interested in studying the judgmental process involved in cadet ratings at the U.S. Air Force Academy. These researchers found that (1) raters were reliable in applying their _individual_ strategies to cadet ratings; (2) the policies among cadet raters varied (predictors received different weights by different raters); and (3) the stated policies of raters differed markedly from their mathematically-derived policies. Taylor and Wilsted reasoned that it was possible for a cadet to receive different ratings depending on his rater. All these findings are consistent with other investigations involving the PC procedure. The composite linear model indicated that overall cadet ratings could be described by knowing the ratings given three variables: performance of duty, cooperation, and leadership (the authors failed to report how much variance was explained by these three variables). The most likely explanation for why other predictors (e.g., acceptance of authority) did not account for much variance is that the rejected predictors were highly correlated with the three predictors retained and were thus implicitly included in the final model.

In a third study involving the PC paradigm, Anderson (1977) had 164 high school teachers evaluate 36 teacher profiles in terms of overall teacher quality. Among other things, Anderson found that raters were no less consistent when given an eight predictor profile than when given profiles with either six or four predictors. Although she states that her evidence is inconclusive, Anderson reasons that no matter how many cues are presented, the judge uses only a limited number in making his or her decision.

Policy capturing has also been successfully employed in research aimed at identifying and weighting the key elements of various Air Force specialities (Wherry and Naylor 1966), group policy toward nuclear safeguards (Brady and Rappoport 1973), psychiatric diagnoses (Goldberg 1965), college students' grade point averages (Wiggens and Kohen 1971), selection procedures for life insurance salesmen (Roose and Doherty 1976), the personnel status of Naval ships (Borman and Dunnette 1974), stockbrokers' decisions regarding the growth potential of stocks (Slovic 1969), and planning decisions involving the designing of transportation and other systems (Hammond et al. 1977). While this list is certainly not exhaustive, it does point out the wide range of applicability of policy capturing.

Summary

Since landscape assessment is still in its infancy, the search is on for reliable and valid methods of quantification and description of the visual resource. One of the more powerful analytic tools used to investigate landscape preference is multiple regression. The landscape preference studies which have used this type of analysis have generally shown favorable results. Multiple R^2 values have ranged from 0.330 to 0.988 with all but two R^2 values above 0.600. This indicates that the predictor variables studied have been able to account for much of the variance in landscape preference. There have even been two successful replications of one of the models (Shafer and Mietz 1970; Shafer and Tooby 1973). Of the eight studies using regression, three reported that judges were using visual information possibly in a nonlinear manner to arrive at their preference evaluations (Shafer et al. 1969; Wohlwill 1968; Buhyoff and Leuschner 1978). Thus, researchers in this field have extended the simple linear regression model to include nonlinear terms and parameters. Such an extension of the analysis has revealed the existence of complex relationships among predictor variables and between predictors and the criterion, landscape preference.

Despite these results, other researchers have voiced their criticisms of the multiple regression approach. These criticisms were summarized on p. 16. While not directly stated, one of the purposes of this chapter has been to suggest that the methodology to be employed in this dissertation, policy capturing, presents a conceptual framework backed by empirical evidence for answering most of the criticisms leveled against the multiple regression approach to landscape assessment. For instance, two of the strongest criticisms of the Shafer model in particular and the multiple regression paradigm in general are the lack of any theoretical foundation and the failure to account for individual differences. According to Daniel (1976) any measurement system should provide a clear conceptualization on which it is based. This allows the validity of the system to be evaluated in terms of how well the index or model it yields conforms to its own conceptualization of what it says it is measuring. Policy capturing, with its roots deeply embedded in the Brunswik Lens Model of behavior and Hammond's Social Judgment Theory, presents a well-developed conceptual rationale for the use of multiple regression in any task involving human judgment. Landscape assessment is obvicusly one such task.

In terms of individual differences, "one of the purposes of using a linear model to represent the judgmental process is to make the judge's weighting policy explicit. Large individual differences among weighting policies have been found in almost every study that reports individual equations" (Slovic and Lichtenstein 1971 , p. 678). Hence, one value of PC is that it allows the inference of an <u>individual</u> judge's policy by requesting an overall evaluation of a total stimulus (e.g., a landscape) rather than requiring evaluations of the elements making up that stimulus. "The focus of the analysis is on the decision of the rater rather than on the rater's interpretation of his/her decision process" (Zedeck and Kafry 1977, p. 287). The exposure of individual differences is crucial to landscape management (Craik and Zube 1976).

Finally, an effort was made to review briefly those studies which indicate the applicability of PC to numerous real-life situations. This "realism" of the PC procedure is a direct legacy from Brunswik and the social judgment theorists (Hammond et al. 1975). Therefore, while PC is not aimed at discovering the laws of human judgment, it is intended as a cognitive aid for those persons having to make everyday decisions with social ramifications. In this respect, PC seems to have practical advantages for those concerned with planning and management of the visual resource. A chief aim of this dissertation is to ascertain to what extent this last statement is true.

Chapter 3

METHODS

The methodology employed in the present study was designed to (1) obtain a representative sample of stimuli; (2) identify and quantify the appropriate independent variables (landscape features), and (3) obtain global preference evaluations of each stimulus as measures of the dependent variable. The steps discussed below generally follow these three goals.

Stimuli

The original stimuli for this study were over 700 35mm color slides of landscape scenes taken from the Blue Ridge Parkway (BRP) between Waynesboro, Virginia and Asheville, North Carolina. These slides represented a wide variety of scenes from panoramic views of distant landforms and rural environments to rather closeup shots of wooden fences, fields, forests and other combinations of natural and man-made (mostly rural) landscapes. The slides were taken by four different individuals (two graduate students in forest recreation, one assistant professor of forestry, and one U.S. Forest Service planner) and represented as closely as possible the range of scenes visually perceived along the BRP.

Daniel and Boster (1976) argue that photographs of landscape scenes should be taken in a random fashion so that the landscape in question is validly represented. No effort was made in this study to randomize the taking of slides. Nonetheless, for the reasons stated above, it is reasonable to assume that the slides were representative of BRP scenery. In addition, monetary and time constraints necessitated the use of slides taken for other landscape research projects and personal use.

Dimensions

In order to carry out the PC procedure, one must first be able to identify and quantify the important elements of landscape preference. The literature in this area consists of many studies in which anywhere from 50 to 100 different variables (a conservative estimate summed over all studies) have been proposed as contributors to landscape preference judgments. Since variable reduction was not one of the purposes of this research, another criterion was used as the basis for selecting predictor variables. In particular, Harvey's (1977) work was used as a guide for selecting landscape dimensions. There are two reasons for using her research as a guide. First, Harvey used five methods (Thurstone's Method of Paired Comparisons, Likert scales, semantic differential scales, grid analysis, and free response) to investigate the elements of landscape preference. Since the results of all methods tended to support each other, there is some convergent validity to her findings. Second, Harvey's slides and photographs were of BRP scenes. Since the same study area is used in this research, the important landscape features identified by Harvey should be nearly identical to the ones used for this study.

This study employed 10 of the most important dimensions identified by Harvey: foreground vegetation, mountains, man-changed area, visible distant landforms, green colors, blue colors, unobstructed expanse of view, sky, clouds, and undisturbed forest. The definitions of these dimensions are given in Table 1.

One reason for choosing only 10 dimensions is due to one of the limitations of PC. That is, to avoid overfitting regression equations it is recommended that each judge rate at least 10 stimuli per dimension (Naylor and Wherry 1964,1965; Hamilton and Dickinson 1978). Ten such judgments are also necessary to insure sufficient variability along each dimension. With 10 dimensions and 10 stimuli per dimension, each judge had to rate 100 slides. With any more than 100 such judgments, subject boredom and fatique could confound the results. Since 15 seconds per judgment were allowed, it took subjects only 25 minutes to make the 100 judgments. Pretesting the procedure indicated that this many judgments would not be unduly burdensome.

Table 1. Definitions of dimensions

Dimension name	Definition
Sky (S)	The proportion of the scene containing clear (unclouded), blue sky.
Foreground vegetation (FV)*	The proportion of the scene containing grasses, shrubs, trees, or other plants in the front-most portion of the scene. The "front-most portion" is that area of the scene where individual leaves, twigs, blades of grass, needles, etc. are clearly discernable.
Green (G)	The proportion of the scene containing any shade of the color green. Blue-green counts as green.
Blue (B)	The proportion of the scene containing any shade of the color blue, except blue-green.
Mountains (M)	The proportion of the scene containing mountains, as opposed to rolling hills, flat land, and sky.
Man-changed area (MCA)	The proportion of the scene which <u>man</u> has visibly altered in some manner. These changes may be either positive (beneficial) or negative (harmful).
Unobstructed expanse of view (UEV)*	The extent to which an observer can clearly see a distant scene beyond the foreground. If a brick wall or a mountain were directly in front of the observer, then the observer's view would be highly obstructed. UEV can also be considered that proportion of the middleground, background, and unclouded sky which is unobstructed.

(continued)

Table 1. Definitions of dimensions (continued).

Dimension name	Definition
Visible distant landforms (VDL)*	The proportion of the scene containing land features (flat land, hills, mountains, etc.) which the observer can clearly see. "Distant" means beyond the foreground of the scene and includes visible landforms in the middleground and background but does not include sky or clouds.
Undisturbed forest (UF)	The proportion of the scene containing forested areas that have <u>not</u> been noticeably changed by either human or natural forces. UF may include foreground if the foreground is forested.
Clouds (C)	The proportion of the scene containing any type of cloud formation. Clouds do not include haze.

*For the definitions of each of these dimensions, a slide on which the foreground was clearly delineated with a black line was shown.

Another reason for keeping the number of dimensions relatively small is reported by Darlington (1968). Darlington (p. 174) argues that "the estimated true validity of a sample multiple regression equation is very low (and the mean square error very high) when the number of predictor variables is large in relation to the number of people in the sample on which the equation was derived." As an example he cites a study in which a regression equation with 84 independent variables and data from 135 subjects was constructed. The correlation with the criterion was 0.73 in the initial sample and 0.04 in the cross-validation sample. Darlington concludes that " it is often better to use fewer predictor variables, or to use a different prediction method altogether, than to use a regression equation with an extremely large number of variables" (p. 174).

One final justification for the use of 10 variables is supplied by Wang and Stanley (1970). In a review of various weighting procedures Wang and Stanley note that differential weighting is probably worthwhile only when there are a small number (3 to 10) of predictors, their average intercorrelation is low (0.50 or less), and the weights have substantial standard deviations. Fewer than 10 dimensions could be used in this study. However, since this is primarily an exploratory study, it is desirable to begin with as many variables as possible given the limitations of the methodology discussed above.

"Expert" Judges: Phase 1

As an initial cut the original 700 slides were reduced to about 300 on the basis of poor photographic quality. Some of the original 700 slides were also eliminated because they were near-duplicates of other slides or because people were contained in the scenes. Finally, about one-fourth of the 700 were discarded because they were taken during seasons other than the season during which the landscapes were to be judged. This latter point needs some clarification. Using a paired comparison routine, Buhyoff and Wellman (1979) have revealed the potential for seasonality bias in landscape preference research. These researchers found a significant interaction between season of photography and season of preference judgments. In other words, if landscape photographs display various seasonal characteristics (e.g., fall colors), then preferences may be biased toward a certain season depending on the season in which preferences are elicited. To illustrate, spring scenes shown in the winter are likely to receive higher ratings than the same scenes shown during the spring. Since subjects in the present study were asked for their preference judgments in late spring, summer, and early fall, slides showing fall colors and winter or early spring characteristics were automatically eliminated.

A panel of judges composed of 11 forest management/ economics graduate students, 3 forest recreation graduate students, and one undergraduate senior majoring in interior design (n= 15) assisted in reducing the 300 slides selected in the first cut down to the 100 needed for the judgment task.

Procedure: Phase 1

The desired result in this phase of the study was a set of 100 slides (10 dimensions x 10 judgments per dimension) varying in amounts of each of the 10 dimensions to be evaluated in Phase 2. Each of the 15 subjects (from the panel of judges just described) was asked to study the 300 slides carefully (all 300 were laid out on a large light table) and pick out 10 which varied in the amount of the dimension that that particular subject was given to evaluate. Thus, the first subject was asked to choose 10 slides from the 300 which varied along a scale from 0 to 9 in amount of mountains, 0 meaning no mountains and 9 meaning a maximum amount of mountains ("maximum" refers to that scene of the 300 slides available which was most mountainous). Each subject was requested to select two slides which clearly represented the 0 and 9 points on the scale and to choose the remaining 8 to represent points between 0 and 9 with most of these 8 representing amounts near the middle of the scale. This procedure took about 15 to 20 minutes per subject. For each of the 10 dimensions, 2 subjects selected 10 slides; some subjects selected slides for more than one dimension. Thus, instead of randomly selecting 100 slides, attention was paid to choosing a set of simulated landscapes which varied in amounts of those predictor variables selected to investigate preference. The 100 slides used are filed with the School of Forestry and Wildlife Resources at VPI & SU.

Subjects: Phase 2

The data for this phase of the study were collected during late spring of 1978. Subjects consisted of 225 undergraduate students enrolled in the introductory psychology class at VPI & SU. This 3,500 person subject pool is maintained by the Department of Psychology and is a statistically representative cross-section of the student population at VPI & SU. Subjects were asked to participate voluntarily in this study with additional credit in the introductory psychology course serving as an incentive. Based upon previous research results showing similarity in landscape judgments across groups (Wenger and Videbeck 1969; Zube 1974; Riesenman 1977; Buhyoff and Leuschner, 1978) it was not considered necessary for this exploratory study to sample actual recreationists or citizen groups.

Of the 225 subjects sampled, 99 were males and 126 were females, the average age was 19 (standard deviation = 1.7), and 36 percent were freshmen and sophomores. Many indicated that it was too early in their college careers to specify their majors, so this information, though collected, was not considered useful.

Procedure: Phase 2

Each of the 10 groups (n = 22 to 24 per group) rated a different dimension according to the definitions given in Table 1. For example, the first group was asked to rate the 100 slides according to the amount of foreground vegetation each scene possessed. The next group was asked to rate the same 100 slides according to the amount of mountains, and so on until all 10 dimensions were rated. A 9-point rating scale[2] was employed with a "1" on this scale representing complete absence of a given dimension and a "9" representing nearly complete coverage of the scene by a given landscape feature. In all cases, subjects were shown 5 practice slides which they attempted to rate according to the definitions given. These practice slides were chosen to vary in the relative amount of each dimension from high to

---

[2] Both Guilford (1954) and Nunnally (1970) state that numerous studies indicate that the reliability of rating scales increases monotonically with the number of scale steps. Reliability increases, very rapidly at first, as the number of steps increases from 2 to 20. Gains in reliability level off at around 7 steps and beyond about 11 steps there is little gain. Thus, the 9-point scale used in this study offers substantial reliability with the added advantage of being less cumbersome than a larger scale (i.e., only one column per slide is required for coding rater responses).

low. After each practice slide subjects were asked to voice any problems they had in making their ratings. Every effort was made to assure that they understood the instructions and the scale. Fifteen seconds were allowed per judgment. Five duplicate slides were placed at the beginning and at the end of the tray of 100 to see if subjects were using the scale consistently. The complete instructions are given in Appendix A. For this phase of the study rater selection for a particular group was based solely on their availability for testing rather than by random assignments.

Subjects: Phase 3

The 78 subjects used in this phase of the study came from the same subject pool described in Phase 2. These subjects consisted of 34 males and 44 females; their average age was 19 (standard deviation = 2.05) and 80 percent were freshmen and sophomores. Again, there were sufficient uncertainties concerning majors and so many different majors listed that this information was not considered useful.

Procedure: Phase 3

All 78 subjects used a 10-point scale ranging from 0 to 9 to rate their overall preference for the same 100 slides shown in Phase 2. This time the low end of the scale represented no preference for the slide (negative preferences for the landscape scenes used in this study were assumed to be nonexistent) while the high end represented a maximum amount of preference. Based on the free responses in Phase 2, which indicated slight confusion among some subjects, it was decided to add the "0" to make the scale more meaningful. That is, a "0" representing no preference makes more intuitive sense than a "1."

The instructions given these subjects were very similar to those presented in Appendix A with the major exception of how the scale was anchored. For this stage of the study, all 78 subjects were told:

> "Your task this evening will be to rate 100 slides of landscape scenes according to how much you prefer a given scene. The rating scale you will be using is on the blackboard. A zero on this scale means that you have no preference whatsoever for a given scene. In other words, you are neutral toward that scene; you have neither positive nor negative feelings for that scene. A nine on this scale indicates the maximum amount of preference you can have for any given scene. In other words, a nine represents the landscape scene that you most highly prefer or that you have the most positive feelings toward. A four or five represents an intermediate amount of preference for a scene. Each unit on the scale represents an additional amount of preference. For each slide shown simply blacken in a number from zero to nine on your op-scan sheet. The number you blacken in represents your amount of preference for each scene."

Analysis

Policy Capturing

Using JAN I, a multiple regression equation was computed for each of the 78 raters. Overall preference ratings served as values for the dependent variable while the means of the 10 dimension ratings (Appendix B) served as values of the independent variables. Thus, for each subject the criterion was a vector of 100 preference ratings corresponding to each of the 100 slides. The raw data for each of the 10 predictors, on the other hand, were represented by a 22 - 24 x 100 matrix (22-24 subjects per dimension and 100 slides per subject); the mean rating for each dimension (n = 22-24) served as the value of each of the 10 independent variables entered into JAN I.

The JAN I output has already been discussed (see p. 32). To reiterate, the output consists of the multiple correlation coefficient, R^2 (a measure of individual subject reliability), and zero-order validity coefficients and beta weights for each landscape dimension. Beta weights indicate the relative importance of each dimension. Theoretically, when the predictor variables are all orthogonal to each other, the zero-order validity coefficients can be used as indices of the importance (percent of variance explained) of each dimension to the rater's overall preference judgment. However, it is reasonable to conclude that some of the dimensions used in this study were relatively highly correlated (e.g., amount of mountains and amount of visible distant landforms). The decision to use such variables was based on the desire to include as much realism in the study as possible. One of the costs of this decision is the increased difficulty in interpreting the zero-order validity coefficients.

In addition to R^2, the reliability of rater judgment was also assessed by deriving a test-retest measure on the ratings of the five slides at the beginning and end of the 100 slides. That is, a Pearson Product Moment correlation between the ratings given the two groups of five slides was computed. Such a measure can be considered a "coefficient of stability" (Lemke and Wiersma 1976, p. 72) since it is the correlation between ratings given at two different times. A stability coefficient of 0.80 means that 80 percent of the variance in ratings remained stable from Time 1 to Time 2; 20 percent of the variance was random and/or due to the instability of the phenomenon being measured.

Clustering

Individual regression equations were clustered using the JAN II program. The output of JAN II has also been previously discussed (see p. 33). To reiterate, at each stage of the grouping process, there is a measure of the composite group efficiency, R^2. If there are 50 raters, the first R^2 is the predictive efficiency of a system with 50 different policies. The second R^2 is the predictive efficiency of a 49 separate policy system or one cluster of the two most homogeneous raters and 48 distinct policies. The last R^2 is the predictive efficiency of a one-policy system, or all 50 judges clustered into one group.

Chapter 4

RESULTS

Individual Regression Analyses

Tables 2 and 3 present the results of the regression analyses (JAN I) performed upon the preference judgments of each rater. Table 2 shows the zero-order correlations (Rxy) between the landscape dimensions and the overall preference judgments of each rater. The last column of Table 2 gives the R^2 values for each rater, an index of the consistency of rater judgment across the 100 slides. As Table 2 indicates, only 17 of the 78 raters can be considered fairly consistent ($R^2 > 0.50$). Thus, one can say that for only about 22 percent of the sample are the preference judgments fairly predictable by linear regression (Borman and Dunnette 1974). There is also a substantial degree of variation in consistency with R^2 values ranging from a high of 0.66 (Rater 77) to a low of 0.08 (Raters 40 and 65).

To exemplify the interpretation of the zero-order correlations, the Rxy between landscape preference and amount of mountains (M) is -0.67 for Rater 1 indicating that as the proportion of the scene containing mountains increases, preference decreases. If the landscape dimensions were reasonably uncorrelated with each other

Table 2. Zero-order validity coefficients (${}^{r}xy$) associated with each of the dimensions.

	Dimension Codes*-										
Rater	S	FV	G	B	M	MCA	UEV	VDL	UF	C	R^2
1	.03	.14	-.37	.10	-.67	.65	-.41	-.24	-.61	-.20	.59
2	.03	.26	-.24	-.14	.38	-.25	-.25	-.25	-.28	-.08	.24
3	-.16	.07	-.03	-.16	-.20	.35	-.09	.08	-.18	.08	.26
4	.04	.32	-.20	.04	-.42	.18	-.15	-.37	-.27	-.03	.28
5	.06	-.19	.01	.00	.24	-.19	.30	.05	.12	.12	.22
6	.08	-.05	-.30	.23	-.38	.49	-.52	-.06	-.43	-.16	.38
7	.18	.04	.12	.12	.08	-.31	.22	-.16	.13	.01	.22
8	-.18	.05	-.08	-.06	.08	-.12	-.03	.08	.16	.24	.16
9	.02	-.10	.13	-.04	-.27	.26	.08	-.03	-.19	-.10	.32
10	-.12	.08	-.34	.00	-.34	.44	-.39	-.10	-.39	.03	.31
11	-.16	-.06	-.06	-.04	-.24	.15	-.11	-.23	.22	.27	.25
12	.19	-.23	.00	.18	.02	-.16	.04	.00	.03	-.04	.20
13	.03	-.21	-.33	.10	-.29	.44	-.23	.01	-.42	-.10	.29
14	.07	-.01	.03	.14	.08	-.18	-.11	-.07	.10	.13	.23
15	.13	.08	-.51	.31	-.56	.50	-.50	-.28	-.56	-.20	.49
16	-.02	.01	-.37	.02	-.38	.57	-.33	-.04	-.49	-.14	.44
17	.28	.09	-.31	.37	-.48	.47	-.25	-.30	-.56	-.33	.44
18	-.04	-.19	-.34	.14	-.29	.49	-.38	-.03	-.42	.02	.41
19	-.12	.29	-.31	.03	-.49	.53	-.52	-.20	-.42	-.03	.49
20	.11	-.02	-.47	.23	-.44	.61	-.63	-.14	-.56	-.19	.57
21	.10	.18	-.32	.12	-.59	.54	-.44	-.30	-.56	-.22	.45
22	.14	.11	-.33	.28	-.21	.12	-.38	-.14	-.21	-.07	.28
23	-.07	-.08	-.23	.06	-.32	.53	-.64	.09	-.40	-.06	.49
24	-.24	.33	-.13	-.10	-.27	.20	-.41	-.18	-.13	.24	.33
25	.13	.06	-.12	.18	-.36	.23	-.20	-.22	-.31	.00	.32
26	.03	-.06	-.33	.11	-.40	.46	-.36	-.03	-.37	-.16	.35
27	-.09	.34	-.10	-.12	-.55	.47	-.27	-.25	-.46	-.13	.59
28	-.13	-.09	-.01	-.02	-.03	.13	-.11	.03	-.08	.15	.12
29	-.17	-.13	.23	-.22	.29	-.20	.14	.25	.28	.24	.16
30	.00	.08	-.21	.10	-.39	.30	-.49	-.16	-.36	.07	.42
31	.11	-.15	-.21	.18	-.29	.36	-.30	-.02	-.32	-.16	.21
32	-.16	-.18	-.07	-.03	.20	-.01	-.23	.28	.14	.21	.26
33	-.06	.18	.11	-.04	.22	-.38	.04	-.02	.36	.24	.29
34	.20	-.16	-.51	.27	-.45	.53	-.40	-.11	-.55	-.19	.46
35	.05	.21	-.45	.19	-.41	.31	-.28	-.30	-.40	-.04	.33
36	.21	.07	-.44	.39	-.55	.35	-.24	-.44	-.51	-.15	.46
37	.00	-.01	-.06	.10	.18	-.18	-.02	.08	.15	.09	.13
38	.14	.06	-.51	.29	-.44	.35	-.35	-.20	-.42	-.18	.42
39	-.01	.16	-.31	.10	-.37	.28	-.21	-.24	-.33	-.04	.22
40	.04	-.06	.06	.07	.06	-.12	.14	-.01	.07	.06	.08
41	.10	.09	-.39	.17	-.41	.42	-.51	-.17	-.43	-.15	.36

(continued)

Table 2. Zero-order validity coefficients (^{r}xy) associated with each of the dimensions (continued).

					Dimension Codes*-						
Rater	S	FV	G	B	M	MCA	UEV	VDL	UF	C	R^2
42	-.24	.19	.07	-.14	-.08	.03	-.21	-.06	.03	.27	.18
43	-.10	-.12	.32	-.14	.13	-.05	.07	.17	.13	-.04	.22
44	-.33	.20	-.18	-.05	-.01	.08	-.50	.07	.06	.37	.53
45	-.12	.00	-.11	.06	.05	.04	-.38	.07	.09	.08	.34
46	.02	-.26	-.13	.04	-.14	.22	-.08	.06	-.22	.04	.18
47	.17	.00	-.44	.29	-.30	.25	-.26	-.27	-.34	-.15	.24
48	.09	-.09	-.13	.12	-.54	.61	-.43	-.09	-.54	-.19	.50
49	-.13	.19	-.43	.04	-.51	.64	-.58	-.12	-.50	-.01	.65
50	.10	-.05	-.42	.20	-.62	.63	-.43	-.21	-.63	-.23	.53
51	-.27	.02	-.10	-.10	.10	-.02	-.20	.08	.05	.31	.22
52	.17	-.05	-.30	.12	-.60	.69	-.31	-.11	-.63	-.36	.61
53	.07	.10	-.36	.05	-.40	.35	-.12	-.20	-.38	-.19	.33
54	-.22	.15	.25	-.15	.46	-.59	.31	.08	.58	.39	.46
55	-.06	.08	-.32	.01	-.45	.59	-.48	-.08	-.47	-.14	.46
56	.30	.38	-.13	-.15	-.10	.06	-.13	-.16	-.01	.24	.26
57	.55	-.07	-.49	.53	-.49	.43	-.15	-.36	-.62	-.59	.61
58	-.08	.25	-.26	.11	-.17	.03	-.32	-.13	-.05	.14	.32
59	.13	.27	-.22	.18	-.30	.09	-.02	-.30	-.20	-.16	.22
60	-.01	.04	-.17	.10	-.23	.20	-.42	-.12	-.21	.13	.28
61	.03	.07	-.43	.10	-.54	.58	-.52	-.17	-.53	-.20	.52
62	.10	.18	-.41	.16	-.63	.61	-.37	-.32	-.64	-.27	.57
63	.06	.22	-.28	.19	-.40	.36	-.41	-.26	-.38	-.08	.29
64	.20	.03	-.35	.24	-.54	.63	-.44	-.20	-.64	-.40	.57
65	.08	-.12	.07	.06	.04	-.01	-.07	.10	.04	-.16	.08
66	.11	.19	-.41	.18	-.63	.65	-.39	-.32	-.62	-.24	.58
67	.23	.20	-.17	-.17	-.34	.44	-.36	-.10	-.31	.06	.34
68	.02	.00	-.54	.23	-.23	.18	-.42	-.14	-.19	.05	.50
69	.29	.02	-.47	.40	-.26	.08	-.13	-.28	-.29	-.20	.27
70	.06	-.02	-.10	.08	-.10	.00	.11	-.08	-.07	.03	.11
71	.00	.01	.00	.01	.27	-.32	.01	.09	.34	.10	.25
72	.15	-.06	-.36	.17	-.38	.46	-.36	-.09	-.45	-.34	.39
73	.01	-.02	-.48	.12	-.54	.69	-.55	-.07	-.62	-.19	.65
74	-.04	.05	-.36	.12	-.39	.51	-.50	-.08	-.40	-.07	.42
75	.02	.13	-.24	.08	-.25	.29	-.28	-.07	-.28	-.05	.20
76	.24	-.04	-.22	.17	-.41	.27	-.07	-.23	-.36	-.22	.29
77	.22	.25	-.32	.20	-.72	.62	-.48	-.42	-.66	-.35	.66
78.	-.30	.12	-.08	-.12	.36	-.31	-.17	.14	.40	.52	.55

*S=sky; FV-foreground vegetation; G=green colors; B=blue colors; M=mountains; MCA=man-changed area; UEV=unobstructed expanse of view; VDL=visible distant landforms; UF=undisturbed forest; C=clouds.

Table 3. Unstandardized regression weights (^{b}x) associated with each of the dimensions.

	Dimension Codes*-									
Rater	S	FV	G	B	M	MCA	UEV	VDL	UF	C
1	-.10	.23	-.24	.06	-.75	.31	.10	.41	.18	-.08
2	-.07	.39	-.18	.29	-.52	.09	.07	.39	.17	.03
3	.36	.27	-.15	.14	-.68	.71	.36	.57	.64	.40
4	-.50	.14	-.08	.01	-.37	-.29	.00	-.10	-.21	-.25
5	.50	-.15	-.03	-.08	.36	.25	.26	-.19	-.01	.49
6	-.17	.04	-.05	.46	-.22	.41	-.21	.22	.24	.02
7	.49	.02	.32	.00	-.05	-.43	.00	-.02	-.35	.44
8	.03	.06	-.24	.32	-.57	.06	.08	.38	.60	.38
9	-.03	-.29	.30	.29	-.88	.55	.39	.08	.66	.22
10	-.34	.10	-.27	-.06	.06	.02	-.14	.06	-.34	-.14
11	.24	.11	-.23	.36	-.30	.07	-.08	.51	.50	.50
12	.39	-.30	.23	.17	-.54	-.34	-.08	.20	.07	.50
13	-.34	-.15	-.21	.21	-.13	.30	.12	.13	-.06	-.06
14	.35	-.13	.33	.28	-.02	-.20	-.30	-.09	-.04	.61
15	-.38	.18	-.26	.38	-.30	.03	-.13	.25	-.14	-.13
16	-.16	.26	-.42	-.15	.16	.38	.07	.22	-.24	-.30
17	-.08	.44	.05	.52	.08	.28	.17	.31	-.50	-.02
18	-.51	-.34	-.05	.54	.15	.70	-.04	-.31	.22	.13
19	-.32	.35	-.22	.18	-.30	.25	-.16	.24	.07	-.14
20	-.12	.07	-.27	.02	.38	.42	-.38	-.12	-.12	-.21
21	.09	.22	-.13	-.28	-.23	-.02	-.19	.09	-.30	-.22
22	.25	.41	-.21	.26	-.21	-.28	-.27	.55	-.20	.22
23	-.34	-.01	.06	.16	-.01	-.10	-.51	.24	-.42	-.12
24	-.10	.08	-.01	.06	-.41	-.01	-.33	-.04	.24	.26
25	.58	.16	.32	.41	-.82	-.04	.06	.55	-.10	.88
26	-.12	.01	-.36	.21	-.68	.53	.05	.35	.75	-.09
27	-.36	.43	.07	-.45	-.31	-.71	-.09	.35	1.14	-.47
28	-.44	-.31	.24	.39	.08	.16	-.05	-.27	-.07	.19
29	.28	-.22	.14	-.12	-.17	-.05	-.02	.07	.20	.37
30	.29	.10	.25	.04	-.58	-.66	-.45	.48	-.70	.55
31	-.08	-.15	-.02	.39	-.46	.43	-.02	.16	.48	.10
32	.02	-.19	-.15	.36	-.14	.33	-.21	.17	.70	.33
33	.68	.15	-.01	-.01	-.38	-.16	-.16	.20	.53	.60
34	.39	.11	-.38	.12	-.37	.48	.04	.43	.24	.29
35	-.06	.38	-.39	.18	-.22	.06	.07	.29	-.11	.07
36	-.32	-.03	-.05	.76	-.56	.36	.21	-.01	.30	.28
37	.10	.19	-.10	.42	-.01	-.04	-.02	.34	.12	.28
38	-.11	.31	-.48	.36	-.62	.06	.06	.60	.23	-.02
39	-.38	.18	-.25	.15	-.28	-.07	.07	.15	-.14	-.14
40	.03	-.09	.21	.38	-.23	-.02	.16	.10	.03	.36
41	.14	.27	-.29	-.18	-.06	-.06	-.34	.25	-.23	-.10

(continued)

Table 3. Unstandardized regression weights (^{b}x) associated with each of the dimensions (continued).

	Dimension Codes*-									
Rater	S	FV	G	B	M	MCA	UEV	VDL	UF	C
42	-.16	-.08	.25	.12	-.30	-.21	-.26	-.08	-.04	.32
43	-.78	-.41	.46	.00	.26	-.34	-.18	-.42	-.40	-.55
44	-.25	.17	-.18	.58	-.31	.11	-.39	.31	.50	.40
45	-.57	-.34	-.16	.38	.14	.51	-.45	-.52	1.02	-.28
46	.18	-.25	.07	.15	-.41	.10	.10	.23	-.03	.47
47	-.22	-.08	-.35	.16	.20	.40	-.06	-.36	.28	-.15
48	.04	-.09	.33	.29	-.71	.22	-.09	.34	-.03	.26
49	-.16	.34	-.43	.24	-.42	.60	-.05	.40	.42	.01
50	-.36	-.09	-.18	.24	-.57	.35	.03	.13	.20	-.13
51	-.63	-.14	.00	.18	.28	-.46	-.31	-.13	-.53	.01
52	.19	.24	-.21	-.08	-.58	.49	.21	.52	.12	-.14
53	-.09	.24	-.52	-.35	-.23	.07	.19	.20	-.04	-.39
54	.03	-.03	.03	.30	-.16	-.07	.09	-.08	.62	.41
55	-.27	.13	-.34	-.16	-.08	.40	-.16	.05	.14	-.40
56	-.60	.23	-.26	.12	.14	.11	.01	-.22	.09	-.19
57	.22	.34	-.34	.12	.24	.46	.24	.12	-.15	-.31
58	.06	.38	-.28	.42	-.64	-.13	-.14	.58	.37	.34
59	-.03	.47	-.25	.21	-.38	-.02	.25	.37	.03	-.09
60	.36	-.05	.17	.10	-.30	-.17	-.42	.11	-.13	.60
61	-.20	.11	-.41	-.16	-.32	.25	-.19	.13	.20	-.36
62	-.35	.36	-.24	-.06	-.08	.02	.04	.19	-.55	-.42
63	.00	.32	-.05	.29	-.07	.24	-.15	.12	-.04	.12
64	-.49	.25	-.07	.10	.26	.08	-.12	.09	-.66	-.60
65	-.25	-.18	.09	.02	.04	-.09	-.19	-.09	.10	-.37
66	-.02	.31	-.32	.09	-.25	.72	.14	.09	.31	-.14
67	-.43	.06	-.17	-.28	.04	.14	-.18	-.19	-.13	-.33
68	-.37	-.16	-.66	.43	-.36	.52	-.16	-.11	1.13	-.01
69	.02	.24	-.40	.38	-.08	-.13	.06	.25	-.13	.01
70	.39	.10	-.07	.24	-.62	.13	.34	.42	.28	.53
71	.48	-.05	-.34	-.10	-.17	.28	-.16	-.05	1.10	.17
72	-.39	.06	-.35	-.23	.19	.02	-.19	-.01	-.26	-.71
73	-.33	.21	-.39	.02	-.19	.15	-.12	.41	-.26	-.30
74	-.19	.08	-.30	.35	-.31	.72	-.10	.12	.69	-.01
75	.20	.48	-.21	.07	-.14	-.02	-.02	.57	-.30	.11
76	.45	-.05	.38	.06	-.70	.00	.10	.23	-.09	.39
77	.16	.32	-.04	-.32	-.20	.04	-.22	.04	-.37	-.27
78	.26	-.02	.08	.43	.02	.10	-.36	-.02	.54	.80

*S=sky; FV=foreground vegetation; G=green colors; B=blue colors; M=mountains; MCA=man-changed area; UEV=unobstructed expanse of view; VDL=visible distant landforms; UF=undisturbed forest; C=clouds.

these zero-order correlations could be used as an index of importance. The conclusion would be that for Rater 1 44.4 percent ($-0.67^2 \times 100$) of the predictable variance in landscape preference is accounted for by the amount of mountains in a scene. Similarly, amount of sky would account for only 0.12 percent of the variance in Rater 1's preference. However, Table 4 indicates substantial correlations among some of the predictors (e.g., the correlation between amount of sky and amount of mountains is 0.85). Thus, the interpretation of zero-order correlations as indices of importance is not strictly appropriate in this study (Zedeck and Kafry 1977). Due to substantial redundancy among certain dimensions, it is difficult to tell from Table 2 which are more important. For example, given that the correlation between mountains and sky is 0.85, are mountains really more predictive of Rater 1's preference than sky, or is the information given by the amount of sky in a scene implicitly included in the information provided by the amount of mountains? Also, Schenk and Naylor (1968) have shown that

> if one assumes a linear model of some specified degree of fit for a subject, any increment in cue intercorrelation will *necessarily* yield an increase in the degree of fit of that model (p. 3).

However, this issue is not entirely settled because in Schenk and Naylor's work the regression weights remained

Table 4. Correlation matrix for landscape dimensions and preference.[a]

	Landscape Preference	S	FV	G	B	M	MCA	UEV	VDL	UV	C
Landscape Preference											
S	.02[b]										
FV	.04	-.31									
G	-.19	-.36	.14								
B	.09	.85	-.22	-.55							
M	-.22	-.34	-.19	.50	-.38						
MCA	.23	.00[b]	-.09	-.31	.02[b]	-.65					
UEV	-.23	.03	-.08	.34	-.15	.43	-.57				
VDL	-.09	-.44	-.46	.43	-.51	.67	-.06	.06			
UF	-.22	-.42	.13	.61	-.46	.85	-.80	.44	.47		
C	-.04	-.83	.27	.20	-.67	.40	-.21	.04	.26	.44	

[a] S=sky; FV=foreground vegetation; G=green colors; B=blue colors; M=mountains; MCA=man-changed area; UEV=unobstructed expanse of view; VDL=visible distant landforms; UF=undisturbed forest; C=clouds.

[b] Not significantly different from zero (p > .03).

relatively stable across sets of predictors varying in amount of correlation and, therefore, much of the change in linear consistency (R^2) may have been a statistical artifact (Slovic and Lichtenstein 1971).

Table 3 provides the raw score regression weights for each independent variable for each subject. The vector of regression weights represents each subject's policy and each of the weights may be thought of as an importance index (Borman and Dunnette 1974; Taylor and Wilsted 1974). Others, (e.g., Naylor and Wherry 1965) argue that raw score regression weights are notoriously hard to interpret due to such effects as cue intercorrelations (discussed above) and suppressor relationships. Nonetheless, based on empirical investigations of 5 different weighting procedures, Claudy (1972) states the following general guideline:

> For samples in which the criterion correlations show little variability and the intercorrelations among the independent variables are on the average either high positive or low negative (Class Y), use equal raw score weights for samples of less than 50 and beta weights for larger samples (p. 320).

Since this is only a general guideline and the parameters of this study closely resemble those of Claudy's recommendation, the interpretation of the raw score regression weights in Table 3 as relative importance indices is not unreasonable (use of standardized beta weights would not be unreasonable either). As such, the preference evaluations of Rater 1 appear to be most heavily influenced (largest weight) by the amount of mountains and least influenced by the proportion of blue colors. Also, note the wide range of variation in individual policies in terms of the relative weights assigned to the 10 dimensions. There was no clear pattern in terms of which dimension received the highest weight. The five dimensions most often receiving high weights were: sky (by 17% of the Ss), mountains (by 15% of the Ss), undisturbed forest (by 19% of the Ss), clouds (by 13% of the Ss), and man-changed area (by 12% of the Ss).

JAN II Clustering

Table 5 indicates that as the number of groups is systematically reduced from 78 separate equations to one overall policy, R^2 went from 0.358 to 0.109. The former R^2 means that about 36 percent of the variance in landscape preference is predicted by the 10 dimensions by using 78 policies. If only one policy is used, only 11 percent of the variation in preference is predicted. Because of the failure to consider information contained in individual equations, there is a loss in predictive efficiency as raters are combined.

The cut-off point for determining how many homogeneous clusters of raters exist is described by Bottenberg and Christal (1968) as being that stage of the iterative

Table 5. Composite predictive efficiency (R^2 from JAN II) and standard error of the estimate (SE) at each stage of the grouping process.

Number of Groups	R^2	SE	Number of Groups	R^2	SE	Number of Groups	R^2	SE
78	.358	--	52	.331	.192	26	.276	.547
77	.357	.028	51	.328	.194	25	.272	.556
76	.357	.048	50	.326	.199	24	.267	.571
75	.356	.051	49	.323	.222	23	.263	.578
74	.355	.056	48	.321	.226	22	.259	.634
73	.355	.058	47	.320	.229	21	.258	.644
72	.355	.093	46	.319	.247	20	.257	.706
71	.354	.099	45	.316	.250	19	.254	.729
70	.353	.100	44	.315	.252	18	.251	.755
69	.352	.104	43	.312	.262	17	.249	.794
68	.351	.104	42	.310	.266	16	.245	.896
67	.350	.108	41	.307	.269	15	.238	.911
66	.349	.113	40	.304	.271	14	.236	.912
65	.349	.114	39	.301	.277	13	.233	.993
64	.349	.117	38	.300	.282	12	.231	1.057
63	.347	.120	37	.299	.282	11	.228	1.388
62	.345	.121	36	.298	.287	10	.218	1.402
61	.344	.128	35	.296	.288	9	.213	1.420
60	.343	.134	34	.294	.291	8	.207	1.474
59	.341	.140	33	.291	.346	7	.201	1.836
58	.341	.142	32	.287	.352	6	.194	2.514
57	.338	.145	31	.284	.366	5	.187	2.749
56	.338	.145	30	.283	.388	4	.167	3.102
55	.336	.161	29	.282	.455	3	.161	4.859
54	.334	.176	28	.279	.469	2	.148	7.854
53	.333	.186	27	.278	.499	1	.109	--

procedure where the inclusion of one more rater policy produces the greatest drop in predictive efficiency (R^2). As can be seen in Table 5, the greatest initial drop in predictive efficiency is in the change from 5 groups (R^2 = 0.187) to 4 groups (R^2 = 0.167) suggesting the existence of 5 meaningful clusters. However, Zedeck and Kafry (1977) did not find this criterion very helpful in terms of differentiating clusters. Likewise, Dudycha and Naylor (1966) found very little difference between the policies of the different clusters that JAN identified using the change in R^2 criterion. JAN II, however, also gives the change in error at each stage of the grouping process. As shown in Table 5, the standard error for this study increased from 0.028 at the 77 system stage to 7.854 at the 2 system stage. This process can be seen even more explicitly in Figure 3. It was decided, therefore, to use the greatest initial increase in error as a criterion for deciding how many separate group policies existed. Since the greatest initial increase in error occurred in the change from 4 groups (SE = 3.102) to 3 groups (SE = 4.859), the existence of 4 meaningful clusters was assumed.

A more complete picture of the grouping process may be obtained by examining Figure 4. This figure shows a sequential history of the grouping of policy equations. The final grouping (far right) is designated by a 1 indicating

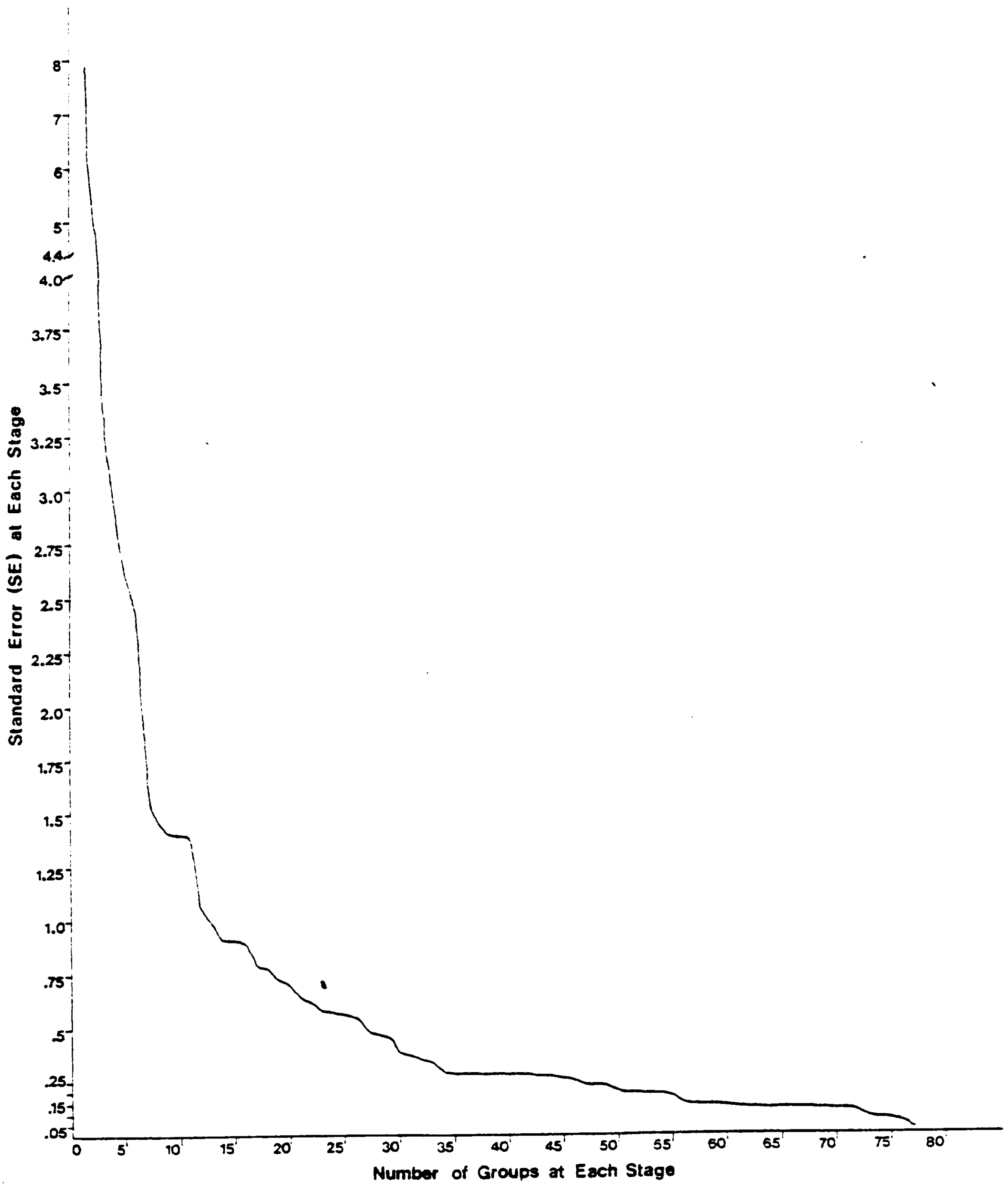

Figure 3. Standard error (SE) values as the number of groups of raters is systematically reduced from 78 to 1 by JAN II.

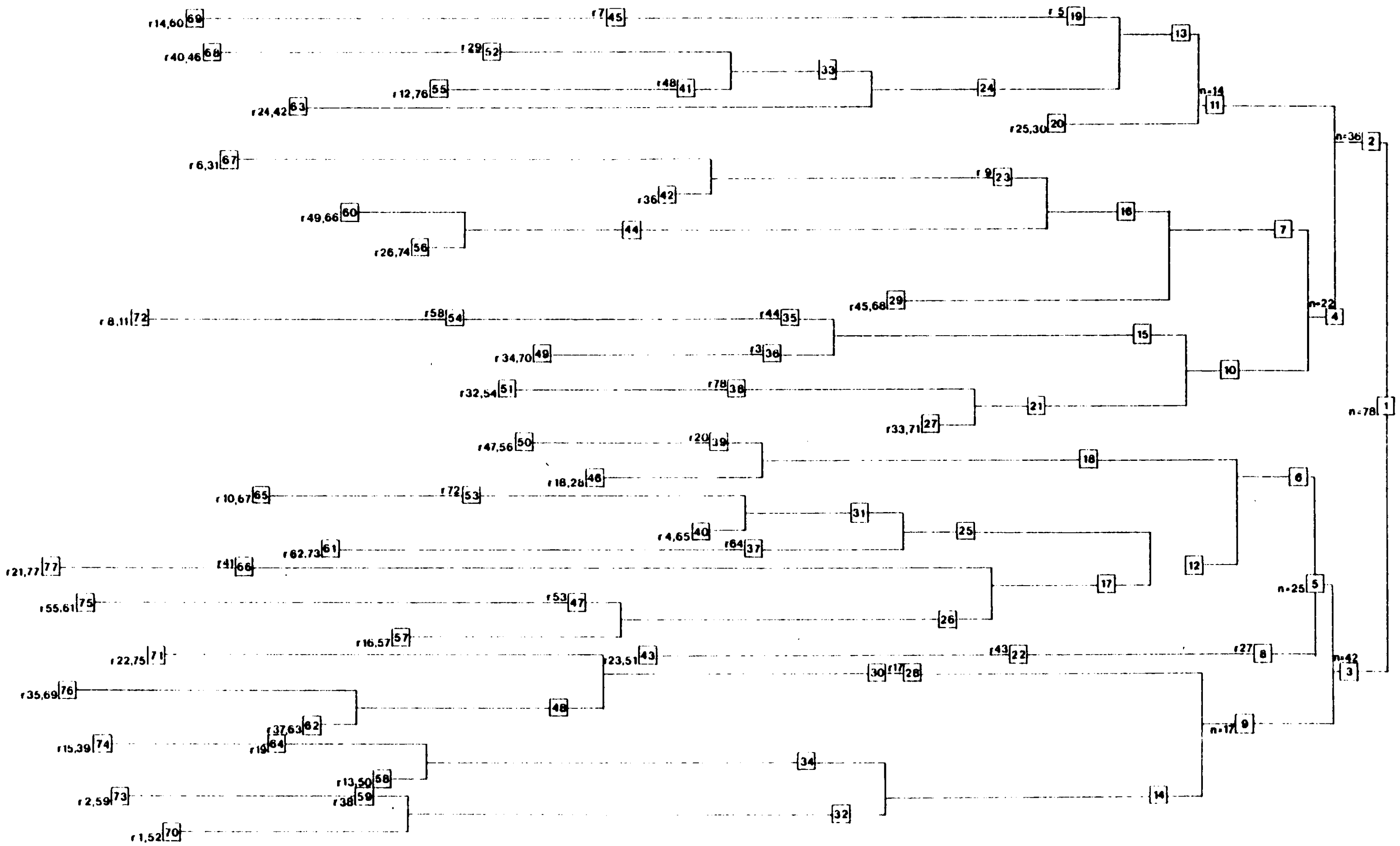

Figure 4. Sequential clustering (JAN II) of the 78 raters from 78 individual groupings to a single composite group (numbers above the boxes refer to individual raters).

that at this stage all raters were grouped into one composite. Following the lines leading to this final composite, one can see that the final grouping consisted of the joining together of a group of 36 raters with a group of 42 raters. This grouping process can be traced throughout the entire "tree" down to the first grouping indicated at the far left with the number 77. At this stage there were 77 policies, 76 single rater policies and one policy representing a pair of raters. Figure 4 also identifies the raters being grouped. For example, the first two raters to be combined were Raters 21 and 77. The combination of Raters 35 and 69 reduced the array to 76 policies, and so on.

Final Models

Once clusters of "homogeneous" raters have been identified, the final two steps in the PC procedure are to derive the composite linear models for each group and attempt to determine why these groups are different. The JAN software does not compute the final composite regression equations. Another statistical package or program must derive the final models. In this study, the stepwise regression procedure available in the SAS (Statistical Analysis System) package was used to generate the final group policies. The results, as shown in Table 6, indicate the existence of four separate group policies. For

Table 6. Final multiple regression models; groups based on JAN II clustering process.[a]

Group	Standardized beta weights[b]												
Categories	S	FV	G	B	M	MCA	UEV	VDL	UF	C	Constant	R^2	SS_E[c]
All Subjects	-0.09[e]	0.14	-0.18	0.23	-0.22	0.19	-0.09	0.21	0.08[e]	0.03[e]	4.47	0.09	33883.23
(n=78)[d]	(8)	(6)	(5)	(1)	(3)	(4)	(7)	(2)	(9)	(10)			
Group One	---	0.44	-0.31	0.28	-0.27	0.20	---	0.45	---	---	3.08	0.17	6331.70
(n=17)		(2)	(3)	(4)	(5)	(6)		(1)					
Group Two	---	---	-0.27	0.50	-0.38	0.53	-0.10	0.20	0.68	0.16	0.30	0.10	9846.37
(n=22)			(5)	(3)	(4)	(2)	(8)	(6)	(1)	(7)			
Group Three	-0.39	0.15	-0.25	---	---	0.15	-0.14	---	-0.18	-0.21	9.31	0.17	10019.56
(n=25)	(1)	(5)	(2)			(6)	(7)		(4)	(3)			
Group Four	0.22	-0.21	0.26	0.15[e]	-0.10	---	-0.10	---	---	0.26	4.19	0.05	5710.13
(N=14)	(4)	(5)	(2)	(6)	(3)		(7)			(1)			

[a] The numbers in parentheses represent the rank order of dimensions within each group.

[b] S=sky; FV=foreground vegetation; G=green colors; B=blue colors; M=mountains; MCA=man-changed area; UEV=unobstructed expanse of view; VDL=visible distant landforms; UF=undisturbed forest; C=clouds.

[c] Residual sum of squares.

[d] "n" refers to the number of subjects, not to the number of observations on which the regression models are based. This model is based on 7800 observations (78 subjects x 100 judgments/subject), the Group One model is based on 1700 observations, etc.

[e] $p > .05$.

instance, the first group is composed of 17 members whose policy reflects heavy reliance on visible distant landforms and foreground vegetation in making overall preference judgments. The members of Group 2 , on the other hand, attach more importance to undisturbed forest, man-changed area, and the proportion of blue color. Note that the regression coefficients of the models generated by stepwise regression are consistently smaller than the R^2 values produced by JAN II. This is because the former coefficients are based on raw score regression weights while the latter are based on standardized weights.

At this point it is worth noting that stepwise regression is a method of narrowing the range of possible candidate models to be used as a basis for data-fitting (Draper and Smith 1966). That is, there may be linear combinations of landscape dimensions other than those identified in Table 6 that fit the data equally as well. Also, there may be other models that are as valid from a prediction standpoint. The best model for data fitting is not necessarily the best for prediction. Basically, stepwise regression begins by introducing that independent variable, X1, which by itself gives the largest R^2. In the next step, the independent variable which gives the largest increase in R^2 in the presence of X1 is entered. However, at each stage, all variables currently in the model are subjected to partial F tests. Those independent variables found to be not significant are removed. This process terminates when no more variables can enter or be removed. Thus, there are two significance levels that the researcher must specify: one for variable entry and one for removal. In this study, both significance levels were set at 0.20 to allow ample opportunity for variables to enter and to be retained in the final model.

Stepwise regression also gives some idea of how much each predictor contributes to variance. That is , each time a predictor is added or deleted, one can observe the resulting changes in R^2. Table 6 reveals that stepwise regression permitted between 6 and 8 dimensions to enter the final models. However, in all cases, the first two, three or four variables entered accounted for most of the variance. For instance, a linear model consisting of 7 variables produces an R^2 of 0.17 for Group 3. However, the third step of the stepwise procedure resulted in a 3-variable model with an R^2 of 0.15. Thus, not much more variance is explained by going from a 3-variable to a 7-variable model. In fact for all the models shown in Table 6, the first 3 variables entered accounted for usually more than 60 percent of the variance explained by the full model. This is not too surprising since prediction is seldom improved by the addition of more than 5 variables (Claudy

1972). Again, this is partly the result of some of the dimensions being highly intercorrelated and provides useful information if the data for one or more independent variables is extremely expensive or time-consuming to collect.

Christal (1968b) contends that it is possible to have low interrater agreement among the entire sample but still be able to cluster raters into two or more groups displaying high agreement. Inspection of Table 6, however, indicates that this did not occur. The multiple R^2 for the full model for all 78 subjects is 0.09 indicating that this particular linear combination of variables accounted for only 9.0 percent of the variance in landscape preference judgments. Clustering the raters into separate groups led to an increase in R^2 in three out of four groups, but the amount of variance explained (10.0 to 17.0 percent) is still considerably lower than the amount explained in other landscape preference research using multiple regression. Likely explanations for these findings are (1) the appropriate landscape dimensions were not measured, (2) The subjects were being highly unreliable in their judgments, (3) the landscape dimensions were not measured accurately, (4) a nonlinear relationship between the dimensions and preference exists, or (5) some combination of all of these. These hypotheses will be explored in greater detail in the next chapter.

Group Differences

Table 7 summarizes the background variables associated with all 78 subjects and with the groups clustered using JAN II. An additional variable, "major," was also assessed but due to the wide variation in types of majors listed by subjects and the substantial number of subjects reporting that they had not decided on a major yet, this variable was not considered. In all cases, with the possible exception of sex for Group 1 (24% males, 76% females), none of the variables performs an adequate job of distinguishing among groups. For each group and all raters combined, males and females were nearly equally represented, over 63 percent of the subjects were either 18 or 19 years old, and over 75 percent were either freshmen or sophomores. None of these results are too surprising given that all subjects were from an introductory psychology class, a rather homogeneous sample. Also, Brush (1976), in his review of a large number of studies, reports that sex, age, education, occupation, and preferred leisure setting account for little variance in scenic beauty evaluations. Two background variables that seem to account for more variance than the ones just mentioned are familiarity with landscapes (Sonnenfeld 1966) and exposure to certain landscapes in childhood (Zube et al. 1974).

Table 7. Background characteristics of groups identified by JAN II (in proportions).

Group	Sex-		Age-					Year in School			
Categories	Males	Females	17	18	19	20	>20	Fr.	Soph.	Jr.	Sr.
All Raters (n=78)	.44	.56	.05	.46	.28	.08	.13	.54	.28	.12	.06
Group One (n=17)	.24	.76	.12	.29	.35	.06	.18	.47	.29	.12	.12
Group Two (n=22)	.50	.50	.00	.36	.41	.09	.14	.41	.41	.09	.09
Group Three (n=25)	.48	.52	.04	.64	.08	.12	.12	.64	.16	.16	.04
Group Four (N=14)	.50	.50	.07	.50	.36	.00	.07	.64	.29	.07	0

Early in the formulation of this study it was decided not to collect more background information than that reported since the primary objective was an investigation of the applicability of the PC procedure for landscape assessment and not the identification of group differences. However, for future landscape preference studies involving the use of PC, knowledge of raters' previous experiences with various landscapes may prove useful because, as Naylor and Wherry (1964) have found,

> JAN does an excellent job of sequential grouping...On the other hand, one finds it very difficult to make an intelligent decision about how many groups are desirable or even to obtain a feeling for what is going on in the grouping process simply by examining the JAN output without any supplementary information (p. 60).

In order to test for possible sex differences, the preference judgments of the 44 females and the 34 males were analyzed using the PC routine. Table 8 summarizes the results of this analysis. Females were generally a little less consistent than males in terms of within-rater reliability (R^2 before any clustering is done). However, the between-rater consistency (drop in R^2 from the separate policies stage to the single group stage) did not vary much according to sex. Thus, one can safely say that sex plays an inconsequential role in determining either individual or group differences in landscape preferences. This finding is in line with previous research reviewed above (Brush 1976).

Table 8. Summary of JAN II results for males and females.

Groups	Before any Combining-		After combining all groups into one -	
	R^2	Error	R^2	Error
Males (n=34)	0.38	0.05	0.12	6.71
Females (n=44)	0.34	0.03	0.11	3.11

Determining Sources of Variation

The purpose of this section is one of "data-snooping" in an attempt to discover (1) why raters were inconsistent and (2) why the final models accounted for such a small percentage in variance. Of course these two results are not mutually exclusive, since the small R^2 produced by the models is partially accounted for by the raters behaving in an inconsistent manner. For future landscape research involving PC, it is necessary to attempt to explain why these two phenomena occurred. One advantage of the JAN programs is that they are flexible enough to allow for the re-analysis of the same data in a variety of ways by simply changing 3 or 4 control statements.

Rater Inconsistency

One logical means of beginning the search for sources of error is to eliminate those raters (Phase 3 subjects) who were unreliable and recalculate the composite policy. In this particular investigation, there are two ways of looking at rater reliability: the R^2 values from JAN I and the correlation between ratings on the 5 duplicate slides. The first method for selecting reliable subjects is fairly simple: choose those subjects whose R^2 values exceed some criterion, 0.50, say.

The second method is also fairly simple: calculate the Pearson correlation (or coefficient of stability) between the ratings given the 5 slides at the beginning and the ratings given the 5 duplicates at the end of the tray of 100 and select those subjects whose correlations are above 0.50. For example, visual inspection of the raw data in Table 9 indicates that Rater 12 was fairly consistent and, in fact, the stability coefficient for Rater 12 is 0.98. However, the results for Raters 35 and 65 cast doubt on the validity of simple correlations as an indices of rater reliability. Scrutiny of Rater 35's ratings for the 5 slides and their duplicates reveals a great deal of consistency, but the coefficient of stability is 0.00 indicating a lack of consistency.

The results of Rater 65 lead to the opposite conclusion. Rater 65's ratings from Time 1 to Time 2 are highly erratic, yet the reliability index is 0.69. A possible explanation for these contradictory results is that, with 5 slides, only 3 degrees of freedom are used in the computation of the stability coefficients. Therefore, the coefficients are more likely to be mathematical artifacts than valid representations of the correlation between ratings given the 5 slides and their duplicates. It should be noted, however, that the correlations between the 5 slide pairs from Time 1 to Time 2 over all 78 subjects were as follows: Pair 1, 0.48; Pair 2, 0.58; Pair 3, 0.42; Pair 4, 0.51; Pair 5, 0.66. These results indicate that

Table 9. Reliabilities of three raters based on their judgments of 5 slides and duplicates of those slides.

Rater	Slide	Ratings on slides near beginning of tray (Time 1)	Ratings on duplicates of same slides near end of tray (Time 2)	Pearson Correlation
	1	3	3	
	2	7	6	
12	3	7	6	0.98
	4	6	5	
	5	8	6	
	1	6	5	
	2	6	7	
35	3	6	7	0.00
	4	5	6	
	5	6	5	
	1	4	1	
	2	8	4	
65	3	8	1	0.69
	4	9	5	
	5	8	4	

when all judges are considered together, the ratings given the five pairs of duplicate slides were moderately consistent from Time 1 to Time 2.

Since there were so many of these spurious correlations, it was decided to use a different index to compare the 5 slides with their duplicates. That index is the sum of absolute differences between the ratings assigned at Time 1 and those assigned at Time 2. To illustrate, the absolute differences in ratings between Time 1 and Time 2 for Rater 35 (Table 9) were 1 ($|6 - 5|$), 1 ($|6 - 7|$), 1 ($|6 - 7|$), 1 ($|5 - 6|$), and 1 ($|6 - 5|$) and the sum was 5 (1 + 1 + 1 + 1 + 1). This index has more face validity than the stability coefficient.

Table 10 shows the sums of absolute differences for each of the 78 subjects. A somewhat arbitrary cut-off value of 7 was imposed. Subjects whose sum of absolute differences in ratings between Time 1 and Time 2 was less than or equal to 7 were considered to be consistent while those whose similar value was greater than 7 were considered inconsistent. Using the cut-off value of 7 as a criterion, Table 10 indicates that 48 out of 78 (62%) of the judges were "reliable." Recall that only 17 out of 78 (22%) of the judges were considered reliable using an R^2 value (Table 2) of 0.50 as a cut-off value. However, it was considered more important in this part of the study to obtain a feeling for

Table 10. Sums of absolute differences between ratings assigned 5 slides near the beginning (Time 1) and the ratings assigned their duplicates near the end (Time 2) of the tray of 100.

Subject	Sum of Absolute Differences	Subject	Sum of Absolute Differences	Subject	Sum of Absolute Differences
1	5	27	4	53	10
2	10	28	7	54	5
3	9	29	2	55	10
4	9	30	5	56	6
5	7	31	6	57	12
6	11	32	9	58	5
7	4	33	6	59	6
8	4	34	2	60	7
9	7	35	5	61	9
10	5	36	5	62	7
11	4	37	8	63	10
12	5	38	3	64	8
13	7	39	2	65	22
14	4	40	10	66	7
15	24	41	9	67	4
16	11	42	8	68	6
17	17	43	10	69	13
18	5	44	5	70	5
19	6	45	7	71	10
20	20	46	4	72	29
21	2	47	14	73	5
22	5	48	7	74	10
23	11	49	4	75	6
24	6	50	2	76	7
25	3	51	12	77	6
26	8	52	8	78	5

the pre-test/post-test reliability of subjects than an average overall reliability. Thus, while the R^2 values obtained from JAN I seem to be more conservative estimates of overall rater reliability, the sum of absolute differences criterion seems better suited to the purpose of comparing differences in rater consistency between Time 1 and Time 2.

The sum of absolute differences criterion provides one means of comparing the reliabilities in overall preference ratings with the reliabilities in values assigned the 10 landscape dimensions. Recall that 10 separate groups of judges ranging from sample sizes of 22 to 24 rated each of the 100 slides according to amounts of the 10 dimensions. Table 11 displays the sums of absolute differences in ratings for these 10 groups. Notice that the percentages of absolute difference sums less than or equal to 7 range from 55 to 100 percent. If it were not for "clouds" the range would be from 74 to 100 percent. This indicates that, with the exception of "clouds", at least three-quarters of the subjects were highly reliable in their ratings of the various dimensions and the dimension ratings were more reliably performed than the overall preference ratings (percentage of absolute difference sums less than or equal to 7 = 62%). It is not known why so many subjects rating the amount of clouds were inconsistent; one can only assume that they did not understand the instructions.

Table 11. Proportions of sums of absolute differences less than or equal to 7, between ratings assigned 5 slides near the beginning (Time 1) and the ratings assigned their duplicates near the end (Time 2) of the tray of 100.

Dimension	Sample Size	Absolute difference sums less than or equal to 7 -	
		Number	Percentage (Number / Sample size x 100)
Sky	22	21	95
Foreground vegetation	22	17	77
Green	23	21	91
Blue	24	23	96
Mountains	22	21	95
Man-changed area	22	22	100
Unobstructed expanse of view	23	17	74
Visible Distant landforms	23	18	78
Undisturbed forest	22	20	91
Clouds	22	12	55
Landscape preference	78	48	62

Slide Variability

Although the 100 slides were randomly placed in the tray, there may be something about the manner in which the slides were presented that led to unexplained variability in the final regression models. For this reason, the effects of slide variability and order of presentation were investigated.

Several subjects indicated that they were uncertain of their ratings on the first few slides. They said that it took them a while to get used to the instructions and that they needed to view a few slides in order to obtain a basis of comparison for their remaining ratings. For these subjects, the practice slides apparently did not serve their function. Because of possible wide variability in ratings of the slides at the beginning of the tray, the ratings on the first 20 slides were eliminated from analysis and then rater policies were "recaptured" using only the last 80 slides.

To investigate the effects of possible fatigue and reliabilities in different portions of the study, the slides were also split in two ways: a 50-50 split and a 33-33-34 split. For each of these splits (i.e., first 50 slides, second 50, first 33, middle 33, and last 34), rater policies were re-captured.

Finally, those slides whose dimension ratings were highly variable were eliminated (Appendix E). As a criterion for selection, only those slides whose standard deviations on 5 or more dimensions were less than 1.00 were chosen. Using this as a criterion, 45 slides were selected. These slides are indicated by stars (*) in Appendix B. In general, with 8 or 9 exceptions, these slides were fairly uncomplicated scenes composed of mainly 4 or 5 dimensions, usually blue colors, clear sky, mountains, and foreground vegetation. The 8 or 9 slides that were more complex showed, in addition to the dimensions just mentioned, various degrees of human influence. The hypothesis here was that those scenes whose dimensions were rated in a fairly consistent manner would also be consistently rated in terms of overall preference; therefore, the re-captured policies would account for more variation in preference. Note also in Appendix B the high degree of variability in overall preference ratings for each slide. In general the standard deviations for the slides rated on preference are larger than the standard deviations associated with the same slides rated on one dimension at a time. This finding indicates the increased difficulty in making the global judgment over making judgments on one cue at a time.

Re-Capturing Rater Policies

The results of all the above analyses are compared in Tables 12, 13, and 14 and Figure 5. Table 12 reveals considerable variation in individual rater reliabilities within the different combinations of subjects and slides. Table 13 summarizes Table 12 by showing the percentage of judges in each combination of slides that is considered reliable (R^2 values > 0.50). A rather obvious result from Tables 12 and 13 is that the R^2's for all the various combinations of slides are greater than the R^2's associated with the 100 slides. For instance, when just the last 80 slides are considered, R^2's are generally higher than those for all 100 slides suggesting a moderate amount of inconsistency in ratings assigned the first 20 slides. Reliabilities from ratings of either the first 50 or last 50 scenes are fairly similar but substantially higher than those from ratings of all 100. The high percentage of reliable raters (41%) judging only the first 50 slides as compared to the percentage of reliable raters judging all 100 indicates the presence of a fatigue effect: judgments of the first 50 slides were more consistent than judgments of the last 50. However, this conclusion is somewhat negated by the 44 percent of reliable subjects that surface when the last 50 slides alone are considered. Thus, there appears to be both a fatigue effect (high variability in

Table 12. Rater reliabilities (R^2 values from JAN I) for various combinations of slides.

	Slide Groupings							
Rater	All 100	Last 80	First 50	Last 50	First 33	Middle 33	Last 34	45 slides chosen due to low variance*
1	.59	.62	.64	.78	.77	.77	.59	.65
2	.24	.23	.30	.33	.57	.49	.50	.35
3	.26	.26	.38	.42	.44	.38	.65	.36
4	.28	.29	.45	.35	.60	.41	.67	.40
5	.22	.22	.30	.41	.42	.30	.48	.37
6	.38	.42	.51	.38	.60	.46	.48	.58
7	.22	.16	.44	.31	.66	.39	.55	.36
8	.16	.15	.30	.31	.47	.21	.49	.28
9	.32	.29	.64	.20	.84	.59	.32	.22
10	.31	.32	.45	.43	.72	.44	.48	.25
11	.25	.31	.39	.34	.56	.58	.40	.30
12	.20	.24	.36	.16	.55	.40	.23	.34
13	.29	.36	.33	.54	.44	.60	.52	.40
14	.23	.22	.48	.23	.62	.57	.29	.32
15	.49	.57	.53	.64	.56	.58	.75	.52
16	.44	.55	.36	.65	.44	.57	.71	.52
17	.44	.46	.49	.56	.56	.65	.56	.60
18	.41	.44	.50	.48	.63	.49	.52	.53
19	.49	.51	.65	.51	.82	.66	.51	.58
20	.57	.67	.52	.72	.65	.74	.63	.61
21	.45	.51	.52	.62	.76	.66	.65	.60
22	.28	.28	.55	.34	.68	.49	.25	.18
23	.49	.54	.52	.61	.66	.73	.55	.48
24	.33	.34	.54	.33	.71	.56	.28	.43
25	.32	.24	.44	.31	.64	.62	.44	.44
26	.35	.35	.38	.51	.49	.23	.64	.44
27	.59	.62	.65	.61	.65	.68	.68	.60
28	.12	.19	.30	.35	.53	.41	.42	.32
29	.16	.18	.34	.24	.65	.60	.29	.39
30	.42	.42	.54	.43	.62	.56	.42	.34
31	.21	.22	.29	.26	.60	.32	.21	.44
32	.26	.25	.45	.36	.76	.33	.35	.40
33	.27	.24	.38	.30	.61	.41	.61	.46
34	.46	.48	.42	.56	.61	.68	.40	.62
35	.33	.44	.32	.53	.55	.39	.72	.42
36	.46	.46	.53	.53	.80	.58	.49	.58
37	.13	.20	.25	.20	.22	.48	.34	.38
38	.42	.44	.38	.57	.57	.42	.55	.45
39	.22	.26	.40	.27	.60	.59	.35	.14
40	.08	.18	.25	.34	.57	.55	.40	.25
41	.36	.34	.61	.41	.74	.51	.27	.44

(continued)

Table 12. Rater reliabilities (R^2 values from JAN I) for various combinations of slides (continued).

	Slide Groupings						
Rater	All 100	Last 80	First 50	Last 50	First 33	Middle 33	Last 34	45 slides chosen due to low variance*
42	.18	.30	.25	.31	.43	.52	.43	.24
43	.22	.26	.37	.23	.40	.64	.40	.27
44	.53	.50	.69	.45	.78	.62	.43	.43
45	.34	.39	.42	.42	.62	.32	.58	.42
46	.18	.26	.26	.34	.35	.46	.26	.37
47	.24	.25	.34	.31	.37	.33	.32	.38
48	.50	.56	.55	.62	.68	.71	.62	.36
49	.65	.66	.70	.69	.74	.78	.72	.66
50	.53	.56	.59	.60	.70	.65	.63	.58
51	.22	.15	.46	.22	.66	.36	.26	.37
52	.61	.73	.45	.77	.43	.78	.72	.75
53	.33	.42	.38	.52	.40	.63	.68	.31
54	.46	.56	.40	.69	.48	.63	.68	.63
55	.46	.51	.57	.55	.66	.46	.53	.63
56	.26	.27	.52	.29	.69	.54	.46	.25
57	.61	.63	.62	.72	.76	.87	.68	.74
58	.32	.31	.39	.49	.56	.62	.45	.46
59	.22	.37	.21	.49	.33	.51	.49	.38
60	.28	.25	.47	.22	.72	.57	.16	.30
61	.52	.53	.55	.61	.64	.73	.55	.47
62	.57	.59	.62	.64	.70	.78	.59	.63
63	.29	.44	.32	.63	.46	.42	.62	.48
64	.57	.62	.61	.63	.61	.84	.63	.64
65	.08	.09	.44	.22	.55	.52	.34	.14
66	.58	.63	.56	.68	.66	.72	.80	.73
67	.34	.40	.47	.38	.45	.67	.42	.33
68	.50	.49	.54	.56	.65	.71	.56	.61
69	.27	.37	.24	.56	.49	.55	.58	.31
70	.11	.10	.30	.17	.53	.28	.29	.32
71	.25	.29	.26	.37	.30	.56	.49	.31
72	.39	.42	.53	.63	.57	.69	.54	.48
73	.65	.69	.57	.74	.62	.67	.70	.70
74	.41	.41	.51	.45	.72	.60	.54	.49
75	.20	.35	.25	.47	.42	.40	.54	.29
76	.29	.30	.33	.42	.47	.16	.56	.33
77	.66	.67	.73	.67	.80	.72	.75	.74
78	.55	.54	.65	.57	.73	.78	.59	.62

*These slides had standard deviations of 1.00 or less on 5 or more of the 10 dimensions (see Appendix B).

Table 13. Number and percentage of reliable (R^2 values ≥ 0.50) raters in each of the various combinations of slides shown in Table 12.

Slide Combinations	Number	Percent (Number/78 x 100)
All 100	17	22
Last 80	23	29
First 50	32	41
Last 50	34	44
First 33	57	73
Middle 33	50	64
Last 34	42	54
45 slides chosen due to low variance*	25	32

*These slides had standard deviations of 1.00 or less on 5 or more of the 10 dimensions (see Appendix B).

Table 14. Summary of JAN II results (change in predictive efficiency, R^2) for various combinations of slides and subjects.[a]

Combinations	Before any combining R^2	Before any combining SE[b]	After combining all groups into one R^2	After combining all groups into one SE[b]
All 100 slides	0.36	0.03	0.11	7.85
Last 80 slides	0.39	0.04	0.13	9.40
First 50 slides	0.45	0.08	0.16	12.30
Last 50 slides	0.46	0.08	0.14	8.61
First 33 slides only	0.59	0.16	0.24	18.05
Middle 33 slides only	0.55	0.19	0.17	23.59
Last 34 slides only	0.50	0.09	0.13	37.09
45 slides chosen due to low variance[c]	0.44	0.11	0.11	21.69
All 100 slides, "reliable" subjects only (n=48)[d]	0.36	0.03	0.10	6.19
First 33 slides, "reliable" subjects only (n=48)[d]	0.62	0.16	0.26	10.77
All 100 slides, 17 subjects only[e]	0.58	0.10	0.34	1.90

[a] Unless otherwise indicated, these data apply to a subject sample size of 78

[b] Standard error of the estimate before the last two groups are combined into one (JAN II does not compute the error term for the single policy system).

[c] These slides had standard deviations of 1.00 or less on 5 or more of the 10 dimensions (see Appendix B).

[d] Subjects whose sums of absolute differences between ratings given the 5 duplicates at Time 1 and Time 2 were $\leq$ 7 (see Table 10).

[e] Subjects whose R^2 values (JAN I) were $\geq$ 0.50 (see Table 2).

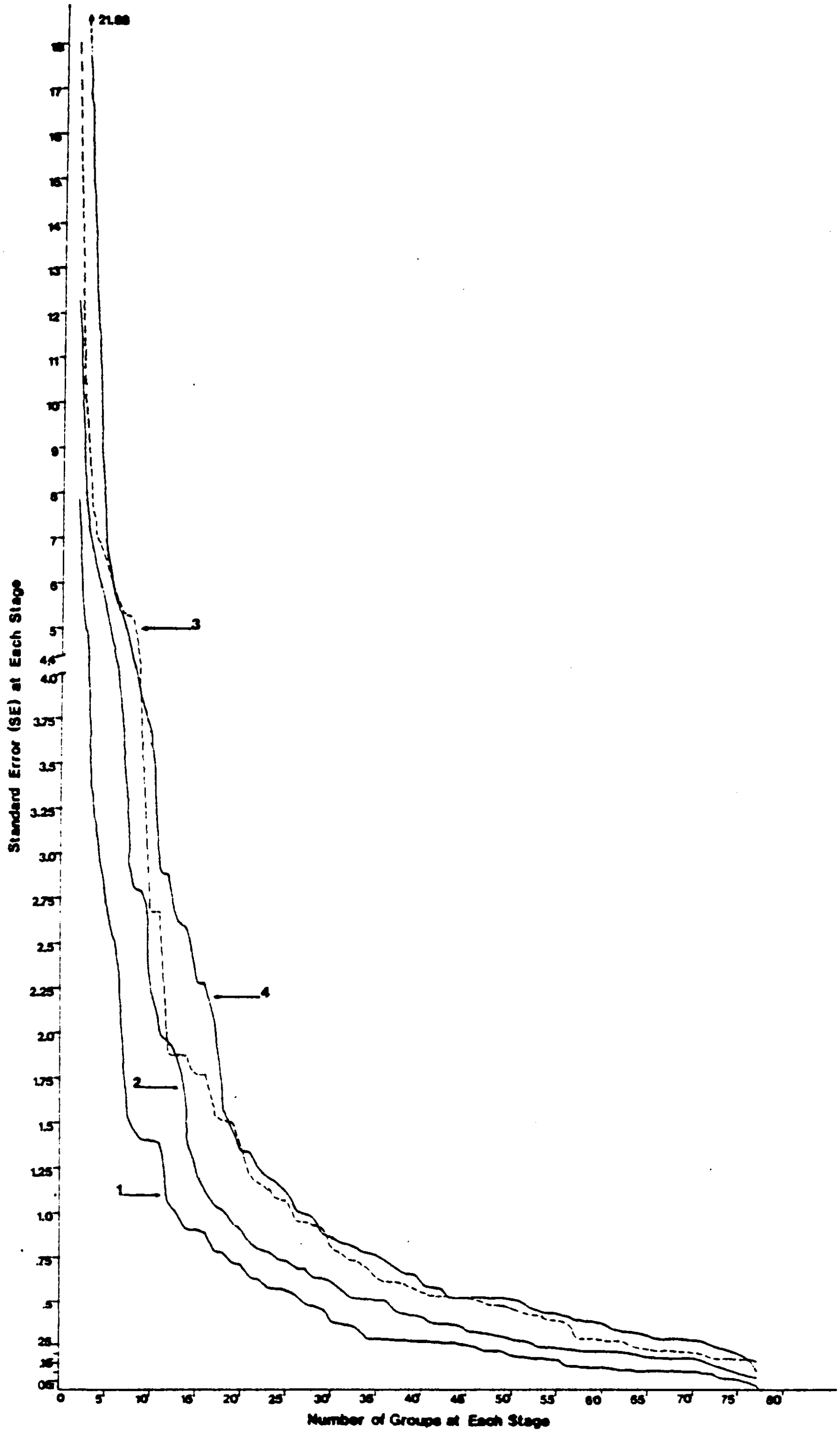

Figure 5. Standard error (SE) values, for various combinations of slides, as the number of groups of raters is systematically reduced from 78 to 1 by JAN II.

Key:
1. All 100 slides (same as Figure 3)
2. First 50 slides
3. First 33 slides
4. 45 slides chosen due to Low Variance on 5 or more dimensions

ratings of slides toward the end of the tray) and an order effect (high variability in ratings of slides toward the beginning of the tray) influencing rater consistency. This hypothesis is further corroborated by observing the R^2's associated with the first 33 and last 34 slides. Seventy-three percent of the subjects are considered reliable when making their ratings on the first 33 slides only. This is quite an improvement over the "All 100 Slides" category and is in line with the percentage of reliable subjects found in other PC studies. Again, the substantial percentage (54%) of reliable subjects in the "Last 34 Slides" category indicates the presence of an order effect when compared to the "All 100 Slides" category. The declining percentage of reliable subjects going from the first to the middle to the last third of slides, however, reveals the dominance of the fatigue effect over the order effect (73% vs 64% vs 54% reliable subjects).

When subject policies were re-captured using only the ratings on the 45 slides chosen by low variance on 5 or more dimensions, individual reliabilities were generally higher than those associated with all 100 slides. However, the increase in number of reliable subjects was not substantial. Hence, there seems to be only a moderate influence on rater consistency due to variability in how the 10 landscape dimensions were rated.

Re-Clustering Raters

The tables in Appendix C present the change in error and the change in composite predictive efficiency at each stage of the JAN II grouping process for the various combinations of data described in this chapter. Table 14 summarizes the results of these analyses. Recall that the "Reliable Subjects Only" category refers to those subjects selected on the basis of their absolute difference sums for 5 slides and corresponding duplicates. The results in Table 14 generally parallel those in Tables 12 and 13. For example, the R^2 values obtained from ratings of all 100 slides and the last 80 slides at the 78-separate policy stage were lower than the R^2 values for any of the other categories, except for one: the category in which all 100 slides were used with only those subjects considered to be reliable also displays a fairly low R^2 before any combining. This indicates either that subject inconsistency had little effect on the overall composite strategies or that the criterion for selecting reliable subjects was not a good one. The final policy of the 17 subjects whose individual reliabilities (JAN I) were 0.50 or greater yields a R^2 of 0.335, the largest of all R^2 values at the one-policy stage. On the other hand, the one-policy R^2 value for the 48 subjects chosen according to the sum of absolute differences criterion is only 0.10. This suggests that the method of selecting "reliable" subjects on the basis of their consistency in rating the duplicate slides may have been ineffective in eliminating those raters who were highly unreliable.

Table 14 also emphasizes the superiority of judgments made during the first third of the slides. That is, when considering only overall preference judgments made on the first 33 slides, approximately 59 percent of the variance is predicted by using 78 separate policies; about 24 percent of the variance is predicted when only one policy is used. When combining the judgments made on the first 33 slides by the "reliable" subjects only, the respective R^2 values before and after grouping were 0.62 and 0.26. The important point to note here is that no matter at which stage of the clustering process policies are observed, the R^2's associated with the first 33 slides are larger than those associated with any other combination of slides. Again, these findings show the existence of a fatigue or other type of effect on ratings made near the end of the tray of slides.

Note also in Table 14 that the R^2 values before any grouping is done are moderately large (0.36 to 0.62) and clearly in line with the R^2's obtained in other landscape preference studies when subjects were not asked to cue on specific dimensions (e.g., see Buhyoff and Leuschner 1978).

Figure 5 graphically displays the increases in error at each stage of the clustering process for the following combinations: all 100 slides, first 50 slides, first 33 slides, and 45 slides chosen by low variance. These graphs show the rapid increases in error occurring near the end of the grouping process, suggesting the existence of more than one homogeneous cluster. Using the same criteria as before (p. 68) for determining how many meaningful groups exist, one can say that the "first 33 slides," "reliable subjects only," and "first 33 slides and reliable subjects combined" categories consist of 4, 3, and 5 homogeneous groups, respectively.

The final regression models for these various groups in each category are summarized in Tables 15, 16, and 17. In some cases stepwise regression yielded a reduced model whose R^2 was considerably smaller than the R^2 of the corresponding full model. When this situation occurred, the model with the larger R^2 was reported and identified in the table. Not all of the categories in Table 14 were analyzed in this manner because the information provided would have been redundant. Note in Table 15 the clear superiority of man-changed area, blue colors, and green colors in clustered policies across all groups except Group 4. Table 15 reveals that groups viewing the first 33 slides were fairly similar in their policies; this finding is consistent with the

Table 15. Final multiple regression models for first 33 slides only; groups based on JAN II clustering process (Appendix C).[a]

Group	Standardized beta weights[b]												
Categories	S	FV	G	B	M	MCA	UEV	VDL	UF	C	Constant	R^2	SS_E[c]
All Subjects (n=78)[d]	---	0.15 (7)	0.41 (3)	0.80 (2)	---	0.96 (1)	-0.07 (8)	-0.18 (6)	0.20 (5)	0.28 (4)	-3.31	0.17	10292.48
Group One (n=33)	---	0.36 (3)	0.28 (4)	0.75 (2)	---	0.82 (1)	---	---	---	0.24 (5)	-3.19	0.20	3792.82
Group Two (n=29)	0.38 (7)	---	0.55 (4)	0.95 (2)	-0.43 (6)	1.29 (1)	---	---	0.75 (3)	0.54 (5)	-9.29	0.22	3567.58
Group Three (n=12)	---	---	---	---	0.71 (3)	0.86 (2)	-0.50 (4)	-0.89 (1)	---	---	7.40	0.24	1800.79
Group Four[e] (n=4)	1.13 (1)	0.86 (4)	0.53 (6)	0.30[f] (7)	0.08[f] (8)	0.06[f] (9)	-0.03[f] (10)	0.89 (3)	-1.11 (2)	0.62 (5)	-6.24	0.25	472.42

[a] The numbers in parentheses represent the rank order of dimensions within each group.

[b] S=sky; FV=foreground vegetation; G=green colors; B=blue colors; M=mountains; MCA=man-changed area; UEV=unobstructed expanse of view; VDL=visible distant landforms; UF=undisturbed forest; C=clouds.

[c] Residual sum of squares.

[d] "n" refers to the number of subjects, not to the number of observations on which the regression models are based. This model is based on 7800 observations (78 subjects x 100 judgments/subject), the Group One model is based on 3300 observations, etc.

[e] Full model - not based on stepwise regression.

[f] $p > .05$.

Table 16. Final multiple regression models for reliable subject only (n=48); groups based on JAN II clustering process (Appendix C).[a]

Group Categories	Standardized beta weights[b]												
	S	FV	G	B	M	MCA	UEV	VDL	UF	C	Constant	R^2	SS_E[c]
All Subjects	---	0.14	-0.10	0.20	-0.30	---	-0.10	0.29	---	0.11	4.53	0.08	19068.61
(n=48)[d]		(4)	(6)	(3)	(1)		(7)	(2)		(5)			
Group One	-0.12[e]	0.38	-0.35	---	-0.27	0.24	---	0.30	---	-0.08	5.45	0.21	6620.83
(n=20)	(6)	(1)	(2)		(4)	(5)		(3)		(7)			
Group Two	0.30	---	---	0.25	-0.35	---	-0.11	0.30	0.20	0.30	1.76	0.06	8355.97
(n=21)	(3)			(4)	(1)		(7)	(4)	(6)	(2)			
Group Three	-0.58	-0.25	---	0.13	-0.21	---	-0.15	---	---	---	7.33	0.11	2916.31
(n=7)	(1)	(2)		(5)	(3)		(4)						

[a] The numbers in parentheses represent the rank order of dimensions within each group.

[b] S=sky; FV=foreground vegetation; G=green colors; B=blue colors; M=mountains; MCA=man-changed area; UEV=unobstructed expanse of view; VDL=visible distant landforms; UF=undisturbed forest; C=clouds.

[c] Residual sum of squares.

[d] "n" refers to the number of subjects, not to the number of observations on which the regression models are based. This model is based on 4800 observations (48 subjects x 100 judgment/subject), the Group One model is based on 2000 observations, etc.

[e] $p > .05$.

Table 17. Final multiple regression models for first 33 slides and reliable subjects combined (n=48); groups based on JAN II clustering process (Appendix C).[a]

Group Categories	Standardized beta weights[b]										Constant	R^2	SS_E[c]
	S	FV	G	B	M	MCA	UEV	VDL	UF	C			
All Subjects (n=48)[d]	0.18[f] (6)	0.28 (5)	0.44 (3)	0.77 (2)	---	0.80 (1)	---	---	---	0.39 (4)	-4.76	0.17	5879.15
Group One (n=17)	---	0.40 (3)	0.25 (5)	0.74 (2)	---	0.81 (1)	---	---	---	0.28 (4)	-3.25	0.22	1829.80
Group Two (n=16)	---	-0.33 (8)	1.01 (2)	1.20 (1)	-0.40 (6)	0.99 (3)	---	-0.34 (7)	0.53 (4)	0.50 (5)	-6.51	0.27	1744.75
Group Three (n=6)	0.37[f] (5)	0.16[f] (7)	---	0.86 (3)	-0.30[f] (6)	1.70 (1)	---	---	1.07 (2)	0.52 (4)	-10.67	0.36	495.01
Group Four[e] (n=4)	1.13 (1)	0.86 (4)	0.53 (6)	0.30[f] (7)	0.08[f] (8)	0.06[f] (9)	-0.02[f] (10)	0.89 (3)	-1.11 (2)	0.62 (5)	-6.24	0.25	472.42
Group Five (n=5)	---	---	---	0.37 (4)	0.49 (3)	---	-0.57 (2)	-0.75 (1)	-0.33[f] (5)	---	13.15	0.21	669.21

[a] The numbers in parentheses represent the rank order of dimensions within each group.

[b] S=sky; FV=foreground vegetation; G=green colors; B=blue colors; M=mountains; MCA=man-changed area; UEV=unobstructed expanse of view; VDL=visible distant landforms; UF=undisturbed forest; C=clouds.

[c] Residual sum of squares.

[d] "n" refers to the number of subjects, not to the number of observations on which the regression models are based. This model is based on 4800 observations (48 subjects x 100 judgments/subject), the Group One model is based on 1700 observations, etc.

[e] Full model - not based on stepwise regression.

[f] $p > .05$.

results shown in Tables 12, 13, and 14. Similar consistency is not observed in Table 16 where policies are noticeably different from group to group. Again note, in Table 17, the dominance of man-changed area, blue colors, and green colors (except for the last 3 groups where sample sizes are very small) when the 45 "reliable" subjects are combined with the first 33 slides.

The overwhelming finding that emerges from Tables 15 - 17, however, is that no matter how raters or slides are combined, the composite regression models account for only a small amount of variation in landscape preference. Even when the preference ratings of the 17 most reliable subjects were used as values of landscape preference, the resulting stepwise regression model accounted for only 26.1 percent of the variance.

Examining Cue Intercorrelations

In light of the multicollinearity exhibited in Table 4, a final PC analysis was conducted to see if more variance in preference could be explained by constructing a composite model using only those independent variables with minor correlations (< 0.45). The four dimensions meeting such a criterion were: foreground vegetation, blue colors, mountains, and unobstructed expanse of view. However, the JAN I analysis resulted in only one rater's having a reliability of greater than 0.50. The composite R^2 from JAN

II clustering was only 0.238 at the 78 individual policy stage and 0.090 after all groups were combined into one. The final 2-variable (mountains and unobstructed expanse of view) regression model accounted for only about 7 percent of the variance in landscape preference. Thus, not only were individual raters fairly inconsistent in their judgments, but the final model did a rather poor job of data fitting. These results parallel research of Dudycha and Naylor (1966) and Schenk and Naylor (1968) who found that decreasing the intercorrelations among regressor variables led to decreases in composite predictive efficiency (R^2). Their warning was that the results of any PC study involving high correlations among regressor variables must be interpreted in light of the effects of multicollinearity.

Chapter 5

DISCUSSION

In their excellent review of the use of unobtrusive measures in the behavioral sciences, Webb and his colleagues (1966) contend that one of the hurdles to overcome on the way to making valid comparisons is the ruling out of "plausible rival hypotheses." One of the "plausible rival hypotheses" that can readily be discarded in this study is the contention that the linear model is inappropriate and, hence, the data may be better fit by some nonlinear function. One of the restrictions of nonlinear regression is that a likely functional form of the model and the starting values for the parameters must be specified before the actual analysis can take place (Draper and Smith 1966). This is not too great of a restriction in some fields where well-formulated theories supported by a substantial amount of empirical evidence exist. However, in areas such as landscape assessment, where theory is still a distant goal and appropriate variables have not yet been identified, the restrictions of nonlinear regression may pose a problem especially when, as in this study, the model is multidimensional. The aim of preference modeling thus far is empirical data fitting, not rational model building.

Yet there are means of testing for nonlinearity of cue utilization. One of the most obvious and simplest means is to plot the data. Since it is physically impossible to plot data in ten dimensions, values for each of the landscape dimensions used in this study were plotted against landscape preference ratings. However, when the data were plotted in this manner, a "shotgun" pattern of data points on each of the 10 graphs resulted. Such a scattering of data is really not too surprising considering (1) the large residual sums of squares for the linear regression models, (2) the low first-order correlations between landscape preference and each of the dimensions (Table 4), and (3) the substantial standard deviations in preference ratings associated with each of the 100 slides (Appendix B), and (4) the moderately low R^2's resulting from each judge's policy equation (Table 2). In brief, the data were too dispersed to even begin an attempt at specification of a functional form.

Another means of testing for nonlinearity is to enter all of the interactive (e.g., $X1 \cdot X2$) and power terms (e.g., $X1^2$) and observe which of these variables are retained in the final regression model. In other words, see if the addition of more complex terms into the basic linear model generates new models that are more accurate representations of the judge's mental process than was the original linear one (Goldberg 1968). With 10 dimensions, there are N(N -

1)/2 or 45 possible interaction terms. When these interaction terms along with the squared terms for each of the dimensions were regressed onto the preference ratings in a stepwise fashion, the resulting model (composed of 8 dimensions, 36 interactive terms, and 7 squared terms) explained 18.2 percent of the variance. Thus, there was an almost inconsequential increase in R^2 over the linear models reported in Table 6. It should be noted that the first variable entering the model was an interactive term (unobstructed expanse of view x undisturbed forest) which accounted for almost 8 percent or nearly half of the 18 percent of variance in the final stepwise model. This suggests the existence of a slight nonlinear effect.

For all these reasons plus the clear superiority of linear analysis in the literature on human judgment (Goldberg 1968; Dawes and Corrigan 1974), it was decided that the simple, additive model was most appropriate <u>for this study</u>. The emphasis in the previous sentence is a mild caution that there does exist some evidence in both the human judgment (Slovic 1969) and landscape assessment (Shafer et al. 1969; Buhyoff and Leuschner 1978) literature for the use of cues in a curvilinear or configural manner. Thus, future studies of landscape preference should consider the possible effects of nonlinear cue utilization.

Individual Variation

Despite these justifications for using the linear model, the fact remains that the regression equations generated in this investigation failed to explain more than 10 to 26 percent of the variance in landscape preference. For future users of the PC procedure, the possible reasons for these low R^2 values must be addressed. First of all, it must be emphasized that low R^2's are not necessarily indicative of a poor rational model, only a poor empirical one. The author agrees with Goldberg (1968) that

> the test of the model is not how well it works as a representation of the state of the world (e.g., how well it predicts landscape preference) but rather how well it predicts the inferential judgments of the judge himself.

This study was concerned with how individual raters assessed information about 10 dimensions in order to form an overall evaluation of landscape preference. The results of the rater analyses are consistent with those of other judgment studies utilizing the PC approach. In their review of a number of PC studies, Zedeck and Kafry (1977, p. 287) note that "the linear regression model has predicted judgments fairly consistently and to a reasonably high degree as indicated by R^2 values ranging from 0.50 to 0.80." In this study the R^2's ranged from 0.34 to 0.62 when separate policies were derived and from 0.10 to 0.34 when composite policies were derived. In many instances the reliabilities (R^2's) for individual policy equations were fairly high (0.50 to 0.80). Thus, the linear regression equation can explain rater policies.

The value of PC is that it allows the inference of a judge's weighting policy by requesting global evaluations of a total stimulus rather than evaluations of individual elements (Zedeck and Kafry 1977; Goldberg 1968). PC analysis is focused on the judgment of the rater rather than on raters' own interpretations of their decision-making processes.

The cues that a rater uses in making a judgment are obviously not equally important and raters do not rate them equally. The purpose of using a linear model is to make the judge's policy explicit. A review of Tables 2 and 3 reveals considerable individual variation; there are at least 2 raters for whom each of the 10 dimensions was most important. This degree of individual variation seems to run counter to previous landscape preference studies. For instance, Fabos (1971) describes a study by Craik (1970) which showed impressive agreement among forestry and conservation students and faculty, landscape architect students and faculty, U.S. Forest Service personnel, and general university students in how they rated the dimensions of various landscape scenes. Similar agreement has been observed among individuals in a variety of subgroups from professional designers to high school students and housewives (Zube et al. 1974; Zube 1974; Wenger and Videbeck 1969; Riesenman 1977; Buhyoff and Leuschner 1978; Buhyoff and Riesenman 1979; Shafer et al. 1969; Coughlin and Goldstein 1970; and Kaplan et al. 1972). However, the degree of individual variation found in this study suggests that the careless adoption of one multiple regression model to represent all raters is bound to lead to contradictions among individual outcomes (e.g., landscape preferences). The PC procedure identifies individual differences and thereby provides a starting point for resolution of this conflict.

This emphasis on individualization is of particular value to those interested in both landscape management and research. For the researcher, PC offers a conceptual basis and methodology for the study of individual differences in landscape preference, the lack of which has been a strong criticism of past research in this field (see p. 16). For resource managers or planners, PC provides a means of gaining self-insight into the reasons for their decisions involving landscape manipulation. For example, Hammond and his associates (1977) have developed an interactive computer system, based partially on the results of PC analysis, which helps individuals evaluate management objectives and planning criteria. Finally, the clustering process enables resource decision-makers to evaluate the policies used by various pressure groups (e.g., the Sierra Club, stripmining advocates, etc.) to arrive at their conclusions concerning management alternatives involving landscape manipulation. Examination of Table 6 reveals the existence of four separate clusters of raters exhibiting various policies. Unfortunately, insufficient background information needed to identify groups was collected. However, the potential still exists within the PC framework for identifying meaningful clusters of raters with similar landscape preference strategies.

Methodological Problems

From their review of PC literature, Slovic and Lichtenstein (1971) conclude that, when asked to evaluate subjectively their own decision processes, judges usually overestimate the importance they place on minor variables and underestimate their reliance on a few major variables. In most of the PC studies reviewed by Slovic and Lichtenstein three variables explained over 80 percent of the variance and one variable explained about 40 percent of the variance. The results of the stepwise regression analyses in this research (Tables 6, 15, 16, 17) are consistent with Slovic and Lichtenstein's conclusions in one respect: most of the variance in the models is explained by one to three variables. The failure of the composite models to account for more than 10 to 27 percent of the predictable variance is not weakness of the PC paradigm but a function of either (1) the amount of individual variation exhibited in this study, (2) the method in which the data were collected, or (3) both.

Dimension Selection

First, another plausible rival hypothesis for the failure of the final model to capture more of the predictable variance is that the wrong independent variables were used. It is realized that the 10 dimensions used are only a subset of the total number of variables that could influence a person's landscape preference. Yet, as previously stated, these variables appeared to be the most valid for the study area involved and there was some empirical evidence for their selection (Harvey 1977). In addition, research by Brush and Shafer (1975), Zube et al. (1974), Carls (1974), Craik (1972a), and Peterson and Neumann (1969) indicates that most of the variables used in this study were found to influence people's landscape preferences in other environments. Thus, failure to employ proper independent variables is not considered to be a serious plausible rival hypothesis.

One variable which may have been important but was given little consideration is photographic quality. Critics of landscape preference research (e.g., Kreimer 1977) have contended that the failure to consider this variable is a shortcoming of such studies. While the author realizes the significance of photographic quality, it must be recognized that realism was a goal of this study. Views from the Blue Ridge Parkway are often hampered by the presence of haze, especially during spring and summer when most of the slides were taken. This haze made perhaps 20 percent of the slides appear rather drab and not very appealing. A haze filter could have been used when the slides were taken, but again the goal in this study was to present random scenes from the BRP environment and not to present representations that had been chosen a priori according to accepted standards of photographic composition. Hence, the effects of photographic quality are recognized but not considered to be a serious limitation.

Slides as Profiles

The emphasis on realism may lead one to question the use of slides. Recall that one of the criticisms of landscape preference research is that photographs or slides are not valid representations of actual landscapes (p. 16). Yet numerous studies provide evidence that people evaluate representations of landscapes in the same manner in which they evaluate actuasl scenes (Coughlin and Goldstein 1970; Shafer and Richards 1974; Zube et al. 1974; Daniel and Boster 1976; Jackson et al. 1978). In any event it would be wise to conduct an on-the-ground validation of any model developed from judgments of slides or photographs before the model is accepted as representative of a given environment.

Though some critics of landscape preference research may argue that slides are not realistic stimuli, the same cannot be said of the use of slides in PC research. In all of the research on human judgment involving PC analysis, the stimuli presented for overall evaluation are either numerical or verbal profiles. An example of a numerical profile was given in Figure 2 (p. 31). Such profiles are stimuli on which all of the cues are presented with their corresponding numerical values. Verbal profiles, on the other hand, consist of written paragraphs in which the value of each cue is described qualitatively. Figure 6 is an example of a verbal profile taken directly from Anderson's (1977) study of rater policies used in assessing teacher quality. Anderson found that high school teachers were significantly more consistent in rating numerical than verbal profiles. She reasoned that it is probably easier to make judgments when cues are reduced to numerical values and that judgments of verbal statements correspond more closely to real-life decisions than evaluations of numerical scales. There are, of course, more sources of variation in real-life situations and this probably explains why people would be less consistent in rating verbal than numerical profiles.

Profile No. 2 Your Rating ___

He prefers not to establish any objectives or plans for his course. His homework assignments are neither very difficult or frequent. He sometimes does not seem to understand the subject he teaches. He shows almost no interest in or concern for individual students. His explanations of material are of moderate clarity. He very seldom involves students in classroom discussions. The amount of enthusiasm expressed in his teaching varies from day to day. If a student starts out poorly, he is sure to get a poor grade in this teacher's course.

Figure 6. Example of a verbal profile (from Anderson 1977).

To this author's knowledge, pictorial profiles (e.g., landscape slides) have never been used to assess human judgment in a study involving PC analysis. Extending Anderson's conclusions one step further one can reasonably argue that the pictorial profile is even more representative of real-life than the verbal profile, especially in light of the fact that a significant portion of the information used by humans for processing and arriving at decisions is obtained visually. Slides of landscapes contain additional sources of variation over verbal profiles: they simulate more closely the cues necessary for people to make landscape preference judgments. The cues are not manipulated, scaled, numbered, or described by the researcher but exist visually in a manner closely resembling reality. This is the reason for the notion stated earlier that landscape assessment critics and human judgment researchers view the reality of pictorial representations of landscapes from opposite ends of a spectrum. Landscape assessment researchers are concerned that slides may not validly represent reality while those involved in PC analysis are concerned that increasing profile realism may lead to additional sources of variation that cannot be accounted for by the linear model. If one accepts Brunswik's plea for a life-relevant theory of human judgment, then the pictorial profile (landscape slide in this case) is certainly a valid means of assessing such judgment.

In sum, the use of pictorial profiles may indicate why some of the variance in landscape preferences was left unexplained. Nonetheless, such a hypothesis probably could not by itself account for the substantial amount of unexplained variance in the models derived in this study. This is especially true with respect to the evidence cited earlier for the superiority of the linear model composed of a few major cues.

Cue Intercorrelations

The desire for realism in this study led to the utilization of several landscape dimensions that were highly correlated. As discussed above, the use of slides as profiles necessarily exposes raters to a wide variety of cues which are intercorrelated. The effects of dimension intercorrelation upon the obtained policies of raters have already been discussed in some detail. The question is "To what extent do cue intercorrelations account for unexplained variance in individual and group models?" The answer appears to be "only to a very small extent." In fact, two studies (Dudycha and Naylor 1966; Schenk and Naylor 1968) indicate that the use of an R matrix with high correlations may inflate the amount of variance explained. Dudycha and Naylor's explanation for this finding is that raters are less able to discriminate among intercorrelated cues, but find the decision-making task much easier, leading to higher

R^2 values than those obtained under low correlation conditions. In line with this reasoning, policies obtained from 4 uncorrelated variables possessed slightly smaller R^2 values than those policies based upon the entire R matrix.

Dudycha and Naylor (p. 602) conclude that

> If one is interested in obtaining information about the judgmental policies of individuals toward a certain class of stimulus objects he needs to be certain that the underlying cue R matrix for his experimental stimuli is representative of the R matrix describing the population of stimuli of which his sample is assumed to be a subset. Otherwise, his ability to generalize is obviously going to be limited.

While generalization was not a major goal of this study, for reasons previously discussed, it is strongly felt that the dimension R matrix in this study is representative of the true R matrix underlying all landscape scenes of the BRP.

Hammond et al. (1975) cite evidence showing that individual rater policies will show more diversity when the predictors are intercorrelated than when they are orthogonal. Such evidence may partially explain the low degree of policy similarity exhibited in Tables 2 and 3.

Darlington (1968, p. 179) also warns that

> When the predictor variables in a multiple regression equation are intercorrelated, the 'contribution to variance' of a predictor variable cannot be interpreted in the same way that it can be interpreted when predictor variables are uncorrelated. In the latter case, the phrase has essentially the same meaning it has in analysis of variance designs.

Therefore, the intercorrelations of some of the landscape dimensions do not explain why the models failed to account for more variance than they did. The real problem lies in the interpretation of the results. Methods exist for interpreting results in light of multicollinearity (i.e., ridge, latent roots, and principal components regression), but these methods have been explored very little in the physical sciences, much less in the behavioral sciences. Exploration of these analytical tools should be a goal of future behavioral research.

Rating Scales

The reduction in importance of nonlinear cue utilization, the inappropriateness of dimensions used, utilization of pictorial profiles, and cue intercorrelations in accounting for unexplained variance leaves only one other plausible rival hypothesis: the use of rating scales to measure dimensions and preference. Rating scales are notoriously prone to such errors as the halo effect and errors of central tendency (Guilford 1954). Rating scales are also easily faked (Blum and Naylor 1968). These methodological considerations partially explain the wide variability in individual subject ratings assigned some of the dimensions and overall landscape preference (Appendix B). Raters may have been assigning values based on a general impression of a given dimension rather than attending to the definitions of the dimensions given in the instructions. In addition, the difficulty of defining some dimensions, such as unobstructed expanse of view, visible distant landforms, and undisturbed forest, is evidenced by the relatively high standard deviations associated with these cues for many of the slides (Appendix E). Such rater inconsistency calls for a more objective measure of landscape demensions, perhaps a grid analysis similar to that used by Shafer and his colleagues (1969).

In fact, an exploratory analysis using grid analysis results from a subsample of the original 100 slides and the same preference ratings has resulted in substantial increases in individual rater reliabilities. In the present study 17 subjects had R^2 values greater than 0.50. However, in the analysis involving grid scores, 69 of the 78 subjects had R^2 values over 0.50; out of these 69, 48 had reliabilities greater than 0.70. The multiple correlation coefficient, R^2 , at the 78 separate policy stage was 0.69. This again represents a considerable increase over the results reported in this study. The R^2 value at the one policy stage, however, was only 0.22 with a much larger error term than that of the final grouping reported in Table 5. These results indicate that although grid analysis of dimensions increases the ability to capture individual rater strategies, enough individual variation still exists to make the identification of homogeneous groups difficult.

Ratings assigned overall landscape preference were even more variable than the values assigned separate dimensions (see data in Appendix B). There is some support in the landscape assessment literature for this finding. Brush (1976) distinguishes between preference judgments (subjects are simply asked to express their likes and dislikes) and comparative appraisals (subjects are asked to make their evaluations relative to some other visually presented or imagined environment). Brush contends that by asking people for their individual preferences, a wide variation in responses is likely to occur. However, by forcing raters to adopt a context for making judgments, comparative appraisals display less response variation. The idea makes intuitive sense and is supported by the work of Buhyoff and Leuschner (1978). Using a paired comparison routine to obtain standardized preference scores, these researchers generated a one-variable model explaining 86 percent of the variance in preference (when subjects were forced to cue in on a single landscape dimension). In the Buhyoff and Leuschner study, subjects were forced to make preference judgments by choosing which of a pair of simultaneously presented scenes they preferred. Thus, the comparative standard was part of the procedure and individual variation was obviously very low ($R^2 = 0.86$).

In addition, a recent literature review by Arthur and her colleagues (1977) reveals support for the argument that scenic beauty differs from scenic preference. In three separate studies (Fines 1968; Rabinowitz and Coughlin 1970; Rutherford and Shafer 1969) there is evidence that scenic beauty ratings are higher than preference ratings, have smaller variances, and are more reliable. Nonetheless, there is some contradictory evidence from two studies (Daniel et al. 1973; Zube et al. 1974). In both of these studies there were no differences between scenic beauty and scenic preference ratings. When such differences did occur they seemed to be related to the method and may be an artifact of the instructions (Zube et al. 1974).

To summarize this chapter, numerous plausible rival hypotheses have either been ruled out or discounted. It appears that the method in which landscape preference and dimensions values were obtained is the largest single source of error and, hence, the reason why individual and composite models did not capture more of the variance in preference than they did. Therefore, the need for future research employing a different method of measuring the dimensions (e.g., grid analysis) is underscored. It also seems necessary to conduct research investigating the effects that different scale anchors (e.g., scenic beauty instead of preference) have on the reliability of rater policies.

Chapter 6

SUMMARY AND CONCLUSIONS

To best evaluate the results of this study, it is first necessary to restate the objectives. These objectives were (1) to determine if Policy Capturing is a viable procedure for modeling the human judgment process regarding scenic beauty preferences, and (2) to identify and determine the relative importance of those landscape features which explain variation in such preference.

The first objective appears to have been well-satisfied. Policy capturing is a conceptually and empirically sound method of investigating individual and group differences in policies concerning landscape assessment. Theoretical soundness and ability to account for individual variation are important advantages of the PC methodology over previous landscape preference research based upon multiple linear regression. According to Wiggins (1973, p. 111),

> the main rationale for an individual difference analysis is that judgments obtained on the basis of a group average will often distort the judgmental strategies of many individual subjects.

The admonition for landscape preference research is clear. The extent of individual differences must somehow be assessed, for if people show much variation in their judgments, decisionmakers must know that any action based on an average index may represent the perceptions of only a minority (Craik and Zube 1976). The broad range of weighting strategies employed by raters in this study is viewed as evidence of the potential for significant individual variations in landscape preferences which somehow needs to be accounted for in future studies of this kind.

If more than one policy is found (as was the case in this study), it is not necessary to ask the raters to arbitrate differences in weights (Christal 1968). Instead, it is possible to use the several equations to rank order the slides and then ask subjects to arbitrate differences in the rank positions of the scenes. Once a rank ordering is agreed upon, one could then simply determine the appropriate set of weights for the final equation. This process would be particularly useful to those interested in developing a scenic quality model for a specific geographic region.

The JAN II program does an excellent job of sequential grouping. In fact this sequential grouping process is a major advantage of JAN II over many factor analytic methods (Naylor and Wherry 1964). Nonetheless, it is difficult to tell how many homogeneous groups are formed, and the appropriate background information is necessary in order to tell what went on in the grouping process. The background information collected in this study did not help distinguish one group from another. In future landscape research using PC, preference for different types of recreation environments and familiarity with the environment being evaluated are two variables that are likely to be helpful in deciding why certain groups are clustered. Other, more extensive background information may also prove useful.

Besides being a useful procedure for evaluating individual and group differences in preference, PC can serve as an aid to the landscape decisionmaker as well. Most federal and state resource agencies currently rely on some sort of descriptive inventory system for assigning scenic beauty values to various environments. Typically, personnel of these agencies assign numerical values to certain environmental features thought to be indicative of scenic quality. These numerical values are then combined is some fashion to arrive at an overall scenic quality index. There is some inherent inconsistency in relying on such individual judgments and in fact there is some empirical evidence that the preferences of resource planners differ markedly from the preferences of the user groups they serve (Buhyoff et al. 1978). Buhyoff et al. (1978) also found, however, that when the professional group was provided information concerning their clients' desires, preferences between professionals and clients showed close correspondence.

Thus, PC can be extremely useful to the resource planner or manager in two different ways:

1. Knowledge of variations in weighting strategies among individuals in various user groups (e.g., recreationists, environmentalists, mining interests, etc.) can help the manager or planner decide what effects various landscape manipulations (e.g., locations of transportation and utility corridors, location of visitor service facilities, and selection of surburban and new town development sites) will have on public acceptance.
2. The resource planner or manager can receive feedback on his or her own rating procedure. Clarification of one's own decisionmaking process (policy equation) can help decisionmakers to be more consistent in their landscape evaluations (Dawes 1977; Anderson 1977).

The second objective of this study, identification and weighting of the most important landscape dimensions, appears to have been only minimally satisfied. The consistently low R^2 values indicate that scenes rated by the same judge tended to be rated on different criteria. This rating of scenes using different dimensions is thought to be a result of having subjects use a rating scale to judge each of the 10 dimensions. The advantage of using a grid analysis as a more objective measure of landscape elements is supported. Employment of a grid analysis of the same 10 dimensions resulted in much higher rater reliability than the subjective rating scale method. However, grid meaurements of the independent variables did little to increase the explanatory power of the composite model, thus indicating that considerable inconsistency in overall preference judgments is still a problem.

On an individual rater basis, amount of clear sky, mountains, man-changed area, undisturbed forest, and clouds were the dimensions most often given the highest weights. There was no clear pattern in terms of which dimensions accounted for most of the variance in preference.

When all raters were combined into one group, blue colors, visible distant landforms, mountains, and man-changed area received the largest weights. The four homogeneous groups clustered by JAN II possessed different weighting strategies, and, as would be expected, there was little consistency among the four groups in terms of which dimensions were most influential. The most explanatory composite models were obtained when only the 17 most reliable subjects were used ($R^2 = 0.335$), when judgments made only on the first 33 slides were considered ($R^2 = 0.238$), and when judgments on the first 33 slides made by the 48 subjects with highest pre-test post-test reliabilities were used ($R^2 = 0.260$). These findings reveal, respectively, the influence of rater inconsistency and an order or fatigue effect. In regards to the order or fatigue effect, there is evidence that judgments made on the first 33 to 50 slides were more consistent than those made on either the last 50 or 67 slides. The R^2's obtained are generally in line with at least one other study (Buhyoff and Leuschner 1978) in which subjects were asked to make a global preference response (in the Buhyoff and Leuschner study, however, only one variable was used). The R^2's clearly fall below several other studies in which either different types of independent variables were used (e.g., Zube et al. 1974), these variables were measured in a different manner (e.g., Shafer et al. 1969), or a different dependent variable, usually scenic beauty, was used (e.g., Arthur 1977). It is not recommended that the models from the present study be used for planning purposes along the Blue Ridge Parkway until future research can increase the explanatory power of the models. As it stands now it is difficult to say which variables influence people's preferences. Nonetheless, the failure to identify important dimensions is not considered to be a function of the original dimensions chosen but of the manner in which they were measured.

A number of plausible rival hypotheses for why the models failed to capture more of the variance in preference were proposed. There is sound empirical and intuitive support for the 10 dimensions chosen, but as Brush (1976) points out, preference may not depend so much on the physical elements themselves but on the configurations of these elements. Some support for this notion emanates from research involving eye movements (Gratzer and McDowell 1971; Macworth and Morandi 1967), Shafer's grid analysis results

(Shafer et al. 1969), and the positive influence of intermediate levels of complexity (Berlyne 1960; Wohlwill 1968; Kaplan et al. 1972). In all of these studies it was the relative number and interaction of elements (e.g., edges between two elements) that either drew attention or increased preference.

Nonlinear utilization of cues is not thought to be a major problem in the present study due to the scatter of the data. However, nonlinear use of landscape dimensions is not ruled out. Even though the human judgment literature overwhelmingly supports the use of the linear model, there is some conflicting evidence (e.g., Berlyne 1960 and Wohlwill 1968 have found evidence suggesting a nonlinear relationship between preference and such factors as novelty and complexity). Thus, the potential exists for finding nonlinear models that explain more variance in preference than the linear model and such models should be considered in future landscape assessment research involving PC.

Multicollinearity poses more of a problem in interpretation of the zero-order validity coefficients and regression weights than it does in reducing the explanatory power of the models. In fact, there is some evidence indicating that high correlations in the cue R matrix may actually inflate the R^2 values obtained in the composite models and lead to lower policy similarity among raters.

Nonetheless, in social science data in general and in landscape assessment in particular, multicollinearity is more representative of real-life phenomena than is the use of completely orthogonal predictors. Certain analytical procedures for the interpretation of regression models in light of multicollinearity have recently become available and should be considered in future landscape preference studies in which nonorthogonality of predictors exists. Until such analytical procedures become more widely available, it seems preferable to rely on orthogonal designs if the researcher is interested only in relative weights (Slovic and Lichtenstein 1971).

From a policy capturing standpoint, the use of color slides as profiles is seen as an extremely realistic means of representing the stimulus to be judged and something that has never been previously attempted. That slides contain all the dimensions used by raters in evaluating landscapes, and that some subset of these dimensions previously scaled by researchers are not presented directly to the rater may be two reasons why the regression models failed to capture some of the unexplained variance. Although these reasons are not believed to have exerted substantial influence on the R^2's obtained, critics of landscape preference research argue that slides are not valid representations of actual scenes. Despite considerable evidence that on-site evaluations agree with judgments of landscape representations, there will probably always remain some doubt concerning the use of slides or photographs. For this reason alone, it would be both necessary and informative to develop a landscape model using the PC procedure and then validate the model by having judges evaluate landscapes on-site. Along these same lines, it seems that the validation of the composite models, through the use of different methods, particularly the study of eye movements, would provide valuable insights into molecular strategies of human judgment (Slovic and Lichtenstein 1971). Eye movements could be investigated either in the lab (i.e., using slides) or on-site. Like any other procedure, PC is not an end in itself but needs to be substantiated with appropriate convergent and divergent methods (Campbell and Fiske 1959).

In spite of these methodological issues, the two largest sources of error in the final models were most likely the use of a rating scale to assess landscape preference and the manner in which the scale was anchored. There was wide variability from rater to rater in how each slide was judged, indicating that subjects may have been reacting more to a general impression of each scene rather than to the instructions. Also, there is some evidence from other research suggesting that scales anchored by "preference" are used more inconsistently than scales anchored by "scenic quality" where some comparative standard of scenic beauty is provided or assumed. This issue is far from settled but it would seem useful to have a group of subjects similar to the ones used in this study make their judgments on the same set of slides using a "scenic quality" rather than a "preferences" scale. Another possibility is the use of one of the psychophysical scaling routines (e.g., signal detection theory, method of paired comparisons) found successful in other landscape research (Daniel and Boster 1976; Buhyoff and Leuschner; Buhyoff and Riesenman 1979) to measure the dependent variable, whether it be preference or scenic quality.

To borrow from computer science jargon, this study has been a "debugging run" to determine the methodological problems inherent in using PC as a method of capturing the strategies used in making landscape evaluations. The results and conclusions described herein suggest a series of recommendations for the future use of policy capturing in landscape assessment. These recommendations can be summarized as follows:

1. To avoid possible fatigue biases, reduce the number of judgments each subject must make from 100 to around 60 or 70. This can be done simply by reducing the subset of landscape elements thought to influence preferences for a particular environment. If the underlying dimensions of a landscape are uncertain or unknown, multidimensional scaling of scenes is an appropriate tool to use to determine which dimensions are important.

2. Once the appropriate dimensions have been identified, use grid analysis or some other objective measure to assign values to each.

3. The use of photographs or slides is considered appropriate as long as some effort is made to show that the photographs or slides are representative of the landscapes from which they were taken. It is also necessary to validate models derived from such representations with judgments of actual landscapes before using the models for management or planning purposes.

4. Collect more background information on the subjects than was collected in this study. Evidence indicates that familiarity with landscapes and exposure to certain landscapes in childhood are two background variables that should be assessed.

5. If the investigation of individual differences is a primary aim, use a scale anchored by preference. However, use a more reliable scaling procedure (e.g., psychophysical scaling) than a simple rating of scenes. If consensus is a goal (e.g., for public decisionmaking), use a scale anchored by scenic quality and instruct subjects to make their ratings according to some given or imagined standard of comparison.

In addition to the potential benefits of PC to resource decisionmakers described in this chapter, there are other practical reasons for using this procedure. Once the important dimensions of landscape preference are known, they may be more effectively managed and quantified, giving scenic quality more nearly an equal status with other forest resources. That is, if scenic quality could be assessed in the same quantitative fashion as the more tangible resources, there would be a common basis for the comparison of scenic beauty benefits with the more traditional timber, water, and forage products (Daniel et al. 1977). Finally, there exists the possibility of using PC in a long term research effort aimed at standardizing the effects of various dimensions on people in different geographic regions of the U.S. Daniel (1976) states that a hazard in using people that actually "consume" the resource (e.g., local people) is that over time they may adapt to seeing and appreciating a landscape that to anyone unfamiliar with the area may seem unpleasant. To use a positive example, the Grand Tetons most likely have a different visual effect on area residents than they do on visitors viewing these mountains for the first time. Thus, there is a need to develop a set of standardized, salient dimensions for everyone. The results of such research may be a taxonomy of similar landscapes and the development of scenic and recreational perceived environmental quality indices, analogous to air quality indices. The need for such indices has recently been supported by Craik and Zube (1976). Policy capturing appears to be an appropriate tool for fulfilling these needs.

LITERATURE CITED

Anderson, B. L. 1977. Differences in teachers' judgment policies for varying numbers of verbal and numerical cues. Org. Beh. and Human Perf. 19(1):68- 88.

Arthur, L. M. 1977. Predicting scenic beauty of forest environments: some empirical tests. For. Sci. 23(2):151-160.

Arthur, L. M., T. C. Daniel, and R. S. Boster. 1977. Scenic assessment: an overview. Landscape Plann. 4:109-129.

Bagley, M. D., C. A. Kroll, and K. Clark. 1973. Aesthetics in environmental planning. Prepared for the Office of Research and Development, U.S. Environmental Protection Agency (EPA-600/5-73-009), Washington, D.C.

Beach, L. R. 1967. Multiple regression as a model for human information utilization. Org. Beh. and Human Perf. 2:276-289.

Berlyne, D. E. 1960. Conflict, arousal, and curiosity. McGraw-Hill, New York. 350pp.

Berlyne, D. E. 1963. Complexity and incongruity variables as determinants of exploratory choice and evaluative ratings. Canadian J. Psych. 17:274- 290.

Blum, M. L. and J. C. Naylor. 1968. Industrial psychology: its theoretical and social foundations. Harper and Row, New York. 637pp.

Borman, W. C. and M. D. Dunnette. 1974. Selection of components to comprise a naval personnel status index (NPSI) and a strategy for investigating their relative importance. Prepared under the Navy R and D Program of the Office of Naval Research under Contract N00014-73-C-0210, NR156-020, 24pp.

Bottenberg, R. A. and R. E. Christal. 1961. An iterative technique for clustering criteria which retains optimum predictive efficiency. (WASS-TN-61-30, ASTIA Document AD-261-615). Lackland AFB, Texas: Personnel Lab., Wright Air Dev. Div.

Bottenberg, R. A. and R. C. Christal. 1968. Grouping criteria: a method which retains maximum predictive efficiency. J. Exp. Ed. 36:28-34.

Brady, D. and L. Rappoport. 1973. Policy capturing in the field: the nuclear safegaurds problem. Org. Beh. and Human Perf. 9(2):253-266 .

Brunswik, E. 1940. Thing constancy as measured by correlation coefficients. Psychol. Rev. 47:69-78.

Brunswik, E. 1952. The conceptual framework of psychology. University of Chicago Press, Chicago.

Brunswik, E. 1956. Perception and the representative design of experiments. University of California Press, Berkeley.

Brush, R. O. and E. L. Shafer. 1975. Application of a landscape preference model to land management. Pages 168-182 in E. H. Zube, J. G. Fabos and R. O. Brush, eds. Landscape assessment: values, perceptions, and resources. Dowden, Hutchinson, and Ross, Stroudsburg, Pa.

Brush, R. O. 1976. Perceived quality of scenic and recreational environments: some methodolcgical issues. In K. H. Craik and E. H. Zube, eds. Perceiving environmental quality. Plenum Press, New York. 310pp.

Buhyoff, G. J. (In press). The reliability of data from systematic observation: a methodological note. J. Leisure Res. 11(4).

Buhyoff, G. J. and W. A. Leuschner. 1978. Estimating psychological disutility from damaged forest stands. For. Sci. 24(3):424-431.

Buhyoff, G. J. and J. D. Wellman. 1978. Seasonality bias in landscape preference research. Leisure Sci. 2(2):181-190.

Buhyoff, G. J. and J. D. Wellman. (In press). Environmental preferences: a critical analysis of a critical analysis. J. Leisure Res. 11(3).

Buhyoff, G. J. and M. F. Riesenman. (1979). Manipulation of dimensionality in landscape preference judgments: a quantitative validation. Leisure Sci. 2(3):221-238.

Buhyoff, G. J., J. D. Wellman, H. Harvey, and R. A. Fraser. 1978. Landscape architects' interpretations of people's landscape preferences. J. Environmental Mgt. 6:255-262.

Campbell, D. and D. Fiske. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol. Bull. 56:81-105.

Carls, E. G. 1974. The effects of people and man-induced conditions on preferences for outdoor recreation landscapes. J. Leisure Res. 6:113-124.

Carlson, A. A. 1977. On the possibility of quantifying scenic beauty. Landscape Plann. 4:131-172.

Christal, R. E. 1963. JAN--A technique for analyzing individual and group judgment. (PRL-TDR-63-3, ASTIA Document AD-403-813). Lackland AFB, Texas: 6570th Personnel Res. Lab., Aerospace Medical Div.

Christal, R. E. 1968a. Selecting a harem--and other applications of the policy capturing model. J. Exp. Ed. 36(4):35-41.

Christal, R. E. 1968b. JAN--A technique for analyzing group judgment. J. Exp. Ed. 36(4):24-27.

Clark, R., J. Hendee, and F. Campbell. 1971. Values, behavior, and conflict in modern camping culture. J. Leisure Res. 3:143-159.

Claudy, J. G. 1972. A comparison of five variable weighting procedures. Ed. and Psych. Meas. 32:311-322.

Coughlin, R. E. and K. A. Goldstein. 1970. The extent of agreement among observers on environmental attractiveness. Reg. Sci. Res. Inst. Discuss. Pap. Number 37, Reg. Sci. Res. Inst., Phila., Pa., 56pp.

Craik, K. H. 1970. A system of landscape dimensions: appraisal of its objectivity and illustration of its scientific application. Report to Resources for the Future, Inc. by Univ. Calif. Inst. Personality Assessment and Res., Berkeley.

Craik, K. H. 1971. The assessment of places. Pages 40-62 in P. McReynolds, ed. Advances in psychological assessment. Vol. 2. Science and Behavior Books, Pao Alto, Calif. 395pp.

Craik, K. H. 1972a. Appraising the objectivity of landscape dimensions. Pages 292-307 in J. V. Krutilla, ed. Natural environments: studies in theoretical and applied analysis. Johns Hopkins Univ. Press, Baltimore.

Craik, K. H. 1972b. Psychological factors in landscape appraisal. Environ. and Behav. 4:255-266.

Craik, K. H. and E. H. Zube. 1976. The development and application of perceived environmental quality indices. In K. H. Craik and E. H. Zube, eds. Perceiving environmental quality. Plenum Press, New York.

Daniel, T. C., R. S. Boster, and P. B. Best, Jr. 1973. Quantitative evaluation of landscapes: an application of signal detection analysis to forest management alternatives. Man-Environ. Sys. 3(5):330-344.

Daniel, T. C. 1976. Criteria for development and application of perceived environmental quality indices. In K. H. Craik and E. H. Zube, eds. Perceiving environmental quality. Plenum Press, New York.

Daniel, T. C. and R. S. Boster. 1976. Measuring landscape esthetics: the scenic beauty estimation method. USDA For. Serv. Res. Pap. RM-167, Rocky Mountain For. and Range Exp. Stn., Fort Collins, Colo. 66pp.

Daniel, T. C., L. M. Anderson, H. W. Schroeder, and L. Wheeler III. 1977. Mapping the scenic beauty of forest landscapes. Leisure Sci. 1(1): 35-52.

Darlington, R. B. 1968. Multiple regression in psychological research and practice. Psych. Bull. 69(3):161-182.

Dawes, R. M. and B. Corrigan. 1974. Linear models in decision making. Psych. Bull. 81(2):95-106.

Dawes, R. M. 1977. Predictive models as a guide to preference. IEEE Transactions on Systems, Man, and Cybernetics SMC-7(5):355-357.

Dickie, G. 1971. Aesthetics: an introduction. Bobbs-Merrill Co., 200pp.

Draper, N. R. and H. Smith. 1966. Applied regression analysis. John Wiley and Sons, New York.

Dudycha, A. L. and J. C. Naylor. 1966a. The effect of variations in the cue R matrix upon the obtained policy equations of judges. Ed. and Psych. Meas. 26:583-603.

Dudycha, A. L. and J. C. Naylor. 1966b. Characteristics of the human inference process in complex choice behavior situations. Org. Beh. and Human Perf. 1:110-128.

Fabos, J. G. 1971. An analysis of environmental quality ranking systems. Pages 40-55 in W. T. Doolittle and R. E. Getty, eds. Recreation symposium proceedings. USDA For. Serv. Northeast For. Exp. Stn., Upper Darby, Pa., 211pp.

Fairfax, S. K. and G. L. Achterman. 1977. The Monongahela controversy and the political process. J. Forestry 75(8):485-487.

Festinger, L. 1957. A theory of cognitive dissonance. Harper and Row, New York.

Fines, K. D. 1968. Landscape evaluations: a research project in East Sussex. Regional Studies 2:41-55.

Goldberg, L. R. 1965. Diagnosticians vs diagnostic signs: the diagnosis of psychosis vs neurosis from the MMPI. Psychol. Monographs 79(9): Whole No. 602.

Goldberg, L. R. 1968. Simple models or simple processes? Some research on clinical judgments. Am. Psychologist 23:483-496.

Goldberg, L. R. 1970. Man versus model of man: a rationale, plus some evidence, for a method of improving clinical inference. Psych. Bull. 73(6):422-432.

Goldberg, L. R. 1976. Man versus model of man: just how conflicting is that evidence? Org. Beh. and Human Perf. 16(1):13-22.

Gratzer, M. A. and R. D. McDowell. 1971. Adaptation of an eye movement recorder to environmental mensuration. Storrs Agric. Exp. Stn. Res. Rep. No. 36, Storrs, Conn.

Green, D. M. and J. A. Swets. 1966. Signal detection theory and psychophysics. John Wiley and Sons, New York. 455pp.

Guilford, J. P. 1954. Psychometric methods. McGraw-Hill, New York. 597pp.

Hamilton, J. W. and T. L. Dickinson. 1978. Validity of policy capturing as a procedure for synthetic validation. Paper presented at the Meeting of the Rocky Mtn. Psychological Association, Denver, Colo.

Hammond, K. R. 1955. Probabilistic functioning and the clinical method. Psych. Rev. 62:255-262.

Hammond, K. R., T. R. Stewart, B. Brehmer, and D. O. Steinman. 1975. Social judgment theory. Pages 271-312 in M. F. Kaplan and S. Schwartz, eds. Human judgment and decision processes. Academic Press, Inc., New York.

Hammond, K. R., J. L. Mumpower, and T. H. Smith. 1977. Linking environmental models with models of human judgment: a symmetrical decision aid. IEEE Transactions on Man, Systems, and Cybernetics SMC-7(5):358-367.

Harvey, H. 1977. Landscape preference quantification: a comparison of methods. M.S. thesis, VPI & SU, Blacksburg, Va. 110pp.

Harvey, J. H. and W. P. Smith. 1977. Social psychology: an attributional approach. C. V. Mosby, St. Louis. 426pp.

Hoffman, P. J. 1960. The paramorphic representation of clinical judgment. Psych. Bull. 47:116-131.

Jackson, R. H., L. E. Hudman, and J. L. England. 1978. Assessment of the environmental impact of high voltage power transmission lines. J. Environmental Mgt. 6:153-170.

Kaplan, R. 1975. Some methods and strategies in the prediction of preferences. In E. H. Zube, J. G. Fabos, and R. O. Brush, eds. Landscape assessment: values, perceptions, and resources. Dowden, Hutchinson, and Ross, Stroudsburg, Pa.

Kaplan, S., R. Kaplan, and J. S. Wendt. 1972. Rated preference and complexity for natural and urban visual material. Percept. and Psychophys. 12(4):354-356.

Kelley, H. H. 1967? Attribution theory in social psychology. In D. Levine, ed. Nebraska symposium on motivation. University of Nebraska Press, Lincoln.

Kreimer, A. 1977. Environmental preferences: a critical analysis of some research methodologies. J. Leisure Res. 9(2):88-97.

Langer, E. J. 1975. The illusion of control. J. Personality and Soc. Psych. 32:311-328.

Lemke, E. and W. Wiersma. 1976. Principles of psychological measurement. Rand-McNally, Chicago. 300pp.

Libby, R. 1976a. Man versus model of man: some conflicting evidence. Org. Beh. and Human Perf. 16(1):1-12.

Libby, R. 1976b. Man versus model of man: the need for a nonlinear model. Org. Beh. and Human Perf. 16(1):23-26.

Lucas, R. C. 1970. User evaluation of campgrounds in two Michigan national forests. USDA For. Serv. Res. Pap. NC-44, North Central For. Exp. Stn., St. Paul, MN.

Mackworth, N. H. and A. J. Morandi. 1967. The gaze selects informative details within pictures. Percept. and Psychophys. 2(11):547-552.

Madden, J. M. 1963. An application to job evaluation of a policy capturing model for analyzing individual and group judgment. (PRL-TDR-63-15). Lackland AFB, Texas: Personnel Res. Lab., Aerospace Medical Div.

McHarg, I. L. 1969. Design with nature. Doubleday and Company, Garden City, New York. 197pp.

Meehl, P. E. 1954. Clinical versus statistical prediction: a theoretical analysis and review of the literature. University of Minnesota Press, Minneapolis.

Meehl, P. E. 1965. Clinical versus statistical prediction. J. Exp. Res. in Personality. 1:27-32.

Mure, G. R. G. 1964. Aristotle. Oxford University Press, New York. 280pp.

Naylor, J. C. and R. J. Wherry, Sr. 1964. Feasibility of distinguishing supervisors' policies in evaluation of subordinates by using ratings of simulated job incumbents. (PRL-TR-64-25). Lackland AFB, Texas: Personnel Res. Lab., Aerospace Medical Div.

Naylor, J. C. and R. J. Wherry, Jr. 1965. The use of simulated stimuli and the "JAN" technique to capture and cluster the policies of raters. Ed. and Psych. Meas. 25(4):969-986.

Newby, F. 1971. Understanding the visual resource. Pages 68-72 in W. T. Doolittle and R. E. Getty, eds. Recreation symposium proceedings. USDA For. Serv. Northeast For. Exp. Stn., Upper Darby, Pa. 211pp.

Nunnally, J. C., Jr. 1970. Introduction to psychological measurement. McGraw-Hill, New York. 572pp.

Peterson, G. L. 1967. A model of preference: quantitative analysis of the perception of the visual appearance of residential neighborhoods. J. Reg. Sci. 7(1):19-31.

Peterson, G. L. and E. S. Neumann. 1969. Modeling and predicting human response to the visual recreation environment. J. Leisure Res. 1:219-237.

Peterson, G. L. 1974. A comparison of the sentiments and perceptions of wilderness managers and canoeists in the Boundary Waters Canoe Area. J. Leisure Res. 6:194-206.

Pickford, R. W. 1972. Psychology and visual aesthetics. Hutchinson Educational, London. 270pp.

Rabinowitz, C. B. and R. E. Coughlin. 1971. Some experiments in quantitative measurement of landscape quality. Reg. Sci. Res. Inst. Discuss. Pap. 43, Reg. Sci. Res. Inst., Phila., Pa.

Riesenman, M. F. 1977. An analysis of the southern pine beetle's impact on aesthetic values of forested landscapes. M.S. thesis, VPI & SU, Blacksburg, Va. 142pp.

Rock, I. and C. S. Harris. 1967. Vision and touch. Sci. Am. 216(5): 96-104.

Roose, J. E. and M. E. Doherty. 1976. Judgment theory applied to the selection of life insurance salesmen. Org. Beh. and Human Perf. 16: 231-249.

Rutherford, W., Jr. and E. L. Shafer, Jr. 1969. Selection cuts increased natural beauty in two Adirondack forest stands. J. Forestry 67(6): 415-419.

Santayana, G. 1907 (c1896). The sense of beauty; being the outline of aesthetic theory. Scribner's Sons, New York.

Sax, J. L. 1971. Defending the environment: a strategy for citizen action. Alfred A. Knopf, New York. 252pp.

Schenk, E. A. and J. C. Naylor. 1968. A cautionary note concerning the use of regression analysis for capturing the strategies of people. Ed. and Psych. Meas. 28:3-7.

Shafer, E. L., Jr., J. F. Hamilton, and E. A. Schmidt. 1969. Natural landscape preferences: a predictive model. J. Leisure Res. 1(1): 1-19.

Shafer, E. L., Jr. and J. Mietz. 1970. It seems possible to quantify scenic beauty in photographs. USDA For. Serv. Res. Pap. NE-162, Northeast For. Exp. Stn., Upper Darby, Pa. 12pp.

Shafer, E. L., Jr. and M. Tooby. 1973. Landscape preferences: an international replication. J. Leisure Res. 5:60-65.

Shafer, E. L., Jr. and T. A. Richards. 1974. A comparison of viewer reactions to outdoor scenes and photographs of those scenes. USDA For. Serv. Res. Pap. NE-302, Northeast For. Exp. Stn., Amherst, Mass. 26pp.

Shafer, E. L., Jr. and R. O. Brush. 1977. How to measure preferences for photographs of natural landscapes. Landscape Plann. 4:237-256.

Slovic, P. 1969. Analyzing the expert judge: a descriptive study of a stockbroker's decision processes. J. Applied Psych. 53(4):255-263.

Slovic, P. and S. Lichtenstein. 1971. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Org. Beh. and Human Perf. 6:649-744.

Sonnenfeld, J. 1966. Variable values in space and landscape: an inquiry into the nature of environmental necessity. J. Soc. Issues 22(4):71-82.

Taylor, R. L. and W. D. Wilsted. 1974. Capturing judgment policies: a field study of performance appraisal. Academy of Mgt. J. 17(3): 440-449.

Thurstone, L. L. 1927. A law of comparative judgment. Psychol. Rev . 34:278-286.

Thurstone, L. L. 1959. Theory and method of scaling. John Wiley and Sons, New York.

U.S. Congress. 1960. The Multiple Use Sustained Yield Act of June 12, 1960. 74 Stat. 215; 16 U.S.C. 528-531, Public Law 86-517.

U.S. Congress. 1969. The National Environmental Policy Act of 1969. 83 Stat. 852; 42 U.S.C. 4321,4331-4335, 4341-4347, Public Law 91-190.

U.S. Congress. 1974. Forest and Rangeland Renewable Resources Planning Act of 1974. 88 Stat. 476; 16 U.S.C. 1601-1610, Public Law 93-378.

U.S. Congress. 1976. Federal Land Policy and Management Act of 1976. 90 Stat. 2743; 43 U.S.C. 1701-1782, Public Law 94-579.

U.S. Department of Agriculture, Forest Service. 1973. National forest landscape management. Vol. 1. U.S. Dep. Agric., Agric. Handbook 434, 77pp., U.S. Government Printing Office, Washington D.C.

U.S. Department of Agriculture, Forest Service. 1974. National forest landscape management. Vol. 2, ch. 1 (The visual management system). U.S. Dep. Agric., Agric. Handbook 462, 47pp., U.S. Government Printing Office, Washington, D.C.

Wang, M. D. and J. C. Stanley. 1970. Differential weighting: a review of methods in empirical studies. Rev. of Ed. Res. 40:663-705.

Ward, J. H. 1961. Hierarchical grouping to maximize payoff. (WADD-TN61-29, ASTIA Document AD-261-750). Lackland AFB, Texas: Personnel Lab., Wright Air Dev. Div.

Webb, E. J., D. T. Campbell, R. D. Schwartz, and L. Sechrest. 1966. Unobtrusive measures: nonreactive research in the social sciences. Rand McNally, Chicago. 225pp.

Weinstein, N. D. 1976. The statistical prediction of environmental preferences. Environ. and Behav. 8(4):611-626.

Wenger, W. D., Jr. and R. Videbeck. 1969. Eye pupillary measurement of aesthetic responses to forest scenes. J. Leisure Res. 1(2):149-161.

Wherry, R. J. and J. C. Naylor. 1966. Comparison of two approaches-- JAN and PROF--for capturing rater strategies. Ed. and Psych. Meas. 26(2):267-286.

Wiggins, N. and E. S. Kohen. 1971. Man vs. model of man revisited: the forecasting of graduate school success. J. Personality and Soc. Psych. 19:100-106.

Wiggins, N. 1973. Individual differences in human judgment: a multivariate approach. Pages 110-142 in L. Rappoport and D. A. Summers, eds. Human judgment and social interaction. Holt, Rinehart, and Winston, New York. 403pp.

Wohlwill, J. F. 1968. Amount of stimulus exploration and preference as differential functions of stimulus complexity. Percept. and Psychophys. 4(5):307-312.

Zedeck, S. and D. Kafry. 1977. Capturing rater policies for processing evaluation data. Org. Beh. and Human Perf. 18: 269-294.

Zube, E. H. 1974. Cross-disciplinary and intermode agreement on the description and evaluation of landscape resources. Environ. and Behav. 6(1):69-89.

Zube, E. H., D. G. Pitt, and T. W. Anderson. 1974. Perception and measurement of scenic resources in the Southern Connecticut River Valley. Inst. Man and His Environ. Publ. No. R-74-1, Inst. Man and His Environ., Amherst, Mass.

APPENDIX A

Instructions for subjects rating the 10 landscape dimensions ("Mountains" used as an example):

"A landscape is a complex visual stimulus composed of many features, or dimensions. Such dimensions include color, topography, amount of water, vegetation, and so on. Your task will be to rate 100 slides of landscape scenes along *only one* dimension: *amount of mountains*. *Amount of mountains* is defined as the proportion of the entire scene containing mountains, as opposed to flat land, rolling hills, and sky."

"The rating scale you will be using ranges from 1 to 9 as follows:

1	2	3	4	5	6	7	8	9
Low								High

A '1' on this scale represents complete absence of the dimension you are rating (that is, no mountains). A '9' means that the scene is almost entirely covered with mountains. A '5' represents an intermediate amount of mountains."

"For each slide shown simply blacken in a number from 1 to 9 on your op-scan sheet. The number you blacken in represents how much mountainous area you think each scene possesses. Remember - rate the slide only on the amount of mountains. Do not rate the scene on any other characteristics it may possess."

APPENDIX B

Table 1. Dimension and preference means and standard deviations associated with each slide.

Landscape Dimensions -

Slide	Sky		Foreground Vegetation		Green Colors		Blue Colors		Mountains		Man-changed area		Unobstructed expanse of view		Visible distant landforms		Undisturbed forest		Clouds		Overall landscape preference	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
1*	3.04	0.37	2.59	0.96	6.39	0.89	2.12	0.80	6.14	0.94	1.91	0.68	7.30	1.72	5.39	1.53	6.73	1.08	3.46	3.20	3.36	1.84
2*	1.46	0.51	1.73	0.70	7.35	2.04	2.17	1.68	8.54	0.51	1.09	0.29	8.44	0.94	8.30	0.56	8.18	0.66	6.50	3.08	3.64	2.13
3	2.09	0.53	4.04	1.17	2.87	0.63	4.38	1.31	3.09	1.34	2.68	0.78	3.83	1.61	3.74	0.75	4.14	1.39	6.96	1.29	5.46	1.90
4	1.68	1.09	5.23	1.54	4.48	0.90	2.42	1.14	3.73	0.88	1.23	0.61	5.96	1.02	3.44	0.66	5.82	1.40	8.18	1.84	5.26	1.82
5	1.14	0.35	2.54	1.57	2.35	1.11	3.71	1.33	4.91	1.34	2.82	1.18	3.96	1.66	5.61	1.44	4.64	1.53	7.68	1.81	5.18	2.42
6	1.73	1.16	2.46	1.26	4.70	0.70	2.42	1.35	4.96	1.00	1.54	0.51	7.39	1.08	4.35	1.03	5.27	1.58	7.96	1.86	3.92	1.79
7	2.91	0.75	5.00	1.63	6.74	0.54	3.08	0.58	4.86	1.08	1.18	0.59	5.30	1.96	4.09	1.08	6.77	1.19	4.46	1.62	6.50	1.67
8	1.14	0.47	1.73	1.52	5.87	1.58	2.08	1.21	6.68	0.78	2.59	1.22	6.52	1.68	7.04	1.85	7.27	1.16	7.86	1.94	4.90	2.15
9*	2.32	0.78	1.59	0.85	2.52	1.31	3.00	0.78	1.73	1.28	5.00	1.66	3.96	2.44	5.26	1.84	2.50	0.91	1.09	0.29	5.15	2.41
10	3.59	1.01	4.00	1.57	5.44	0.94	3.33	0.76	3.68	1.43	2.50	1.06	3.61	1.47	4.52	1.50	5.82	1.33	2.68	2.61	5.06	1.65
11	3.96	0.84	2.73	0.63	5.17	1.03	3.42	0.97	3.27	1.32	3.41	0.91	3.17	1.50	5.70	1.02	4.46	1.18	3.09	2.37	5.53	1.69
12*	6.18	0.80	1.86	0.99	3.39	0.89	6.21	0.83	3.41	0.59	2.09	0.53	6.39	1.80	3.00	1.04	4.09	1.72	1.68	1.76	5.87	1.69
13*	1.91	1.41	2.27	0.70	5.09	2.50	2.38	0.71	7.68	0.89	1.32	0.57	7.13	1.63	7.22	1.20	7.91	0.53	7.54	2.13	4.70	2.20
14*	3.82	0.80	4.73	1.72	4.30	1.33	4.42	0.93	4.23	1.31	1.86	0.47	6.70	0.97	3.52	0.79	5.82	1.01	2.91	1.82	5.09	1.67
15	2.00	1.54	4.04	1.65	3.96	1.22	3.58	1.82	4.09	0.68	1.36	0.58	6.91	1.08	3.48	0.66	5.54	1.47	8.32	1.64	5.00	1.84
16*	2.96	0.72	4.59	1.44	6.70	0.56	2.62	0.71	5.82	1.37	1.04	0.21	8.52	0.79	5.00	1.51	7.18	0.91	2.73	1.16	4.41	1.88
17	2.41	0.67	4.41	1.45	6.17	1.19	3.21	1.22	5.50	1.37	1.73	0.55	6.04	1.36	5.04	0.88	7.14	0.71	4.91	1.82	5.87	1.66
18*	2.73	0.70	2.91	1.38	6.83	0.89	2.62	0.58	5.00	1.34	3.64	0.66	6.61	1.78	6.52	1.31	5.04	1.09	1.32	0.72	4.96	1.87
19*	6.68	1.04	1.32	0.94	3.30	0.64	6.75	0.53	3.68	0.65	1.09	0.29	8.30	1.49	3.96	1.40	4.27	1.70	1.04	0.21	4.94	1.99
20*	4.82	1.01	1.59	0.59	3.61	0.72	4.92	0.83	4.18	0.66	3.18	0.80	7.52	1.41	5.13	1.22	4.09	1.31	1.46	1.71	5.31	1.98
21	4.32	0.78	2.27	1.08	5.22	0.67	3.25	1.03	4.14	1.28	1.23	0.43	8.17	1.19	5.44	0.94	6.36	1.56	2.77	2.35	3.00	2.08
22	1.59	0.73	1.73	1.45	5.44	1.31	2.75	1.11	5.46	1.14	2.50	0.60	4.13	1.74	7.09	1.20	6.86	1.12	7.64	1.14	6.23	2.06
23	1.91	0.29	2.32	1.13	5.83	1.03	3.12	0.80	6.18	1.01	2.09	0.29	5.61	1.41	6.87	1.10	7.00	0.76	7.00	1.16	6.27	2.06
24*	1.27	0.77	2.14	0.99	4.30	0.82	1.67	1.05	4.73	0.94	1.27	0.46	8.04	0.98	5.44	1.08	5.82	1.62	8.54	1.71	3.53	1.97
25*	4.64	0.90	1.91	0.61	4.78	0.74	4.67	1.01	4.77	0.53	1.14	0.47	8.13	1.22	4.96	0.88	5.77	1.57	1.77	1.77	3.49	1.81
26	2.23	0.75	3.82	1.14	2.74	1.18	3.04	0.86	6.14	1.81	2.86	1.21	4.30	1.30	5.87	1.42	5.36	2.01	5.64	2.38	4.90	2.62
27	1.27	0.46	4.09	1.11	6.78	1.56	1.58	0.72	6.64	1.14	2.09	0.68	5.52	1.47	6.44	1.12	7.54	0.80	8.18	1.84	5.30	1.71
28*	4.96	1.17	2.14	0.71	3.22	0.60	6.21	0.66	3.82	0.96	2.77	0.61	5.22	1.48	3.74	1.05	4.23	1.57	2.41	1.92	6.13	1.76
29*	1.32	0.89	3.36	1.53	3.48	1.16	2.92	1.14	2.32	0.48	4.04	1.00	5.65	1.85	3.78	1.00	3.18	1.05	8.23	0.72	8.04	1.20

(continued)

Table I. Dimension and preference means and standard deviations associated with each slide (continued).

	Landscape Dimensions -																			
	Sky		Foreground Vegetation		Green Colors		Blue Colors		Mountains		Man-changed area		Unobstructed expanse of view		Visible distant landforms		Undisturbed forest		Clouds	: Overall landscape preference	
Slide	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
30	1.32	0.65	5.73	1.42	4.56	0.84	2.67	1.24	2.73	1.12	1.59	0.59	6.30	1.15	3.30	1.15	6.00	1.20	8.91	0.29	5.86	2.22
31	1.27	0.55	5.64	1.65	4.52	0.85	2.62	1.31	2.64	1.14	1.59	0.50	6.39	1.12	3.04	1.02	6.04	1.17	9.00	0.00	4.97	2.06
32	3.00	0.82	2.86	1.17	5.78	1.08	1.88	0.90	4.96	1.53	3.68	1.56	7.52	1.12	4.61	1.64	4.46	1.47	6.50	3.56	6.10	1.81
33	2.41	0.67	3.59	1.30	5.30	1.43	2.62	0.58	1.18	0.66	6.09	2.09	3.74	2.54	5.35	1.53	1.64	0.90	3.41	2.58	6.88	1.74
34*	2.09	0.53	3.91	1.38	6.39	0.89	2.75	0.74	4.82	1.44	1.23	0.43	5.00	2.04	5.17	1.34	7.14	0.71	5.73	1.64	5.59	1.80
35*	2.23	1.02	1.46	0.96	6.70	0.88	1.71	0.91	6.96	0.38	1.04	0.21	7.48	1.88	6.70	1.33	7.64	0.90	6.18	3.51	4.18	2.36
36	1.91	0.61	4.41	2.40	6.70	0.93	1.96	0.36	1.27	1.28	3.41	1.26	7.04	2.50	3.70	2.42	6.14	1.25	1.77	1.66	5.78	1.90
37	4.73	1.03	2.36	1.97	4.83	0.58	4.79	0.72	3.64	1.29	1.86	1.04	7.96	1.15	4.00	1.45	5.64	1.56	2.77	2.67	3.50	1.65
38*	3.64	0.66	4.09	1.27	6.17	0.89	3.38	0.71	4.36	0.90	2.59	0.73	7.78	1.04	4.39	0.99	5.04	1.59	2.09	2.27	3.63	1.80
39*	3.68	0.48	3.50	0.80	6.04	0.71	4.17	0.56	4.73	0.77	1.27	0.55	7.48	1.12	4.65	1.03	6.36	0.95	2.41	0.59	4.44	1.76
40*	5.96	0.78	2.23	0.61	4.30	0.82	5.42	0.65	3.96	0.78	1.18	0.40	6.74	1.25	3.65	0.94	5.27	1.78	1.23	0.53	6.01	1.66
41	2.23	0.61	3.18	2.24	5.56	1.16	2.42	0.50	1.77	1.34	3.91	1.15	7.87	1.42	4.26	2.53	4.14	1.81	2.64	1.71	6.72	1.88
42	1.09	0.29	3.27	1.24	4.83	1.19	1.92	1.02	3.77	1.44	2.91	0.75	4.78	2.04	5.52	1.41	5.23	1.27	8.64	1.50	5.55	1.88
43	1.09	0.29	5.68	1.36	5.61	1.37	2.17	1.40	4.00	1.41	2.09	0.87	6.17	1.37	3.70	0.88	6.54	1.14	8.64	0.73	4.69	1.78
44	2.41	0.80	3.82	1.30	3.44	1.67	4.04	1.00	2.96	1.17	2.68	0.89	4.78	1.59	4.44	1.31	5.00	1.75	5.18	1.56	6.19	1.87
45	1.04	0.21	2.18	1.53	7.17	1.30	1.62	1.01	7.32	1.89	2.00	0.76	7.04	0.98	7.39	1.99	8.00	0.54	7.82	2.24	3.50	1.86
46	2.50	0.86	5.27	1.75	5.04	1.19	3.17	1.01	4.64	1.65	1.14	0.64	7.44	0.90	3.22	1.13	6.50	1.22	5.54	2.04	4.30	1.69
47	1.14	0.47	4.96	1.50	4.04	1.19	2.38	1.64	3.77	1.34	1.82	0.66	6.83	1.19	3.61	1.20	6.64	1.29	8.54	1.40	4.50	1.98
48	1.27	0.46	2.96	1.21	4.56	1.38	2.00	0.83	1.86	1.28	5.64	1.73	6.74	2.05	3.83	2.21	2.09	0.92	7.73	1.72	4.78	2.13
49*	1.18	0.40	2.73	0.98	6.09	0.85	2.08	0.97	5.96	1.25	2.18	0.73	6.44	1.08	6.26	1.01	6.86	1.04	8.91	0.29	5.24	1.91
50*	1.18	0.40	3.54	1.01	5.96	1.02	1.92	0.88	5.50	0.96	1.77	0.43	7.56	0.79	5.48	1.16	6.91	0.87	8.96	0.21	4.08	1.84
51*	1.27	0.77	3.82	0.91	4.87	1.32	1.88	0.99	5.23	0.87	1.68	0.57	7.56	0.73	4.91	1.12	6.50	1.22	8.91	0.29	4.36	1.74
52	1.18	0.50	3.50	1.22	4.83	1.11	2.92	1.06	3.54	1.10	2.45	0.51	5.65	1.82	4.48	1.44	5.91	1.15	8.82	0.40	5.30	1.98
53*	1.09	0.29	4.59	1.33	6.56	1.44	1.71	0.75	6.36	1.40	1.96	0.65	6.70	1.30	5.91	1.08	7.82	0.59	8.86	0.35	4.68	1.98
54*	1.96	1.65	3.09	0.68	4.65	1.23	1.88	1.33	4.59	0.50	1.36	0.49	8.09	0.52	4.04	0.98	5.82	1.53	8.77	0.53	4.23	1.92
55*	1.46	0.74	2.54	0.67	6.04	1.15	2.00	1.02	6.18	1.10	2.18	0.73	7.61	0.99	6.17	0.98	6.91	1.11	8.73	0.55	4.68	2.07
56*	1.32	0.57	4.00	1.31	5.65	1.26	2.17	1.31	5.46	1.22	2.41	0.59	7.04	0.64	5.30	1.10	6.64	0.85	8.59	0.96	5.22	1.84
57*	1.00	0.00	2.18	0.73	7.04	3.01	1.17	0.48	7.77	2.78	1.04	0.21	8.56	1.16	8.22	0.67	8.91	0.29	4.18	3.20	3.87	2.18
58*	6.64	0.66	1.64	0.66	3.22	0.67	5.58	1.44	3.54	0.60	1.00	0.00	7.91	1.62	3.44	1.20	4.32	1.94	1.68	1.46	4.44	1.83

(continued)

Table 1. Dimension and preference means and standard deviations associated with each slide (continued).

	Landscape Dimensions -																		: Overall landscape preference	
	Sky		Foreground Vegetation		Green Colors		Blue Colors		Mountains		Man-changed area		Unobstructed expanse of view		Visible distant landforms		Undisturbed forest		Clouds	
Slide	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
59	3.96	0.58	3.27	1.70	5.74	1.21	3.75	0.90	3.18	1.05	2.54	1.01	6.87	1.46	4.26	1.39	4.64	1.56	2.18	2.11	4.55	2.20
60*	1.64	0.66	1.27	0.63	6.52	1.97	1.67	0.92	7.96	0.38	1.18	0.40	7.83	1.72	7.70	0.64	7.91	0.75	7.96	1.96	3.95	2.04
61	4.64	1.18	2.46	1.30	2.70	1.58	4.83	1.27	1.46	0.51	5.23	0.97	3.30	2.36	3.83	1.19	1.82	1.68	3.68	1.46	7.10	1.98
62	1.04	0.21	1.86	1.64	6.91	1.38	1.54	0.88	7.88	0.66	2.82	0.85	6.78	1.68	8.00	1.62	7.46	1.06	8.04	2.38	4.50	2.34
63*	1.36	0.66	5.91	1.82	6.00	1.38	2.58	0.72	2.77	1.02	1.73	0.70	6.87	1.14	2.74	1.14	7.14	0.89	8.73	0.55	4.83	2.16
64*	1.14	0.35	3.86	1.25	6.87	1.25	1.54	0.78	6.14	1.28	2.23	0.61	6.65	0.98	6.30	1.02	7.64	0.66	8.50	0.96	4.65	1.97
65*	2.14	0.56	2.68	0.84	5.52	2.39	2.62	0.65	7.00	0.82	1.14	0.35	8.13	0.55	6.65	0.65	7.54	0.91	5.23	0.92	4.90	2.09
66*	2.36	0.73	1.91	0.53	4.52	1.20	3.21	0.59	5.82	0.73	1.36	0.49	6.48	1.62	5.70	1.10	6.73	1.55	6.32	1.17	5.81	2.08
67*	6.86	0.71	2.09	0.68	3.35	0.65	6.12	0.90	3.18	0.59	1.14	0.35	7.17	1.72	3.52	1.38	4.18	1.99	1.27	0.55	5.69	1.96
68*	1.59	0.50	1.23	0.53	4.74	1.25	2.38	1.17	4.91	1.27	2.73	0.70	4.87	2.07	6.30	0.93	6.09	1.34	8.23	0.53	4.96	2.07
69*	1.50	0.86	3.18	1.53	3.87	0.92	2.04	1.40	3.50	0.96	2.68	0.72	6.09	1.12	4.65	0.83	4.73	1.39	8.86	0.47	5.10	1.65
70*	5.86	0.64	2.50	1.01	3.48	0.79	5.42	0.83	3.54	0.60	2.32	0.65	7.35	1.07	3.96	0.82	4.00	1.38	1.36	0.73	4.94	1.76
71*	4.00	0.93	3.50	1.14	2.30	1.02	4.79	0.93	1.77	0.43	5.86	1.17	2.30	1.18	4.22	1.08	1.86	0.83	2.68	1.00	4.53	2.62
72*	3.46	0.60	3.46	1.34	4.74	1.10	3.67	0.82	1.04	0.21	5.91	1.15	5.17	2.23	4.65	1.67	1.59	0.59	1.23	0.68	5.28	2.31
73	1.91	0.43	3.64	1.09	4.09	1.00	3.50	1.14	2.14	0.47	3.73	0.77	4.04	1.58	4.39	1.08	3.72	1.52	6.73	1.45	6.51	1.84
74	1.32	0.48	3.54	1.30	4.61	1.16	3.04	0.96	2.14	0.71	4.96	1.46	3.17	1.30	5.13	1.36	2.91	0.97	8.32	1.04	6.95	1.72
75	3.96	0.38	2.36	1.14	3.70	1.02	5.17	0.96	2.96	1.05	3.68	1.43	5.78	1.24	4.39	1.08	2.73	1.16	3.64	1.43	5.96	2.07
76	2.18	0.66	1.73	0.83	2.26	1.94	2.58	1.06	1.91	0.92	5.64	1.79	5.13	1.94	6.09	1.08	2.23	1.69	5.18	2.32	6.26	1.94
77*	4.18	0.50	2.50	0.96	4.74	1.39	4.42	0.93	1.73	0.55	5.41	1.05	2.61	1.67	5.52	1.16	2.64	1.29	1.32	0.78	6.87	1.85
78	1.91	0.43	3.82	1.10	6.13	1.25	2.46	0.72	5.36	1.22	1.91	0.43	6.56	1.56	4.70	1.18	6.96	0.90	6.46	2.24	4.82	1.79
79*	3.04	0.58	1.91	0.68	4.56	1.41	3.42	0.65	4.00	0.82	3.09	0.68	7.61	1.41	4.96	1.06	4.68	1.36	5.09	1.23	5.15	1.61
80*	1.04	0.21	2.32	0.72	4.65	1.94	1.96	0.96	6.23	1.48	2.27	0.55	6.48	1.16	6.30	1.74	7.18	0.85	8.59	1.22	5.24	1.96
81	1.23	0.43	5.04	1.76	3.96	1.43	2.25	0.68	5.41	1.50	1.96	0.49	6.09	1.34	5.26	1.74	7.64	0.79	7.54	2.54	5.35	2.02
82*	3.27	0.88	4.36	1.26	3.35	0.78	2.42	1.10	1.41	0.50	3.09	1.31	6.35	1.72	4.17	1.03	4.04	1.46	5.09	3.41	6.26	1.81
83	6.14	0.64	2.23	1.41	3.04	0.56	5.83	1.13	3.27	0.63	2.46	0.67	6.56	1.38	3.39	0.89	4.23	1.63	1.59	1.05	4.72	1.83
84	2.96	1.05	2.91	1.41	3.26	1.14	2.00	1.18	1.46	0.67	5.36	1.40	5.04	1.82	3.87	1.60	2.04	1.33	4.68	3.87	6.56	2.03
85*	1.86	0.35	2.14	0.71	5.26	1.82	2.42	0.50	6.04	1.09	2.00	0.54	6.04	1.30	6.44	0.90	7.14	0.77	6.36	1.68	5.53	2.01
86	1.96	0.38	2.18	1.14	1.26	1.14	3.00	1.25	6.14	1.39	2.50	0.74	5.26	1.54	6.26	1.36	6.46	1.37	6.04	1.25	7.53	1.93
87	1.32	0.48	2.96	0.65	4.78	2.13	1.83	1.46	2.00	1.07	4.50	1.54	6.48	1.93	7.13	1.22	3.91	1.54	5.96	3.40	6.80	1.89

(continued)

Table 1. Dimension and preference means and standard deviations associated with each slide (continued).

	Landscape Dimensions -																				Overall landscape preference	
	Sky		Foreground Vegetation		Green Colors		Blue Colors		Mountains		Man-changed area		Unobstructed expanse of view		Visible distant landforms		Undisturbed forest		Clouds			
Slide	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
88	2.59	1.01	3.54	1.94	3.26	1.48	4.33	1.24	1.04	0.21	5.18	1.79	6.70	1.92	4.09	1.68	2.04	1.68	2.82	1.94	7.56	1.63
89	1.46	0.60	3.14	1.08	4.70	2.01	4.38	1.74	7.14	1.21	1.18	0.40	5.56	2.19	6.78	1.31	7.46	1.79	7.50	2.36	5.17	2.12
90	2.18	0.59	4.09	1.54	5.91	1.12	2.79	0.72	5.18	1.01	1.18	0.59	6.48	1.78	4.39	1.08	6.82	1.10	6.54	1.18	4.95	1.64
91	3.32	0.94	3.46	1.06	5.22	1.35	3.04	0.75	3.09	1.11	3.41	0.73	5.22	2.09	5.39	0.94	5.18	1.65	3.82	2.58	5.17	1.79
92	2.59	0.67	3.41	0.80	4.22	1.04	4.08	1.10	2.41	1.33	2.36	0.58	3.65	1.92	4.09	1.00	5.23	1.07	4.91	1.34	5.37	1.79
93*	3.14	0.71	2.86	0.94	5.39	0.89	3.42	0.58	2.68	1.29	2.27	0.70	4.56	1.56	5.00	1.35	6.14	1.12	3.73	1.64	3.88	1.67
94	3.27	0.77	2.91	1.38	4.35	1.43	3.92	0.88	3.68	1.46	1.91	0.53	4.44	1.78	4.70	1.18	6.27	1.12	2.64	1.43	6.04	1.95
95	2.36	0.85	2.82	2.08	2.70	1.84	3.29	0.69	4.77	0.68	1.09	0.29	8.39	1.64	4.74	1.36	5.54	1.79	5.82	1.47	6.24	2.12
96	3.23	0.61	2.64	1.73	4.83	1.34	3.58	0.78	2.36	0.58	3.04	0.84	7.26	1.21	4.52	1.56	4.36	1.65	2.64	1.05	7.30	1.56
97	4.50	0.96	3.59	1.18	3.52	1.16	4.25	0.90	3.27	1.08	1.54	0.67	5.83	1.61	4.44	0.94	5.59	1.44	2.27	1.42	6.44	1.91
98	1.82	1.22	4.18	1.01	3.70	2.20	2.12	1.08	6.41	1.33	1.91	0.61	6.87	1.42	5.78	1.24	7.32	1.04	7.45	2.94	5.53	2.14
99	2.82	0.50	2.82	1.33	6.44	1.16	2.46	0.78	5.04	1.50	2.14	0.77	7.74	1.74	6.04	1.49	6.46	1.37	4.18	.3.84	3.72	2.02
100	3.09	0.53	1.46	1.10	4.83	1.11	2.92	0.72	2.32	0.57	4.00	1.16	4.35	1.56	5.26	1.63	4.14	1.73	3.46	3.17	4.86	2.36
Sample size	22		22		23		24		22		22		23		23		22		22		78	

*These slides had standard deviations of 1.00 or less on 5 or more of the 10 dimensions.

APPENDIX C

Table II. Composite predictive efficiency (R^2 from JAN II) at each stage of the grouping process for various combinations of slides and subjects.[a]

Number of Groups	All 100 slides	Last 80 slides	First 50 slides	Last 50 slides	First 33 slides	Middle 33 slides	Last 34 slides	45 low variance slides[b]	100 slides, "Reliable" S_s only[c]	First 33 slides, "Reliable" S_s only[c]
78	.358	.389	.452	.457	.591	.550	.503	.442	---	---
77	.357	.389	.450	.456	.588	.547	.502	.441	---	---
76	.357	.388	.449	.455	.587	.545	.501	.441	---	---
75	.356	.386	.449	.454	.586	.545	.500	.439	---	---
74	.355	.386	.448	.453	.585	.539	.497	.437	---	---
73	.355	.385	.447	.453	.584	.538	.496	.433	---	---
72	.355	.384	.446	.452	.582	.535	.494	.431	---	---
71	.354	.384	.445	.451	.578	.532	.493	.429	---	---
70	.353	.383	.443	.449	.577	.529	.492	.427	---	---
69	.352	.383	.442	.448	.576	.527	.490	.426	---	---
68	.351	.382	.442	.446	.575	.525	.489	.424	---	---
67	.350	.380	.440	.445	.570	.524	.488	.422	---	---
66	.349	.380	.437	.442	.569	.521	.485	.423	---	---
65	.349	.379	.435	.437	.566	.519	.484	.420	---	---
64	.349	.378	.434	.436	.565	.517	.481	.418	---	---
63	.347	.377	.434	.435	.563	.516	.479	.414	---	---
62	.345	.375	.432	.432	.559	.515	.475	.413	---	---
61	.344	.374	.431	.431	.556	.511	.474	.411	---	---
60	.343	.373	.427	.430	.556	.507	.472	.410	---	---
59	.341	.372	.424	.429	.554	.505	.470	.408	---	---
58	.341	.371	.421	.427	.552	.503	.467	.404	---	---
57	.338	.370	.420	.426	.551	.502	.464	.402	---	---
56	.338	.369	.417	.425	.550	.501	.463	.398	---	---
55	.336	.368	.414	.424	.548	.498	.462	.396	---	---
54	.334	.367	.413	.422	.546	.495	.460	.394	---	---
53	.333	.366	.409	.420	.544	.492	.457	.394	---	---
52	.331	.365	.408	.418	.544	.491	.455	.392	---	---

(continued)

Table II. Composite predictive efficiency (R^2 from JAN II) at each stage of the grouping process for various combinations of slides and subjects (continued).[a]

Number of Groups	All 100 slides	Last 80 slides	First 50 slides	Last 50 slides	First 33 slides	Middle 33 slides	Last 34 slides	45 low variance slides[b]	100 slides, "Reliable" S_s only[c]	First 33 slides, "Reliable" S_s only[c]
51	.328	.362	.406	.417	.542	.488	.453	.390	---	---
50	.326	.361	.404	.414	.539	.480	.451	.389	---	---
49	.323	.359	.401	.410	.537	.478	.449	.385	---	---
48	.321	.358	.399	.408	.536	.477	.444	.382	.363	.620
47	.320	.357	.396	.406	.531	.474	.442	.377	.363	.616
46	.319	.355	.394	.404	.528	.471	.438	.375	.362	.614
45	.316	.352	.393	.404	.525	.467	.435	.374	.361	.611
44	.315	.350	.391	.398	.524	.465	.432	.372	.360	.603
43	.312	.347	.387	.398	.519	.463	.428	.372	.358	.601
42	.310	.347	.386	.396	.515	.457	.422	.368	.357	.596
41	.307	.346	.385	.394	.513	.453	.420	.364	.355	.593
40	.304	.343	.384	.390	.512	.447	.416	.361	.354	.589
39	.301	.340	.381	.385	.509	.444	.415	.359	.352	.587
38	.300	.338	.376	.383	.506	.440	.410	.358	.349	.583
37	.299	.336	.373	.381	.504	.438	.406	.355	.347	.580
36	.298	.336	.370	.378	.501	.436	.404	.353	.346	.576
35	.296	.334	.366	.375	.498	.431	.400	.349	.342	.573
34	.294	.331	.365	.372	.492	.429	.396	.346	.340	.568
33	.291	.330	.363	.369	.489	.424	.394	.344	.337	.564
32	.287	.329	.361	.368	.487	.421	.391	.341	.336	.561
31	.284	.328	.359	.364	.484	.410	.386	.338	.332	.553
30	.283	.326	.354	.361	.479	.410	.382	.336	.328	.549
29	.282	.324	.350	.357	.478	.405	.380	.329	.326	.546
28	.279	.321	.349	.352	.472	.402	.370	.321	.317	.545
27	.278	.314	.347	.350	.468	.392	.361	.319	.314	.543
26	.276	.312	.345	.342	.464	.389	.354	.318	.310	.538
25	.272	.310	.341	.340	.458	.385	.349	.316	.307	.533

(continued)

Table II. Composite predictive efficiency (R^2 from JAN II) at each stage of the grouping process for various combinations of slides and subjects (continued).[a]

Number of Groups	All 100 slides	Last 80 slides	First 50 slides	Last 50 slides	First 33 slides	Middle 33 slides	Last 34 slides	45 low variance slides[b]	100 slides, "Reliable" S_s only[c]	First 33 slides, "Reliable" S_s only[c]
24	.267	.302	.336	.334	.454	.383	.346	.312	.301	.525
23	.263	.298	.331	.331	.450	.378	.338	.309	.298	.522
22	.259	.296	.328	.324	.446	.373	.334	.306	.293	.518
21	.258	.294	.326	.320	.441	.370	.329	.298	.291	.511
20	.257	.292	.323	.313	.436	.368	.323	.295	.289	.502
19	.254	.282	.320	.311	.430	.360	.315	.291	.281	.490
18	.251	.279	.314	.309	.426	.356	.309	.285	.278	.488
17	.249	.276	.310	.305	.422	.352	.306	.281	.274	.479
16	.245	.271	.307	.302	.417	.345	.301	.279	.266	.471
15	.238	.267	.304	.286	.412	.340	.297	.269	.263	.461
14	.236	.264	.301	.279	.404	.338	.289	.265	.260	.454
13	.233	.260	.296	.269	.399	.326	.280	.263	.255	.448
12	.231	.254	.290	.266	.393	.320	.274	.251	.250	.444
11	.228	.249	.282	.258	.388	.295	.269	.246	.243	.437
10	.218	.245	.280	.254	.383	.288	.258	.299	.237	.429
9	.213	.237	.272	.247	.375	.271	.248	.224	.232	.422
8	.207	.236	.267	.241	.357	.264	.244	.221	.228	.414
7	.201	.226	.263	.235	.351	.257	.239	.205	.222	.401
6	.194	.198	.255	.231	.340	.249	.212	.172	.218	.389
5	.187	.190	.237	.212	.314	.242	.204	.167	.208	.374

(continued)

Table II. Composite predictive efficiency (R^2 from JAN II) at each stage of the grouping process for various combinations of slides and subjects (continued).[a]

Number of Groups	All 100 slides	Last 80 slides	First 50 slides	Last 50 slides	First 33 slides	Middle 33 slides	Last 34 slides	45 low variance slides[b]	100 slides, "Reliable" S_s only[c]	First 33 slides, "Reliable" S_s only[c]
4	.167	.176	.231	.196	.300	.229	.186	.152	.191	.344
3	.161	.165	.216	.189	.294	.219	.176	.142	.185	.333
2	.148	.140	.182	.164	.276	.189	.155	.130	.170	.313
1	.109	.133	.155	.144	.238	.171	.128	.113	.102	.260

[a] Except for the last two columns, these data apply to a subject sample size of 78.

[b] These slides had standard deviations of 1.00 or less on 5 or more of the 10 dimensions (see Appendix B).

[c] Forty-eight (48) subjects whose sums of absolute differences between ratings given the 5 duplicate slides at Time 1 and Time 2 were ≤ 7 (see Table 10).

Table III. Standard error of the estimate at each stage of the JAN II grouping process for various combinations of slides and subjects.[a]

Number of Groups	All 100 slides	Last 80 slides	First 50 slides	Last 50 slides	First 33 slides	Middle 33 slides	Last 34 slides	45 low variance slides[b]	100 slides, "Reliable" S_s only[c]	First 33 slides, "Reliable" S_s only[c]
78	---	---	---	---	---	---	---	---	---	---
77	.028	.038	.075	.078	.162	.194	.086	.109	---	---
76	.048	.043	.086	.090	.164	.226	.092	.171	---	---
75	.051	.054	.095	.092	.166	.235	.119	.181	---	---
74	.056	.059	.096	.104	.177	.264	.124	.205	---	---
73	.058	.068	.129	.108	.178	.280	.142	.239	---	---
72	.093	.093	.140	.110	.181	.304	.147	.244	---	---
71	.099	.096	.160	.117	.202	.327	.162	.261	---	---
70	.100	.100	.172	.138	.205	.330	.190	.270	---	---
69	.104	.100	.172	.148	.210	.332	.190	.271	---	---
68	.104	.102	.173	.159	.218	.353	.191	.280	---	---
67	.108	.103	.174	.181	.224	.354	.202	.292	---	---
66	.113	.107	.185	.202	.228	.366	.226	.294	---	---
65	.114	.117	.186	.217	.231	.367	.233	.298	---	---
64	.117	.123	.188	.219	.241	.391	.237	.306	---	---
63	.120	.124	.203	.229	.250	.402	.246	.323	---	---
62	.121	.124	.207	.237	.276	.408	.252	.339	---	---
61	.128	.130	.207	.247	.277	.414	.257	.342	---	---
60	.134	.130	.207	.249	.278	.472	.262	.364	---	---
59	.140	.132	.218	.251	.289	.472	.265	.372	---	---
58	.142	.135	.229	.262	.290	.479	.268	.384	---	---
57	.145	.155	.232	.268	.298	.498	.268	.390	---	---
56	.145	.166	.242	.276	.373	.508	.269	.412	---	---
55	.161	.174	.244	.292	.394	.511	.284	.431	---	---
54	.176	.174	.245	.292	.399	.533	.286	.436	---	---

(continued)

Table III. Standard error of the estimate at each stage of the JAN II grouping process for various combinations of slides and subjects (continued).[a]

Number of Groups	All 100 slides	Last 80 slides	First 50 slides	Last 50 slides	First 33 slides	Middle 33 slides	Last 34 slides	45 low variance slides[b]	100 slides, "Reliable" S_s only[c]	First 33 slides, "Reliable" S_s only[c]
53	.186	.179	.249	.296	.420	.534	.294	.448	---	---
52	.192	.185	.258	.298	.422	.557	.300	.452	---	---
51	.194	.206	.260	.305	.438	.609	.324	.491	---	---
50	.199	.216	.268	.315	.452	.622	.345	.502	---	---
49	.222	.224	.293	.344	.466	.657	.352	.505	---	---
48	.226	.227	.303	.350	.469	.791	.356	.520	---	---
47	.229	.327	.305	.354	.472	.799	.370	.521	.028	.162
46	.247	.260	.321	.358	.501	.807	.373	.523	.074	.178
45	.250	.271	.329	.361	.509	.808	.407	.524	.093	.205
44	.252	.283	.364	.370	.523	.824	.419	.528	.099	.224
43	.262	.289	.370	.373	.527	.840	.437	.534	.104	.228
42	.266	.290	.372	.383	.533	.849	.520	.580	.114	.231
41	.269	.293	.377	.388	.536	.853	.520	.587	.120	.250
40	.271	.294	.410	.414	.551	.859	.548	.594	.128	.358
39	.277	.299	.411	.461	.582	.881	.550	.611	.134	.364
38	.282	.301	.438	.488	.602	.916	.643	.652	.136	.395
37	.282	.302	.448	.492	.610	.970	.646	.695	.137	.406
36	.287	.327	.504	.525	.614	.990	.662	.723	.142	.413
35	.288	.354	.514	.553	.640	1.002	.664	.755	.161	.450
34	.291	.379	.517	.572	.697	1.087	.690	.774	.176	.452
33	.346	.398	.521	.628	.735	1.094	.714	.790	.185	.458
32	.352	.414	.533	.641	.749	1.176	.721	.829	.208	.466
31	.366	.414	.564	.684	.772	1.270	.775	.848	.214	.472
30	.388	.432	.606	.710	.805	1.303	.802	.869	.222	.485
29	.455	.478	.638	.737	.423	1.368	.832	.888	.247	.524
28	.469	.508	.648	.742	.942	1.462	.833	.960	.249	.551
27	.499	.530	.684	.762	.952	1.486	.860	1.006	.250	.560

(continued)

Table III. Standard error of the estimate at each stage of the JAN II grouping process for various combinations of slides and subjects (continued).[a]

Number of Groups	All 100 slides	Last 80 slides	First 50 slides	Last 50 slides	First 33 slides	Middle 33 slides	Last 34 slides	45 low variance slides[b]	100 slides, "Reliable" S_s only[c]	First 33 slides "Reliable" S_s only[c]
26	.547	.432	.685	.835	.957	1.533	.947	1.022	.250	.590
25	.556	.534	.738	.863	1.076	1.630	1.031	1.131	.288	.640
24	.571	.542	.748	.919	1.084	1.632	1.132	1.155	.309	.712
23	.578	.542	.788	.923	1.147	1.736	1.134	1.208	.376	.738
22	.634	.577	.791	.934	1.165	1.771	1.152	1.247	.387	.749
21	.644	.598	.845	.942	1.212	1.847	1.169	1.350	.419	.866
20	.706	.680	.909	1.006	1.381	2.068	1.238	1.495	.429	.928
19	.729	.708	.945	1.043	1.515	2.151	1.250	1.566	.447	.978
18	.755	.713	1.004	1.202	1.520	2.245	1.263	1.573	.455	1.023
17	.794	.869	1.038	1.203	1.562	2.316	1.299	2.011	.540	1.138
16	.896	.983	1.115	1.227	1.762	2.339	1.312	2.262	.619	1.266
15	.911	1.009	1.242	1.586	1.768	2.581	1.512	2.271	.644	1.420
14	.912	1.222	1.409	1.594	1.875	2.872	1.692	2.596	.646	1.520
13	.993	1.253	1.827	1.672	1.887	3.181	1.860	2.614	.703	1.558
12	1.057	1.455	1.951	2.015	1.898	3.192	2.378	2.872	.718	1.657
11	1.388	1.517	1.976	2.105	2.672	4.010	2.402	2.881	.855	1.701
10	1.402	1.564	2.214	2.118	2.673	5.003	2.548	3.622	1.093	1.791
9	1.420	1.564	2.790	2.261	3.712	5.345	2.559	3.840	1.244	1.910
8	1.474	1.632	2.816	2.460	5.247	7.465	2.794	4.417	1.266	2.545
7	1.836	1.955	3.261	3.889	5.380	7.939	3.603	5.186	1.305	3.840
6	2.514	2.356	4.488	3.998	5.760	9.196	3.890	5.628	1.577	3.851

(continued)

Table III. Standard error of the estimate at each stage of the JAN II grouping process for various combinations of slides and subjects (continued).[a]

Number of Groups	All 100 slides	Last 80 slides	First 50 slides	Last 50 slides	First 33 slides	Middle 33 slides	Last 34 slides	45 low variance slides[b]	100 slides, "Reliable" S_s only[c]	First 33 slides "Reliable" S_s only[c]
5	2.749	2.787	5.388	4.917	6.461	9.958	5.317	6.924	1.776	3.874
4	3.102	3.007	6.319	5.898	7.057	17.716	12.324	7.226	2.869	6.386
3	4.859	3.779	7.290	6.528	9.714	19.253	17.690	17.738	2.890	8.118
2	7.854	9.397	12.302	8.611	18.054	23.589	37.092	21.689	6.191	10.770

[a] Except for the last two columns, these data apply to a subject sample size of 78.

[b] These slides had standard deviations of 1.00 or less on 5 or more of the 10 dimensions (see Appendix B).

[c] Forty-eight (48) subjects whose sums of absolute differences between ratings given the 5 duplicate slides at Time 1 and Time 2 were ≤ 7 (see Table 10).

POLICY CAPTURING WITH THE USE OF VISUAL STIMULI: A METHOD OF QUANTIFYING THE DETERMINANTS OF LANDSCAPE PREFERENCE

by

Dennis B. Propst

(ABSTRACT)

Policy Capturing, a potential methodology for evaluating landscape preference, was described and tested. This methodology, tested and applied in industrial psychology since 1960, results in a mathematical model that theoretically represents the human decision-making process. Under experimental conditions, judges were asked to express their preferences for scenes of the Blue Ridge Parkway (Virginia and North Carolina). A multivariate linear model simulating each judge's decision was computed by regressing landscape preferences onto 10 dimensions thought to influence such preferences. An equation which "captures," or defines the policy of each judge was generated. Individual equations were then clustered and a composite strategy calculated at each step until one overall policy was obtained. Coefficients of determination (R^2's) for some individuals were generally large (> 0.50). However, composite R^2's were fairly low (< 0.25). Methodological problems concerning the use of policy capturing for landscape assessment along with practical management applications are discussed.