RATER FAMILIARITY IN SIMULATION VALIDITY STUDIES

by

CARROLL OWEN MCCORMICK

B.A. (HONOURS) ACADIA UNIVERSITY, 1983

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Department of Psychology

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

May, 1986

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of **Psychology**

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date **May 6, 1986**

Abstract

Previous research on photographic simulation validity is reviewed. Evidence supporting simulation validity is found, but methodological flaws seriously compromise the results of many of the studies. In the present study 408 University of British Columbia undergraduates rated the affective quality of two building interiors, using actual site, color slide, and written description media. Groups rating each site at each medium were equally represented by familiar and unfamiliar raters. The hypothesis that compensating efects of rater familiarity with stimulus sites produce results that are misinterpreted as supportive of simulation validity was tested. Results show that valid ratings of color slides were obtained, but results that appear supportive of simulation validity can be confounded with rater familiarity effects and building prototypicality.

Dr. R. S. Corteen

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

I would like to thank Frank Flynn for writing the scramble and descramble programs for the dependent measure, and Marsha Schroeder for her able and generous advice on matters statistical. I wish to acknowledge the contributions of the many people who assisted in making this project possible, including the professors who granted me class time for the collection of data, the student participants, and the Psychology department staff. My thanks also to my graduate student colleagues for their expert advice, and for their encouragement and friendship that has been so vital during the past three years.

To my advisor Ray I wish to extend my heartfelt thanks for his advice and guidance during the course of this project, and for his assistance in successfully overcoming the bewildering array of challenges that have had to be met in the course of the past three years.

To two special people - Jim and Diane - thank you for the sharing.

Introduction

The last time I went camping I shot a few rolls of film to bring back with me as a reminder of where I had been and what I had done. They were meant mainly for my own use, but I also hoped to share my experiences by showing them to friends. My goal was no different from that of many others: To enable other people to experience vicariously, with the aid of photography, what it would have been like, and how they would have felt, had they been there themselves.

It is taken for granted that photographic representations can capture the essential qualities of reality. This faith in the medium is responsible for countless slide shows of "our vacation", and for the tours through photograph albums that endlessly delight their proud owners and often dismay their guests.

In scientific and other professional endeavours, photographic representations of places and events are also extensively employed, their use guided as much by the same common sense assumptions that are held by people who take photographs for their own pleasure as by empirical evidence of their usefulness. These methods of representing places and events are usually referred to as simulations, although others with a bent for nomeclature insist they be called proxies (Coughlin & Goldstein, 1970) or surrogates (Daniel & Ittelson, 1981; Shuttleworth, 1980).

The photographic media, including all those based on silver nitrate, and electronic video equipment, are among a large variety of simulations used in many disciplines. Also in use are scale models, movies of scale models, line drawings, short verbal labels, detailed verbal descriptions, and computer modeling techniques. Mekechnie (1977) makes a distinction between static (e.g., photographic), dynamic (e.g., movie), concrete (e.g., scale models), and abstract (e.g., computer modeling) forms of simulations, to which can be added the category of verbal descriptions.

Of primary interest to this discussion is the use of static, and to a lesser degree verbal descriptions, in obtaining user preference ratings in a variety of applications such as environmental cognition research, outlined by Evans (1980), and Moore (1979), architectural design, and landscape aesthetics.

A small sample of such research includes the study of riverscape aesthetics in the Canadian Praries (Pomeroy, Green & Fitzgibbon, 1983), waterscapes (Herzog, 1985), landscape appraisal (Craik, 1972), visual air quality research (Rowe & Chestnut, 1983), the development of landscape preference models (Daniel & Boster, 1976; Brush & Shafer, 1975), obtaining affective appraisals of environments (Russell & Lanius, 1984), and determining basic dimensions of affect (Russell & Pratt, 1980; Russell, 1979, 1978).

The practicality of simulations in research makes it very popular. It is far less expensive and time consuming to have people rate simulations than to take them to actual sites. Valuable user responses to buildings can be obtained before construction begins (e.g., Carpman, Grant, & Simmons, 1985), and variables such as the weather and the time of day can be reliably controlled by the researcher (Howard, Mlynarski, & Sauer, 1972). Without simulations the undertaking of many research projects would be financially prohibitive.

Surrounding the widespread use of simulations in research is the important but insufficiently understood issue of simulation validity; whether ratings obtained using simulations are the same or acceptably similar to ratings obtained from actual sites themselves. A number of simulation validity studies have been done, but knowledge about simulation validity comes largely from one-shot studies. There does not exist an integrated and systematically developed literature of in-depth research on the validity of simulations issue. It also appears that those studies that have been done have been underutilized as sources of information on designing better, methodologically improved simulation validity experiments.

These studies provide few empirically derived procedures for using simulations, a shortcoming also noted by Nassauer (1983). To be fair, advice can be culled from the literature on issues such as color versus black/white film, normal versus wide angle lenses, slides versus prints, and some viewer variables. In fact, most authors of these studies have endorsed to some degree the validity of using simulations. Whether the endorsements are justified or not will be addressed in detail later.

Studies employing simulations occasionally refer to simulation validity studies for an endorsement of the technique. It is not the purpose of this discussion to chronicle the application of knowledge obtained from simulation validity studies, but in the course of researching this topic some observations were made that bear mentioning.

It has been my impression that researchers use simulations in their studies as they see fit, typically employing only routine controls such as randomization procedures in presenting photographs. It is common to note that researchers make no mention of the validity of the method they employed (e.g., Carpman, Grant & Simmons, 1985; Chalmers & Knight, 1985; Herzog, 1985; Calvin, Dearing & Curtin, 1972), an oversight noted by Shuttleworth (1980). In another study (Pomeroy, Green & Fitzgibbon, 1983) the researchers, aware of the issue, stated that the use of simulations had been validated, but without revealing the source of the revelation. Williams (1985), citing some validity studies, received the impression that they had unequivocally established the validity of the procedure. It was typical to read studies that discharged their obligation to the question by interpreting the qualified conclusions of validity studies as endorsements of the method.

I do not know how widespread such practices are, but it is clear that the validity literature has had the effect of creating the impression that simulations can adequately represent real environments. During the course of this review two points will be made clear. First, a simulation study is very complex, a point that has not

been made salient, nor has been sufficiently understood in the simulation validity literature. The second point is that there are sufficient methodological shortcomings in the validity studies reviewed here to limit the number of sound conclusions that can be reached regarding the validity of simulations issue.

A validity study can be seen as being comprised of four components. This should not be regarded as an immutable or exhaustive list of categories, but as an aid in understanding the process and complexity of conducting simulation validity studies.

Every simulation validity study must have one or more 1) stimulus sites, 2) media, 3) raters and, 4) rating scales or dependent measures. When designing such a study, a wide variety of candidates is available to serve as each component, and each can be expected to have an impact on its conclusions.

When discussing simulation validity, particularly with reference to its application in a study, the nature of these components become important. With so many of them, the generalizability of simulation validity results becomes rather complicated. Therefore various items that have been used as components in these studies will be reported.

Many different stimulus sites have been used, and it is possible that the use of simulations may be more valid for some environments than for others. There are diverse media to select from, complicated by additional choices of how to present them. The choice of a normal or wide angle camera lens, the size of a print or projected slide image, the use of single or multiple slides or panoramic photographs, could increase or decrease the validity of ratings of environmental stimuli.

The outcome of any study depends on the choice of dependent measure, which in simulation studies are usually rating scales. Although many of the rating scales reviewed here were designed to measure the same thing -- typically affective responses and visual preferences -- none of the instruments here are quite the same.

In some of the studies there exist generalizability problems because very few, and in some cases only one, scales were used. For some of the instruments nothing is known about their psychometric properties, nor about their usefulness in user preferences studies. When available the origin and nature of the dependent measures will be reported.

Buildings. Buildings used in simulation studies include "prototypical" houses, institutional and commercial buildings (Hershberger & Cass, 1974), buildings and their interior rooms (Howard et al., 1972), and a full scale mock up of a study bedroom (Lau, 1969). Media include single and multiple color slides (Danford & Willems, 1975; Hershberger & Cass, 1974; Howard et al., 1972), detailed verbal descriptions (Danford & Willems, 1975), scale models (Seaton & Collins, 1972; Lau, 1969), Super 8 mm color and black/white film, black/white video tape (Hershberger & Cass, 1974), black/white slides (Howard et al., 1972), color and black/white photographs (Seaton & Collins, 1972). Howard et al. (1972) used multiple slides covering approximately 150^{o} of visual angle, taken with a 35 mm wide angle lens. The projected size of the slides was 26 in x 39 in.

Hershberger and Cass (1974), using five simulation media, found that they all performed very well, but that the ratings obtained from the color film were most similar, and that those obtained from the multiple slides were least similar, to the actual site ratings. In their second experiment they expressed no preference between the two simulation media that were used -- single 35 mm color slides and Super 8 mm color film. Howard et al. (1972) noted little difference in the way the color and black/white slide media used in their study were rated, and Seaton and Collins (1972) were more satisfied with the color than the black/white photographs as simulation media.

With respect to viewer variables Hershberger and Cass (1974) used architectural and pre-architectural students. Seaton and Collins (1972) used visitors to a 1970 University of British Columbia open house, assuming that they were unfamiliar with the stimulus sites. Only Danford and Willems (1975) and Howard et al. (1972) included viewer variables in their study designs. In the Danford and Willems (1975) study, half of each of the first three treatment groups were familiar with the building, and half were not. In the fourth group none of the subjects were familiar with the building. Each group was equally represented by males and females. No main effects were found for gender or familiarity.

Howard et al. (1972) included familiarity as an independent variable in their study. All of the viewers were familiar with the two public buildings, and none were familiar with the two private buildings, used as stimulus sites. The viewers in each of the three media conditions -- actual site, color slides, and black/white slides -- rated all four buildings and interior rooms.

The scores for the familiar-public buildings showed some regression toward the mean (less extreme ratings) and those for the unfamilar-private buildings showed a shift toward the negative end of the scales. The authors thought it possible to reverse regression and negation by increasing the brightness of the slides.

No effects for familiarity were found. This is not suprising since viewer familiarity was confounded with the buildings: Since all of the subjects were familiar or unfamilar with the same pairs of buildings, there were no familiar and unfamiliar ratings of any one building to compare to each other. The only way that effects of viewer familiarity on ratings of stimulus sites can be tested is to have both unfamiliar and familiar viewers rate the same building, and then compare the ratings. This was not done, so there was no way to seperate effects due to the properties of the buildings from those due to the different levels of viewer familiarity.

All of the studies used verbal rating scales in unipolar and semantic differential formats as dependent measures. Danford and Willems (1975) used 34 single adjective unipolar scales, selecting from previous studies adjectives thought to be descriptive and affective in nature. They reported high within-group rater reliability. Hershberger and Cass (1974) used 30 scales designed to measure presentational, affective, and evaluative areas of architectural meaning, and 30 "factor representative" scales. Howard et al. (1972) used 28 bipolar affect scales: Based mainly on data reported in Osgood, Suci, and Tannenbaum (1957) they selected scales so that about 1/3 of them loaded highly on Evaluation, 1/3 on Potency, and 1/3 on Activity. Seaton and Collins used five unipolar scales, chosen from adjectives that loaded on factors reported in other studies.

Enthusiasm for the usefulness of photographic simulations varied greatly among the studies, although on the whole considerable support was generated in favor of photographic simulation validity.

The conclusions of Danford and Willems (1975) were the most unequivocally negative. They concluded not only that validity had not been supported in their study, but that simulations in general are not even amenable to validation attempts because of problems with investigations of a cognitive nature, i.e., having raters assess anything with rating scales. If correct, these would be serious claims, but their study design suggests less ominous explanations that might account for the similar ratings that were obtained between the media.

A closer look at their study is in order: Group 1 was given a 16 minute guided tour of the building concurrent with information about the building. Group 2 was given a simulated tour with 62 color slides and the same information. Group 3 was simply asked to imagine what a law school -- the stimulus site -- would look like, and how it

would make one feel. Group 4, of which none of its members were familiar with the building, just saw the slides.

Danford and Willems (1975) believed that both convergent and divergent responses had to be present as preconditions for concluding that simulations were valid. This is why they did not regard the extremly high rate of agreement between Groups 1 and 2 by itself as an endorsement of the validity of using slides as simulations. Convergent responses were sought between media -- Groups 1 and 2 -- but in addition, divergent or differing responses were required between the first two and the second two groups. Their reasoning was as follows: Due to the absence of color slides in Group 3, the absence of information about the building in Group 4, and the unfamiliarity of Group 4 with the building, the ratings by Groups 3 and 4 should differ significantly from those of Groups 1 and 2. Because the responses between all four groups were so similar, they concluded that, "the convergent responses -- or absence of divergent responses -- by Groups 3 and 4 clearly suggest that the responses were determined or constrained by the technique of presentation and measurement, and, therefore, are useless for practical purposes" (p. 512).

It must be emphasized that the authors presented no convincing rationale for why the responses of Group 4 had to be different from those of Groups 1 or 2 just because the raters were unfamiliar with the stimulus buildings. A more cautious stance would have been to simply have concluded that familiarity had had no effect on viewer ratings (which was the case with the other groups): Divergent findings should not have been expected between Group 4 and Groups 1 and or 2. Then there is the matter of their dependent measure. Apparently, little was known about the appropriateness of the adjectives for rating buildings, and nothing was indicated about the dependent measure's ability to distinguish between different buildings. Looking at the scales it is not at all difficult to imagine any reasonably clean and well lit

concrete and glass building receiving similar ratings. This could account for why Group 3 obtained ratings so similar to those of the other groups. It is quite possible that Danford and Willems (1975) did fail to demonstrate simulation validity, but they certainly did not build a good case for the futility of testing simulation validity, or the uselessness of rating anything with rating scales.

In their first experiment Hershberger and Cass (1974) found very similar factor structures for the real buildings and each of the five media types. Multiple comparison procedures revealed significant differences between actual site and simulation ratings for only 5/30 scales. There were 2 with the multiple color slides and video tape, 3 with the black/white film, and 1 with the color slides. There were no differences in the ratings of the actual site and the color film.

In the second experiment six of eight factors were very similar across the three media, with warmth and lighting factors absent from the simulation media. Multiple comparison procedures revealed significant differences between the two simulation media and the actual site on 12 scales. The buildings viewed in person were judged as significantly more good, beautiful, pleasing, friendly and unique than when judged from the slides or film, and as more quiet and safe.

The finding that the simulations were judged more negatively than the actual sites is in agreement with those of Howard et al. (1972). These findings notwithstanding, Hershberger and Cass (1974) concluded that the results of their preliminary evaluation suggested that color film and color slides can be used to simulate actual designed environments for the purpose of obtaining evaluative judgements.

Howard et al. (1972) factor analyzed their data and concluded that, "... the subjects in each condition used the semantic-differential scales in the same way conceptually" (p. 6-6-3). Despite some main effects for media, they concluded that

simulations can be used for rating actual buildings if it is kept in mind that the simulations were rated more negatively than the actual sites.

Seaton and Collins (1972) suggested that the evaluative judgements made from the simulations were similar to those made from the actual buildings, but they did express a certain amount of skepticism about their use. Lau (1969) found that full scale mock-ups and scale models of a study bedroom were rated similarly on dimensions of pleasantness and gloom. He mentioned that reduced scale can enhance the quality of an object or of a building interior.

There are aspects of the Hershberger and Cass (1974), Howard et al. (1972), and Seaton and Collins (1972) studies that make it difficult to assess just how much they learned about simulation validity. Their conclusions certainly suggest that the ratings from the simulations were valid, but were the dependent measures appropriate?

Hershberger and Cass (1974) obtained ratings on a total of 24 buildings. It would have been helpful had they reported data on how well the scales distinguished between buildings. After all, if a lot of the buildings were rated similarly, it suggests that the instrument might not rate buildings differently, let alone different different media of the same building. For the same reasons Howard et al. (1972) should have reported whether there were building main effects.

There are indications that if in fact the response instruments did work well, the simulations may have performed better than either Hershberger and Cass (1974) or Howard et al. (1972) reported. Hershberger and Cass (1974) performed multiple analyses of variance, followed up by analyses of variance and multiple comparison procedures. A visual inspection of the profile of scores across media that they reported gives the impression that they did not use a familywise error rate such as the Bonferroni calls for by setting the alpha for each scale at .05/30=.002. Alpha levels were not

reported, and it is quite possible that they obtained more significant differences than they would have had they used more conservative and appropriate alphas.

Howard et al. (1972) reported significance levels set at .05 for each of 28 scales. This certainly led to a greatly inflated Type I error rate. Furthermore, t-tests were conducted on all posible pairs of media for each scale that was significant at .05. The alpha for each t-test was not reported, but was likely set at .05. Such an alpha, combined with an inappropriate use of t-tests for multiple comparison procedures (see Glass and Hopkins, 1984) would have greatly inflated the Type I error rate, or the number of false positives or differences obtained in their results.

In sum, the studies using buildings show promise with respect to simulation validity. The unknown quantity that limits the extent to which the results can be endorsed is the dependent measure.

Landscapes. A variety of outdoor settings has been used. In addition to places simply referred to as landscapes, wilderness campsites (Shelby & Harris, 1985), scenic highway corridors (Evans & Wood, 1980), coastal landscapes (Nassauer, 1983), and small streams (Rabinowitz & Coughlin, 1971) have served as stimulus sites. Simulation media include aerial photographs (Zube, 1974), color photographs (Dunn, 1976; Zube et al., 1975; Shelby & Harris, 1985), color, and black/white prints (Shuttleworth, 1980), color slides (Rabinowitz & Coughlin, 1971), and written descriptions (Shelby & Harris, 1985).

Seung-Bin Im (1984) reported taking his color slides and prints with a 24 mm wide angle lens. It was not reported whether the prints were color or black/white. Using the prints he constructed panoramic photographs subtending a visual angle of about 240^{o}. Shelby and Harris (1985) showed raters 5 in x 7 in photographs taken with a 28 mm wide angle lens. Shuttleworth (1980) showed subjects 7 in x 5 in prints taken with a 35 mm wide angle lens.

Nassauer (1983) compared wide angle slides taken with a 35 mm lens and covering 63^{o} of view with sets of three slides taken with a 50 mm lens comprising a panoramic view covering 140^{o}. They were projected to form images measuring 100 cm x 147 cm. He found that the slides forming the panoramic view and the single wide angle slide were given different visual preference ratings for 9/17 landscapes, but that the rank ordering of the landscapes in the two simulation media correlated highly. He reported no significant differences between the two simulation media on ratings obtained on 10 adjective pairs. For reasons which will be discussed later, it is not possible to ascertain which simulation best represented the actual sites.

Shuttleworth (1980) found that ratings of color prints more closely approximated the ratings of the actual sites than did those of the black/white prints. Subjects gave more extreme and varied responses to the black/white prints than to the color prints. Seung-Bin Im (1984) concluded that subjects' ratings of the panoramic prints and color

slides were the same. This must be regarded with caution, since the two simulation media were rated with two different rating techniques -- paired comparisons, and a single visual preference scale -- and thus cannot be compared with any confidence.

Seung-Bin Im (1984) included viewer familiarity as a subject variable in his design. All of the subjects were in residence at the university from where the stimulus sites were selected, and all were familiar with the campus. Their length of residence and the number of visits to each site was correlated with their visual preference scores -- the correlations were very low. This was understood to mean that their level of familiarity had had no effect on their ratings. Since all of the subjects were familiar with the campus, the effects of unfamiliarity on ratings could not be ascertained.

Shuttleworth (1980) used students, all of whom were familiar with the landscapes used in the study. Zube, Pitt, and Anderson (1975) used eight professional, non-professional, and student groups. They found a high overall rate of agreement among these expert and non-expert groups. Rabinowitz and Coughlin (1971) used upper class suburban housewives. Using students and non-students, Evans and Wood (1980) found no significant differences between them in their ratings of the actual and simulated scenes.

A variety of dependent measures were used in the studies. Seung-Bin Im (1984) had subjects rate scenes on a single 10-point like-dislike scale, with 10 representing the highest preference, and 1 the lowest. Subjects also made paired comparisons of the panoramic photographs. By having a second group of subjects rate a new set of slides of the same places, he concluded that the visual preference scale was reliable. The scores of the two groups correlated .927, $p.=.0027$. Shuttleworth (1980) used seven 7-point semantic differentials and a rank order questionnaire. Dunn (1976) had

respondents answer four questions about their relative preference for the six landscapes represented in the photographs, and the reasons for their preferences.

Zube, et al. (1975) used three dependent measures -- 18 landscape description and evaluation semantic differential scales, a landscape feature checklist, and four questions about the landscapes, the first three of which concerned desire for residency, and the fourth intended as a scenic evaluation question. Rabinowitz and Coughlin (1971) used 14 rating scales, 15 semantic differentials, and 3 basic preference questions. Nassauer (1983) used a single 10-point visual quality scale, 10 being very high visual quality, and 1 very low visual quality. He also administered a mixture of 10 affective and descriptive adjective pairs ranked on a 7-point scale. Evans and Wood (1980) selected 15 bipolar adjective scales. They were chosen because they had loaded highly on landscape assessment factors in previous research, or had been shown to be sensitive to different levels of scenic quality. An additional high-low scenic quality scale was added. A Crombach's alpha coefficient of .87 indicated high measurement reliability.

In general, the results of the landscape validity studies led their authors to conclude that simulations can serve as valid representations of real environments. The Zube et al. (1975) results indicated that the color photographs were effective simulations for eliciting evaluative responses, at least for the Connecticut River Valley landscape.

Shuttleworth (1980) concluded that there were very few significant differences between the ratings of photographs of the sites and the ratings of the actual sites. Ratings of the black/white photographs were more extreme and highly differentiated than the color photographs, which more closely approximated the actual site ratings. It was concluded that the use of photographic simulations are valid provided that color photographs are used, and are taken with a wide angle lens.

Evans and Wood (1980) compared ratings of roadside scenes to ratings obtained from viewing a series of slides that simulated a drive along the same highway. This was achieved by using two projectors to fade in and fade out 100 slides so as to create an effect similar to that of a movie camera. A correlation of .79, $p.<.001$ was obtained between the on site and simulated drive ratings. They concluded that the use of simulations was valid. They went on to stress that the ecological validity of their findings would be strengthened if they were replicated with other scenic roads, and with other types of scenic highway users. Zube (1974) found a high rate of agreement between field ratings of landscapes and aerial photographs.

Shelby and Harris (1985), in a comparison of methods for determining visitor evaluations of ecological impacts, had subjects rate 20 wilderness campsites, color photographs and written descriptions of them. Three dependent measures were used: Two 5-point scales to rate the acceptability of specific impacts found at each campsite and the overall desirability of each campsite, and an overall ranking of campsites in order of preference. Photographs and written descriptions agreed with the on site evaluations 90% and 80% of the time, respectively. Evaluations of more general characteristics showed less agreement -- 75% for the photographs and 65% for the written descriptions. The preference rankings of the written descriptions was slightly higher at 80% of the sites, than the photographs. They were most satisfied with the simulation media for evaluating specific impacts. Clearly the generalizability of these findings are greatly restricted since only two scales and a rank ordering procedure were used. Though the written descriptions worked quite well for rating specific impacts, it would be incautious to assume this finding to hold for other stimuli or more complex tasks, or with other rating scales.

Although Nassauer (1983) concluded that his study demonstrated simulation validity, there are two bases upon which this must be contested. When the panoramic

and wide angle photographs were compared, 9 of the 17 landscapes were rated as significantly different from each other on the visual quality scale, indicating that at least on this scale there was less than 50% agreement between the two media. The actual sites were not rated on the visual quality scale, so the rate of agreement between the simulation media and actual site condition was not determined. Clearly the rate of agreement would not have exceeded 50% of the landscapes: This follows because the simulation media ratings differed on 50% of the landscapes, and on those landscapes only one of the ratings could have been the same as the actual site rating.

On the whole, the simulation media of the landscapes were not rated as significantly different from each other on the 10 bipolar adjective scales. Note though, that since these scales were not used to rate the actual sites, the question of simulation validity using these scales must be left unanswered.

The methodological problems in this study disqualify it from making any conclusions about simulation validity. Most unacceptable was that different scales and rating procedures were used for the actual site and the simulation raters: Baseline or control group data in simulation validity studies must come from actual site ratings. In its absence, or when it is not comparable to data collected from other media, questions of simulation validity cannot be addressed.

Seung-Bin Im (1984) reported a high correlation (r=.923, p.<.003) between the visual preference values of on site and slide evaluations. An interrater reliability rating was obtained between two groups that had each rated the same 12 slides -- the preference values correlated .927, p.<.0027. Although he concluded that photographic simulations are acceptable for evaluating visual preferences, the use of a single scale to obtain ratings severely limit the generalizability of the results.

Dunn's (1976) conclusion that his admittedly crude and primitive attempt at assessing validity provided evidence supporting the validity of using photographic

simulations cannot be taken seriously: Respondents were asked to compare the site where they had just been interviewed with the photographs of the other sites that they had not seen. This is not an acceptable experimental design.

Rabinowitz and Coughlin (1971) tentatively concluded that the rating of color slides is a valid technique. The best results were reported to have come from the multiple color slides shown subjects in their homes, presented with a hand viewer. The low amount of variance accounted for by the landscapes was attributed to a wide variation of slide quality. They also suggested that explicit instructions to rate the site represented by the slide increased the validity of the ratings.

The above studies represent diverse approaches to testing simulation validity. A number of studies with no obvious methodological drawbacks provide evidence in favor of simulation validity. An encouraging aspect of all of the studies taken together is that using a wide variety of stimulus sites, media, and raters, results were mainly in the direction favouring simulation validity. On the other hand many researchers restricted the generalizability of their findings by using too few rating scales. Many of the results are of unknown value because of an absence of information on the properties and performance characteristics of the dependent measures.

There are some technical issues to be addressed in a systematic way. In particular it would be of great value to know what effect the size of a projected slide image or photograph has on the validity of ratings. Further elucidation of such issues will await further research.

Rater familiarity. A number of rater variables were examined, including viewer familiar with stimuli. Nassauer (1983) alluded to possible influences of familiarity with stimulus sites on ratings, writing "That the photograph is acting as a symbol or a trigger to memory of field experience seems possible ..." (p. 1). Other researchers have specifically controlled for viewer familiarity in the selection of subjects for their

studies. Seaton and Collins (1972) recruited visitors to a 1970 UBC open house as judges because it was decided that "... it would be better if judgements were not contaminated by previous knowledge about the particular buildings" (p. 6-10-3). It was not explained what was meant by the word "contaminate".

Shuttleworth (1980) held the opposite view when selecting subjects. He reasoned that those familiar with the landscapes might be better able to descriminate between a site viewed in the field and the same site viewed in a photograph, thus maximizing the possibility of differences in the results obtained from the different media.

Danford and Willems (1975) and Howard et al. (1972) tested for effects of viewer familiarity. No effects for familiarity, and a confound of familiarity with buildings, respectively, were the results. Shafer and Richards (1974) did not control for subject familiarity, noting only that many of the scenes were familiar to most of the subjects. Seung-Bin Im (1984) used only subjects who were familiar with the sites, albeit with varying levels of familiarity. This study was not designed to test for effects of unfamiliarity on ratings.

Various viewpoints have been put forward on subject familiarity, but they have produced little about possible effects of familiarity of viewers with sites on their ratings of simulations.

The effects of subject familiarity with stimulus sites on their ratings may be of considerable importance in the interpretability of any data obtained from validity studies. Familiarity with stimulus sites may enable raters to compensate for information that is absent from a simulation, an effect opposite to the effect predicted by Shuttleworth (1980).

If raters who are familiar with the stimulus sites are used, any absence of main effects for media could be due to compensating effects of familiarity, and not to qualities of the medium that make it a valid simulation of the actual environment.

Until more is known about what effects familiarity has on how simulations are rated, the interpretability of results obtained in validity studies is in doubt when familiar judges are used.

The present study was designed to investigate three problems. First, it served as a simulation validity study in which the ratings from two actual building interiors were compared with ratings obtained from color slides and written descriptions of the building interiors. Second, half of the subjects were completely unfamiliar with the stimulus sites, and the other half were familiar with the stimulus sites. In this manner the data could be analyzed for effects of viewer familiarity on ratings. Third, the study design allowed the testing of the claim of Daniel and Ittelson (1981) that ratings obtained from verbal rating scales have "nothing (or virtually nothing)" to do with the stimulus being rated.

The precondition for the selection of the two stimulus sites was that they be rated as significantly different from each other on the dependent measure. This enabled testing for convergent and divergent validity between buildings at each medium: The stimulus sites should be rated differently in the actual site and slide conditions, and be rated the same in the written description condition. In addition the use of two measurably different sites would increase the generalizability of the study findings.

The first experiment was a slide study conducted to determine whether the two sites were different as measured by the dependent measure. Following it is the description of the main experiment in which the three issues outlined above are investigated.

Pilot Study

Method

Subjects. Forty-one University of British Columbia summer session students volunteered to participate in the pilot study.

Site selection. The atrium of the D. T. Kenny Psychology building and the lobby of the Scarfe Education building served as the stimulus sites. These buildings were chosen because they are frequented by many of the students from which participants in the study were recruited.

Photographic sampling. Color slides were taken of the stimulus sites using a 35 mm camera with a 28 mm wide angle lens. Vertical and horizontal slides were taken of the atrium in the D. T. Kenny Psychology building from the second floor and from the third floor landing. Ten people who have offices in the building were asked to judge which one of the slides best represented the atrium. The vertical slide shot from the third floor landing was selected for the study. A horizontal shot of the lobby in the Scarfe Education building was taken.

A series of slides was taken of the approach and entry to each building. Three slides were taken along the approach to the Kenny building, as well as one of the interior of the stairwell leading to the third floor landing. Three slides were taken along the approach and entry to the Scarfe building. All of the slides were taken on a sunny day.

Dependent measure. A 40-adjective instrument designed to measure the affective quality of places was chosen. It was selected because of demonstrable reliability, and because some of its psychometric properties were known. A 105-item version was constructed by Russell, Ward, and Pratt (1981), and refined into its present 40-item form by Russell and Pratt (1980). In both forms the instrument has been field tested in several hundred environments. Properties of the instrument and instructions for its

use are reported in detail in Russell and Pratt (1980). For each copy of the dependent measure the adjectives were arranged in random order.

Design. The pilot study was a simple two group design, using the aforementioned Kenny and Scarfe building interiors as stimulus sites. A single simulation medium -- color slides -- was used.

Materials. Booklets were designed for each building, and were identical except for the building names appearing in them. They contained a short explanation of the experiment, instructions, the dependent measure, and a short questionnaire being field tested for the Experiment. The booklets are reproduced in the Appendix.

The slides were shown in classrooms, using a Kodak Ektographic Slide Projector Model AF-2 with an f:3.5 Kodak Projection Zoom Ektamar Lens. The slides were projected onto a portable silver lenticular projection screen. The projector was positioned approximately 3 m from the screen.

Subjects were seated between 3.5 m and 5 m from the screen. The experimenter sat next to the projector. A stopwatch was used to time the presentation duration of the slides.

Procedure. By prior arrangement the experimenter went into classrooms at the end of scheduled lectures. Subjects were given booklets and were instructed to read the first page while the projection equipment was being set up. Rooms were darkened, leaving only enough light so that subjects could read the booklets. Groups were then shown the slides simulating the approach and entry into the building (These approach slides served no specific function in the pilot study; only in the Experiment would they serve their purpose to better equalize the amount of visual information available to actual site and slide condition subjects).

Each slide was shown for the same amount of time that the actual site participants in the Experiment were to take to walk from the spot where one

approach slide was taken to the place from where the next slide was taken. When a slide of a stimulus site was projected, participants were asked to read page two of their booklets. They then viewed the slide for an additional 45 sec before completing the dependent measure and questionnaire. The last slide continued to be shown while the dependent measure was being completed. Participants were then debriefed. It took approximately 15 min to run each group.

Results

A MANOVA was performed on the data, mainly to control for the inflation of Type I error that can result when analyzing correlated multiple dependent measures. The MANOVA was also justified because of the overall theoretical concept of affect that the Russell and Pratt (1980) instrument is presumed to measure.

Before performing the MANOVA, the 40 scale items were combined to form 4 bipolar scales of affective quality, using the procedure outlined in Russell and Pratt (1980). They are Scale 1) arousing-sleepy, Scale 2) exciting-gloomy, Scale 3) pleasant-unpleasant and, Scale 4) distressing-relaxing.

It has been proposed that affect can be represented by two orthogonal dimensions -- pleasure and arousal. They can be plotted in two-dimensional space, with the arousing-sleepy scale forming the vertical axis, and the unpleasant-pleasant scale forming the horizontal axis. The distressing-relaxing and exciting-gloomy scales represent combinations of pleasing and arousing qualities, forming axes bisecting the vertical and horizontal axes. Thus the eight adjectives which form the four bipolar affect scales can be seen as being evenly spaced 45^{o} apart around the perimeter of a circle. For example, distressing is seen as both highly arousing and unpleasant, and exciting as both highly arousing and pleasant. This conceptualization of the underlying dimensions of affect has its origins with Osgood, Suci and Tannenbaum (1957), and has been further developed and refined in Russell, Ward, and Pratt (1981) and Russell

and Pratt (1980). The arrangement of these affective descriptors are reproduced in Figure 1.

.....

Place Figure 1 about here (See p. 24)

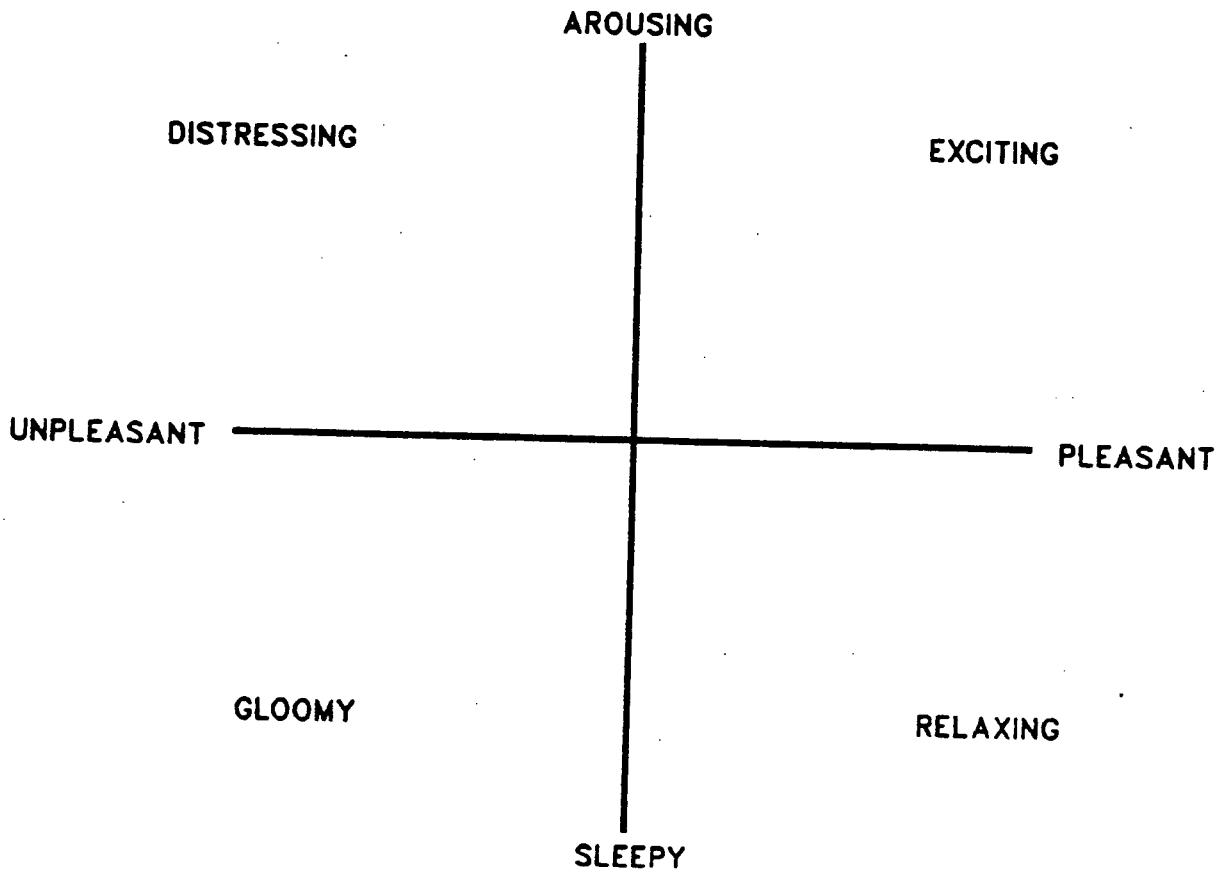
.....

The Wilks Lambda was used for the multivariate F test of significance of the building main effect. A significant multivariate F was obtained for B, $F(4,36)=22.01, p.=.000$, indicating that the two sites have significantly different affective qualities. Univariate F tests were performed on each scale. The critical value of F was determined using the Bonferroni inequality. An alpha of .0125 was set for each univariate F test, preserving a familywise error rate of .05. Significant univariate F's were obtained for Scale 1 ($F(1,39)=53.98, p.=.000$), Scale 2 ($F(1,39)=50.57, p.=.000$) and, Scale 4 ($F(1,39)=16.71, p.=.000$). Scale 3 was not significant ($F(1,39)=.33, p.=.569$).

Discussion

The highly significant multivariate F and the three out of four significant univariate F's provided satisfactory evidence that the sites were significantly different as rated by the dependent measure. It was therefore decided to use the two sites in the Experiment.

Figure 1: Affective descriptors



Experiment

Method

Subjects. Participants were 408 University of British Columbia students enrolled in undergraduate psychology courses, who received course credit for their participation. Students completed preselection questionnaires in class. Those who had never been to one of the two buildings used in the study, or who had been in one of the buildings four or more times were contacted by telephone for participation in the study. Of those who completed the question on gender there were 177 males and 205 females in the study. The mean age of the participants was 20.2 years. The mean year of studies in which they were enrolled was 1.98.

Sites. The sites used in the Pilot Study were used as the stimulus sites in the Experiment.

Photographs. The same slides that were used in the Pilot Study were used in the Experiment, except for the following change: The origional slide of the Scarfe lobby showed very few students in the lobby. To reflect the number of students that typically occupied the lobby during the period in which actual site subjects were to be run, a new slide was taken from the same location, with about one half of the couches occupied by students. This increased the similarity between the color slide and actual site stimuli. During the actual site ratings roughly the same number of students were in the lobby.

The slides simulating the approach and entry to the stimulus sites were used so that participants in the actual site and slide conditions would receive roughly equivalent amounts of visual information about the buildings in which the ratings were to take place.

Roughly three quarters of the actual site participants were run during sunny weather, most of the remainder during overcast and partly sunny weather, and a very few in rainy weather.

Dependent measure. The same dependent measure used in the Pilot Study was used in the Experiment. For each copy of the dependent measure the items were arranged in random order.

Design. The study was a 2 x 3 x 2 factorial design. There were two stimulus sites. The three media were actual site, color slides, and brief written description. There were two levels of familiarity: Completely unfamiliar -- never before in the building and; familiar -- had been in the building four times or more.

Materials. Familiar and unfamiliar participants were recruited by administering a preselection questionnaire to undergraduate classes. The 2 buildings in the experiment were imbedded in a list of 13 other buildings so as not to alert the students to the buildings of interest in the study. Students could indicate on the questionnaire which buildings they had ever been in before, and how many times they had been there. The preselection questionnaire is reproduced in the Appendix.

Booklets were designed for each building and media condition -- a total of six 4-page booklets. Each booklet is reproduced in the Appendix. The content and format of each type of booklet was identical except for the following details: Building names, instructions for completing the experimental tasks specific to each medium, and the written descriptions of the stimulus sites.

The questionnaire on the last page of the booklet was included to obtain descriptive information about the participants, as well as to serve as a check on their familiar or unfamiliar status. In this manner the data from 10% to 15% of the people recruited to participate in the study were rejected because their unfamiliar status had

changed to familiar (but less than four visits) in the time between completing the preselection questionnaire and participating in the study.

The written descriptions were deliberately constructed to be information poor. Since it was possible that the slides would contain enough information that familiar participants would not have any advantage over unfamiliar participants when rating them, it was thought that a written description might be a more sensitive medium with which any compensating effects of familiarity could be demonstrated. They contained only enough information to enable participants familiar with the stimulus sites to locate themselves mentally in the exact place from which actual site and slide condition participants rated the stimulus sites.

The slide condition was run in a 6 m square room. A Kodak Ektographic Slide Projector Model AF-2 with an f:3.5 Kodak Projection Zoom Ektamar Lens was used to present the slides. The two-way projection lamp was adjusted to the brightest setting. The projector was set 1 m off the floor and 5.9 m from one of the flat white walls which served as the projection surface. The horizontal images measured 1.64 m by 1.09 m. The vertical images measured 1.09 m by 1.64 m.

Participants were seated in groups of 1 to 8 persons at four 2-person tables located 2.6 m and 3.9 m from the projection surface, and from .6 m to 1.5 m to each side of the projector.

Lighting by which the participants could work was provided by a 100 watt floor lamp. The experimenter sat next to the projector. A stopwatch was used to time the presentation duration of the slides.

Procedure. All of the participants were run during a six week period during the fall semester. Groups in all conditions were usually comprised of a mixture of familiar and unfamiliar participants. Participants were run in groups ranging in size from 1 to 8, depending upon their availibility.

It took approximately 15 min to run each group of actual site subjects. They were met at a prearranged time in the lobby of a building in the vicinity of the stimulus buildings, and led to a spot near the building to which they had been assigned. They were given booklets, and were asked to read the first page.

Participants in the Kenny building condition were then led through a ground floor entrance and up a stairwell to the third floor landing. They could not see the atrium until they reached the third floor landing. They were then instructed to read the second page of their booklets. They viewed the atrium for an additional 45 sec, after which they were instructed to complete the dependent measure and the questionnaire. Afterward, all participants were given debriefing forms and course credit slips.

Participants in the Scarfe building condition were then led inside and directly to a point from where they could see the lobby. They were run in the same manner as the participants in the Kenny condition.

Participants in the slide condition were led into the room and were seated. They were given booklets and asked to read the first page. Then they were shown the slides simulating the approach and entry into the building for their experimental condition. Each slide was shown for the same amount of time that the actual site participants took to walk from the place where one approach slide was taken to where the next slide was taken. When the slide of the stimulus site was projected, participants were asked to read the second page of their booklets. They then viewed the slide for an additional 45 sec before completing pages three and four, during which time the slide of the stimulus site was still shown. Afterward they were given debriefing forms and course credit slips. It took 10 to 15 min to run each group.

Preselection questionnaires were not used to select participants for the written condition. A mixture of booklets for both buildings were given to a total of three undergraduate psychology classes, ranging in size from about 30 to 80 students. The

students were asked to read the first page (those who did not wish to participate were excused). They were then instructed to read the written description on the second page. They were given 45 sec to visualize the scene that was described, after which they completed the dependent measure and the questionnaire. They were then debriefed and recieved course credit slips.

Considerably more booklets were administered than were required for the written conditions. After they had been completed the booklets were examined. Those that met the requirements for familiarity or unfamiliarity with the stimulus sites were retained.

Results

Before performing the MANOVA, the 40 scale items were combined to form 4 bipolar scales of affective quality in the same manner as in the Pilot Study.

A three-way MANOVA was performed on a 2 x 3 x 2 factorial design with two levels of building **B** (Kenny, Scarfe), three levels of media **M** (actual site 'a', slide 's', written 'w') and two levels of familiarity **F** (familiar, unfamiliar). With large and equal cell n's (n=34), the sampling distribution of the means was assumed to be normal. Additionally, the variance-covariance matrices were assumed to be homogeneous. If the assumption is valid, the significance tests are robust (Tabachnick & Fidell, 1983, p.233).

The conscientiousness and accuracy of subjects' ratings were checked using a procedure outlined in Russell and Pratt (1980). Two regression equations were used to calculate predicted exciting-gloomy and distressing-relaxing scores, based on obtained arousing-sleepy and pleasant-unpleasant scores. Then the predicted and actual cell means on Scales 2 and 4 were compared.

Correlations between the actual and predicted scores were calculated, and compared with correlations between actual and predicted scores in Russell and Pratt (1980). Theoretically, the predicted and actual scores should correlate 1.0. There is no

firm criterion for determining when the correlations are acceptably close to the theoretical values, but they closely approximate those reported in Russell and Pratt. These values are reported in Table I.

Place Table I about here (See p. 35)

Another check on the reliability of subjects' ratings was obtained by producing a correlation matrix of the ratings on the four scales, and comparing these values with those reported in Russell and Pratt (1980) from their actual data, and with the theoretical correlations among the scales. Some of the correlations are closer to the theoretical values than those obtained by Russell and Pratt, and some are not, but on average the values are within the range expected from actual data (Russell, J., personal communication, February, 1986). These values are reported in Table II.

Place Table II about here (See p. 36)

The Wilks Lambda was used for the multivariate F tests of significance of the main effects and interactions. Significant multivariate F's were obtained for **M** x **F** x **B**, $\underline{F}(8,786)=2.97$, $\underline{p}.=.003$; **F** x **B**, $\underline{F}(4,393)=2.63$, $\underline{p}.=.034$; **M** x **B**, $\underline{F}(8,786)=6.17$, $\underline{p}.=.000$ and; **M** x **F**, $\underline{F}(8,786)=2.13$, $\underline{p}.=.031$. Significant multivariate F's were also obtained for **B**, $\underline{F}(4,393)=32.11$, $\underline{p}.=.000$ and; **M**, $\underline{F}(8,786)=5.85$, $\underline{p}.=.000$. **F** was not significant, $\underline{F}(4,893)=.47$, $\underline{p}.=.76$. The significant multivariate F's show that all three independent variables, with the exception of **F** by itself, have a strong influence on how affect is rated. To determine more precisely the effects of the experimental manipulations on the dimensions of affect represented by the four scales, the MANOVA was followed up by multiple ANOVAs.

For each significant multivariate main effect and interaction, univariate F tests were performed on each scale. The critical value of F was determined using the Bonferroni inequality. An alpha of .0125 was set for each univariate F test, thus preserving a familywise error rate of .05. The results of the univariate F tests and the MS_e associated with each scale, are reproduced in Table III.

Place Table III about here(See p. 37)

The Neuman-Keuls (NK) multiple comparison (MC) procedure, using a per-contrast Type I error rate of .0125, was used to perform pairwise comparisons among cell means for scales for which significant univariate F's for **M** were obtained. The Dunn MC method (Glass and Hopkins, 1984), with a familywise error rate set at .01, was used to make C planned contrasts among specified cell means in the scales for which significant univariate interactions were obtained.

Scale 1

For the **M** x **B** interaction nine planned pairwise comparisons were made, with the critical t value, $.99_t.396,9=3.26$. The only significant difference between media for Kenny was between the slide and written conditions (a=s, a=w, s≠w). There were no significant differences between the media for Scarfe. The last three pairwise comparisons were of Kenny and Scarfe at each medium. The sites were rated as significantly different in the actual site and slide conditions, and as the same in the written description condition.

For the **M** x **B** x **F** interaction, 12 planned pairwise comparisons were made, with the critical t value, $.99_t396,12=3.289$. Only the slide and written description conditions were rated as different for Kenny familiar (a=s≠w, a=w), and for Kenny unfamiliar the actual site was rated as different from the slide condition (a≠s=w,

a=w). For both the familiar and unfamiliar Scarfe conditions all three media were given the same ratings. The 12 cell means for Scale 1 are reproduced in Figure 2.

Place Figure 2 about here (See p. 38)

Scale 2

There were no significant differences among the three media for the **M** main effect (q_1=4.03<.9875^{q}396,3=4.07). Examining the **M** x **B** interaction, the critical t value, .99$_t$396,9=3.26. For Kenny the actual site and slide conditions were rated as the same, and both were rated as different from the written condition (a=s≠w, a≠w). The media for Scarfe were all rated as the same. Kenny and Scarfe were given significantly different ratings in the actual site and slide conditions, and were rated as the same in the written description condition. The 12 cell means for Scale 2 are reproduced in Figure 3.

Place Figure 3 about here(See p. 39)

Scale 3

Comparison of the media for the **M** main effect indicated no significant difference between the ratings of the actual site and slide conditions (q_2=2.61<.9875^{q}396,2=3.59). The actual site condition was rated as significantly different from the written description condition (q_1=6.61>.9875^{q}396,3=4.07), as was the slide condition (q_3=4.0>.9875^{q}396,2=3.59).

Two planned pairwise comparisons were made for the **F** x **B** interaction, with the critical t value, .99$_t$396,2=2.806. For both the Kenny and Scarfe interiors the ratings

by the familiar and unfamiliar subjects were not significantly different. The 12 cell means for Scale 3 are reproduced in Figure 4.

Place Figure 4 about here (See p. 40)

Scale 4

The media for **M** were compared. The actual site and written description conditions were rated as different (q_1=7.14>.9875^{q}396,3=4.07). The slide and written description conditions were not rated as different (q_2=3.45<.9875^{q}396,2=3.59). The actual site and slide conditions were rated as different (q_3=3.68>.9875^{q}396,2=3.59).

Nine planned pairwise comparisons were made for the **M** x **F** interaction with the critical t value, .99$_t$396,9=3.26. For the familiar condition the three media were rated as the same. In the unfamiliar condition the actual site and slide conditions were rated as the same, and both were rated as different from the written condition (a=s≠w, a≠w). The ratings of the familiar and unfamiliar groups at all three media were the same.

For the **M** x **B** interaction, the critical t value, .99$_t$396,9=3.26. There were no significant differences among media for Kenny. For Scarfe the actual site and slide conditions were rated as the same, and both were rated as different from the written condition (a=s≠w, a≠w). In a comparison of the two sites at each medium the ratings of Kenny and Scarfe were significantly different in only the written description condition. The 12 cell means for Scale 4 are reproduced in Figure 5.

Place Figure 5 about here (See p. 41)

Table I

Correlations among actual and predicted scale scores

Present study

	Scale (actual)	
Scale (predicted)	2	4
2	.84	
4		.80

Russell and Pratt (1980)

	Scale (actual)	
Scale (predicted)	2	4
2	.84	
4		.87

Table II

Correlations among scales

	Scale		
Present study			
Scale	2	3	4
1	.65	.21	.5
2		.66	.05
3			-.5
Russell and Pratt (1980)			
Scale	2	3	4
1	.47	-.08	.62
2		.65	-.14
3			-.65
Theoretical correlations[a]			
Scale	2	3	4
1	.71	.0	.71
2		.71	.0
3			-.71

[a]Obtained from Russell and Pratt (1980)

Table III

Probability values for each univariate F test

		Scale			
Effect	d.f	1	2	3	4
M x F x B	2,396	.008*	.052	.343	.059
F x B	1,396	.565	.313	.002*	.033
M x B	2,396	.000*	.000*	.215	.001*
M x F	2,396	.031	.963	.323	.003*
M	2,396	.034	.008*	.000*	.000*
B	1,396	.000*	.000*	.000*	.025
MS_e		108.3	132.9	111.8	117.9

*significant at p.<.0125

Figure 2: Cell means for Scale 1 (arousing-sleepy)

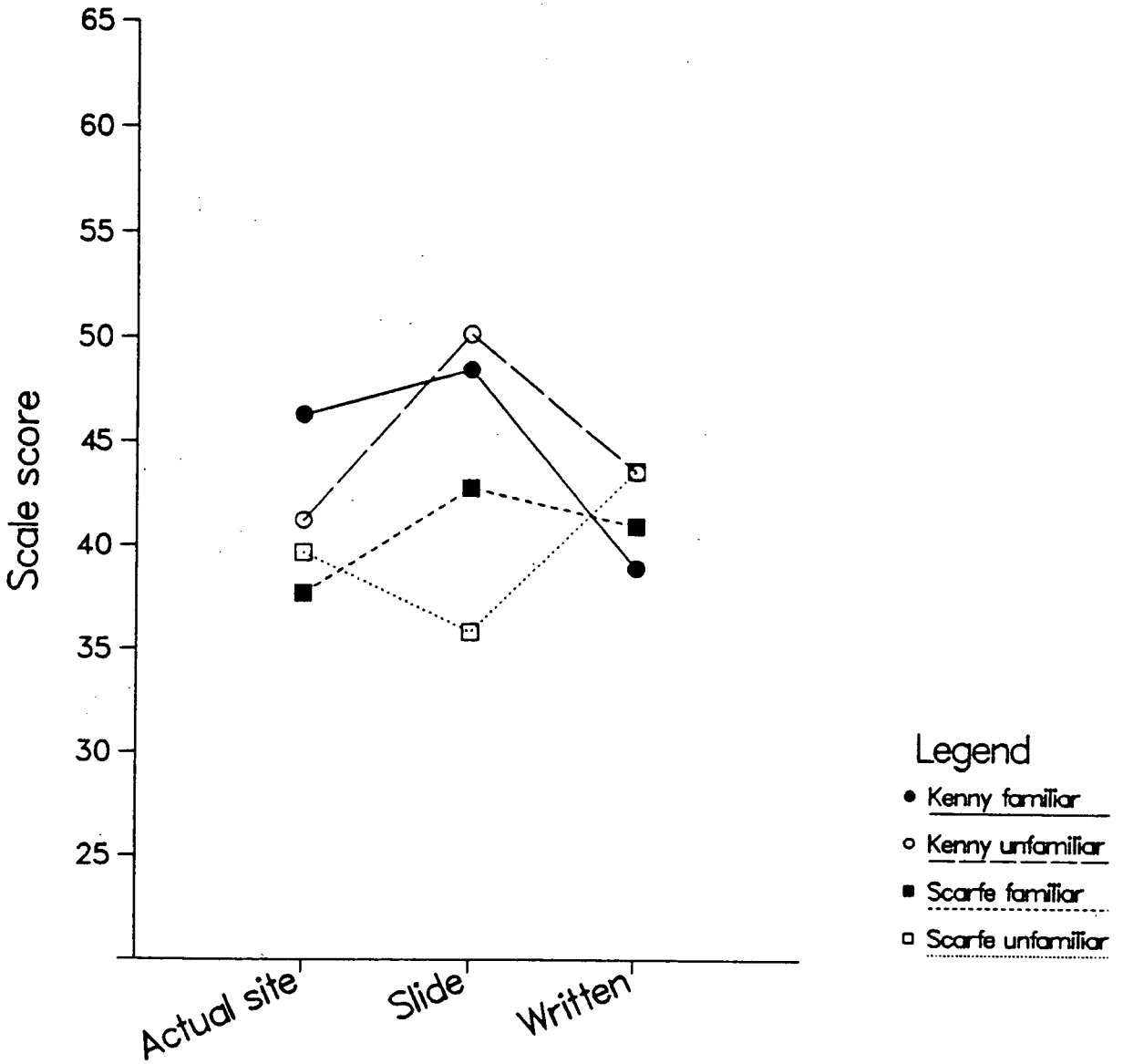


Figure 3: Cell means for Scale 2 (exciting-gloomy)

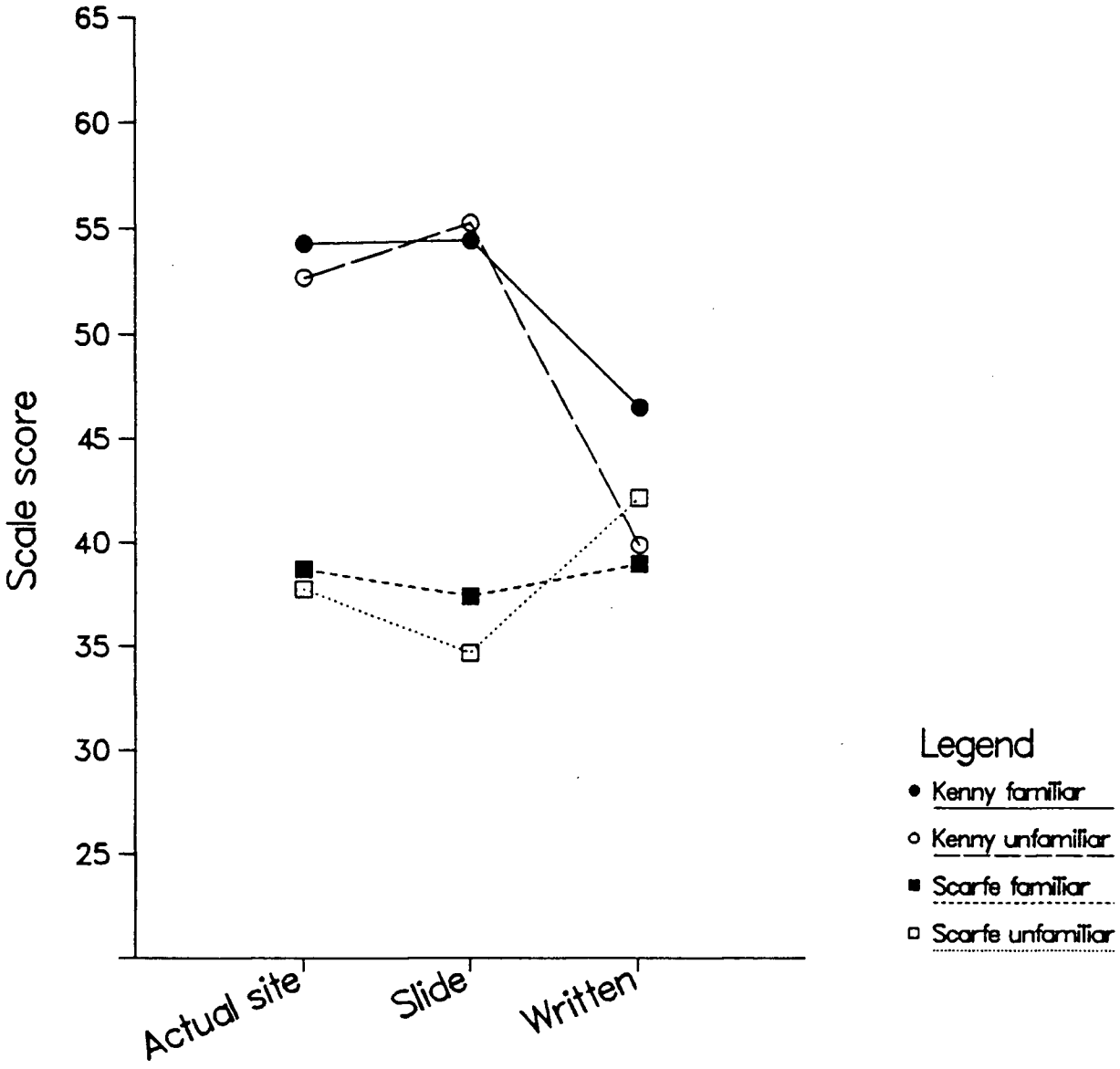


Figure 4: Cell means for Scale 3 (pleasant-unpleasant)

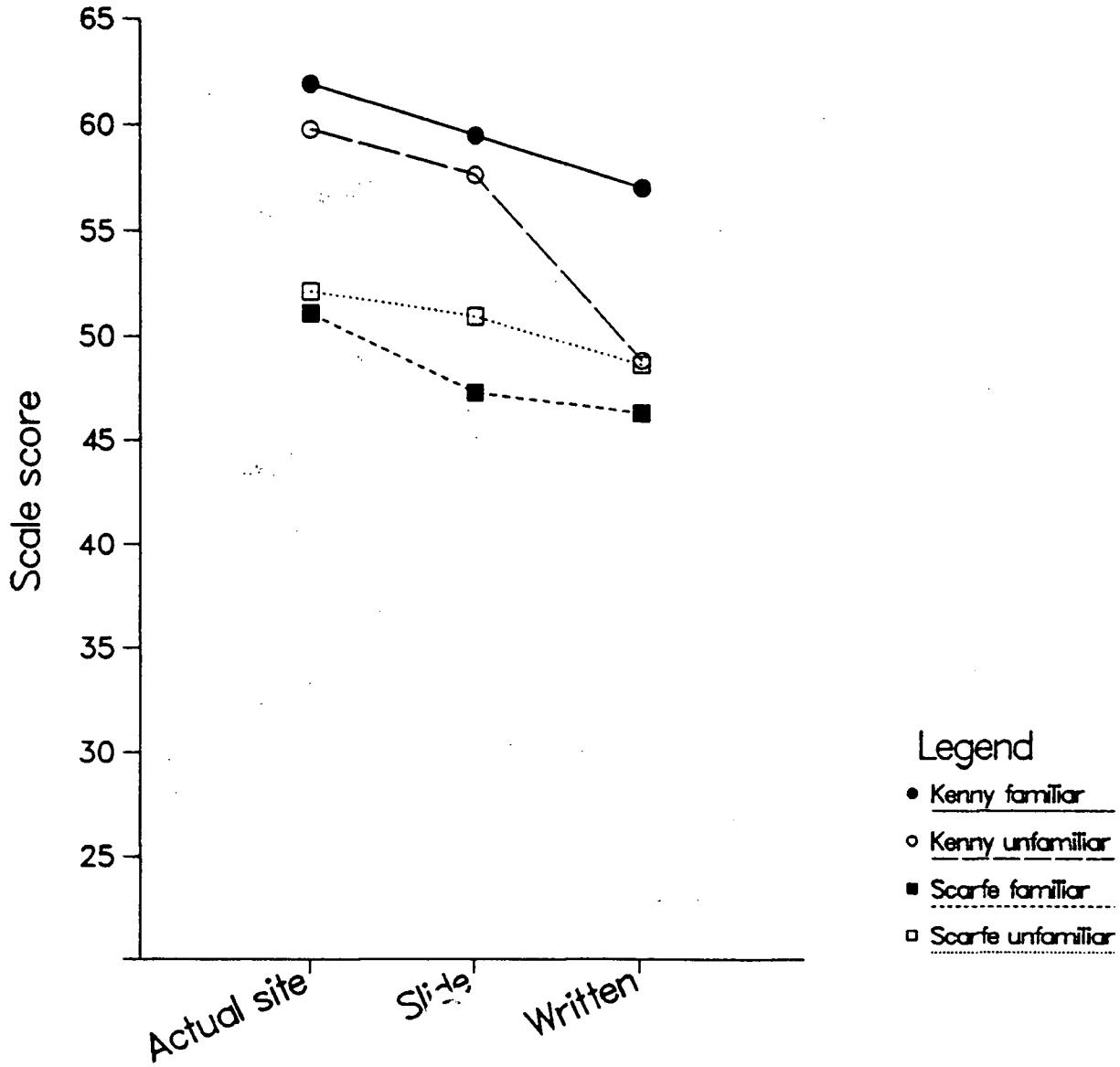
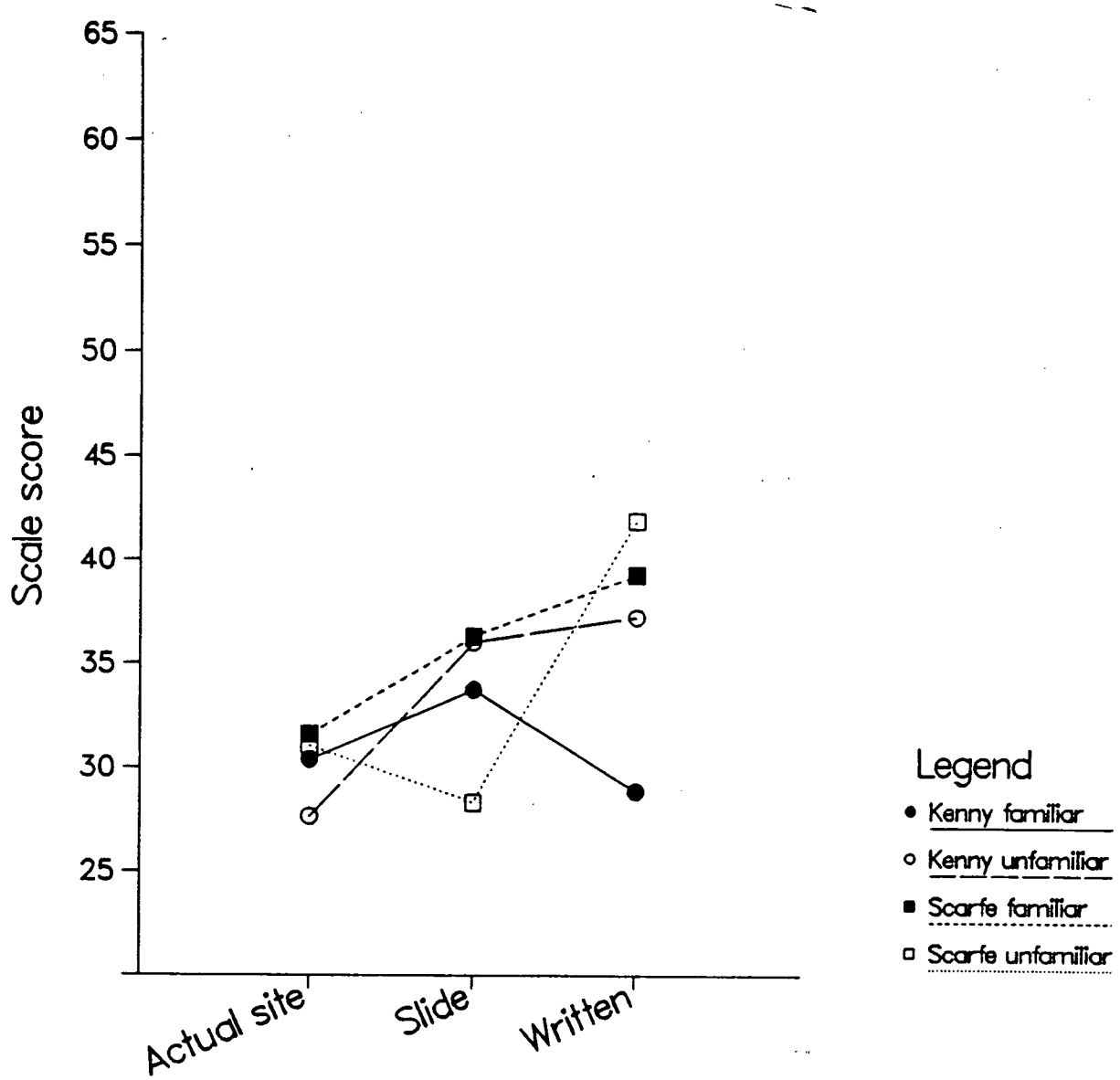


Figure 5: Cell means for Scale 4 (distressing-relaxing)



Discussion

The present study sought to answer two main questions. Would people rate color slides of two building interiors in the same way as they rate the actual interiors, and would those familiar with the stimulus sites rate slides of the building interiors more similarly to the actual site than those unfamiliar with the stimulus sites?

It was speculated that color slides might contain sufficient information about the actual sites to enable both familiar and unfamiliar judges to rate them in the same way -- a sort of information floor effect. Therefore, an additional medium was included in the study design that might more sensitively discriminate familiar from unfamiliar raters. A written description with only enough information to enable familiar subjects to mentally go to the same location from where the raters in the other conditions saw the sites, was added as a third medium. Otherwise the descriptions for each building interior were identical and extremly information poor.

Such a description also provided the opportunity to address the charge levied by Daniel and Ittelson (1981) that verbal scaling techniques or semantic differentials "... do not differentiate between actual environments, color slides, verbal discriptions, or even simple verbal labels when applied to diverse environments" (p. 156).

The results clearly show that determining the validity or non-validity of the simulation technique used in this study is not a cut and dried matter, a problem that has arisen in other validity studies. The differences among the scales in the evidence produced for or against validity, and the powerful main effects and interactions indicated by the many significant multivariate and univariate effects, make the task of drawing conclusions quite complex.

Reliance on the media factor alone yields varied results -- only on Scale 3 does the expected pattern of the actual site and slide conditions being rated as the same, and both being rated as different from the written description ($a=s\neq w$, $a\neq w$), occur.

The pairwise contrasts in **M x B** show that there is a considerable difference in how the media for each building interior were rated. On Scales 1 and 2, the pairwise comparisons among media for Scarfe show no significant differences between actual site, slide, and written, but there are differences between Scales 1 and 2 in how similarly or different the media for Kenny were rated. On Scale 4 the pattern of no differences between media for Scarfe, and differences among media for Kenny, is reversed: No significant differences were found between the three media for Kenny, but for Scarfe the actual site was rated in the same way as the slide condition, and the written description was rated differently from either of the other two conditions.

As could be expected from the contribution that the three independent variables made to the interactions, the greatest amount of information is revealed in examining the cell means for **M x B x F** (See Figures 2 through 5).

Only on Scale 1 do the three factors each contribute sufficiently to produce a significant three-way interaction. The Dunn MC method show no significant differences among media for either the familiar or unfamiliar Scarfe conditions. For Kenny the results are ambiguous and conform to the anticipated outcome for neither the familiar nor the unfamiliar raters.

Although the factors do not all contribute sufficiently to yield significant three-way interactions for Scales 2, 3, and 4, pairwise comparisons of the cell means, again with the critical t value, $.99_t 396,12=3.289$, reveal an important pattern. Making the same pairwise comparisons as for Scale 1, it was found that differences in cell means were associated only with the unfamiliar conditions: For Kenny unfamiliar on Scales 2 and 3, the actual site and slide conditions were rated as the same, and both were rated as different from the written description condition. On Scale 4 the actual site was rated as different from the written description condition. For Scarfe unfamiliar at Scale 4, the pattern of ratings was the same as was found with Kenny unfamiliar on

Scales 1 and 2. This pattern is supportive of the validity of using simulations, and the absence of differences among cell means across media for the familiar conditions strongly suggests that familiarity has a compensating effect. The effect becomes noticable when rating an information-poor medium. It appears that familiarity with stimulus sites enables subjects to rate the written descriptions in the same way as the slides and the actual sites.

That unfamiliar subjects are rating the actual site as the same as the slide, but rating the written description as different from either the actual site or the slide, provides two types of evidence for the validity of using the slide simulation. Convergent results were found between the two media that were expected to have been rated similarly -- the actual site and slide -- demonstrating convergent validity. Divergent results were obtained between the actual site and slide media, and the written condition ratings, where different ratings had been expected, thus providing the complimentary evidence of divergent validity.

In answer to the research question that seemingly valid responses might be spurious, and attributable to compensating effects of being familiar with stimulus sites, ratings of the unfamiliar subjects demonstrate that the slides -- of these two building interiors -- contain sufficient information to enable a majority of valid ratings.

There are clearly differences between Kenny and Scarfe with respect to the effects of media and familiarity. There is something quite different in the way that Scarfe was rated, in that there were fewer significantly different cell means between the actual site and written description conditions or the slide and written description conditions than there were for Kenny.

One way to account for this is by referring to what might be regarded as the prototypical or average building interior on the UBC campus in 1985. In relation to other UBC building interiors, the Scarfe lobby is very average, with nothing that

makes it architecturally distinctive. The atrium in Kenny, on the other hand, is distinctly unusual compared to other UBC building interiors, and the extensive use of glass, concrete, and intensely blue piping create a striking visual impression.

If, when required to rate a written description that in fact describes next to nothing, unfamiliar judges visualize and rate a typical building interior, the ratings would be similar to those obtained from familiar or unfamiliar judges rating typical slides or sites. If this is what happened when the Scarfe media were rated, it would account for why the ratings of the unfamiliar judges resulted in all three media receiving the same ratings. With no prior knowledge of what the atrium in Kenny looked like, and considering that the atrium is quite unusual compared to other UBC buildings, it seems very unlikely that an unfamiliar judge could make ratings similar to those made of the Kenny slides or actual site.

In addition to the pairwise comparisons of cell means across media, a visual inspection of **M** x **B** x **F** figures support the above conjecture in two places: The ratings of the written descriptions by unfamiliar Scarfe and Kenny subjects are almost identical on Scales 1 and 3, and are very similar for Scales 2 and 4. There is also a pattern consistent across all four scales of a greater difference between the familiar and unfamiliar ratings of the written description of Kenny than of the familiar and unfamiliar ratings for the written description for Scarfe. The pattern of statistically equal cell means across media for unfamiliar Scarfe subjects lends support to a prototypicality hypothesis.

The performance of verbal rating techniques has raised concern among some researchers (e.g., Daniel and Ittelson, 1981; Danford and Willems, 1975). The results of the present study do not suggest that the dependent measure constrained the rater responses. The powerful main effects for **B** are conclusive evidence that on both the multivariate and univariate levels, in this study at least, the dependent measure

differentiates between actual environments. The multiple comparisons in the **B x M** interaction demonstrates that the dependent measure does differentiate between different environments represented by different color slides. Contrary to the claim of Danford and Willems (1975), the results do not indicate that the responses were constrained by the dependent measure.

Danford and Willems (1975) believe that both convergent and divergent responses must occur in order to demonstrate that the dependent measure is working. That the rating scales demonstrated convergent and divergent responses in the present study can easily be seen by examining the pairwise comparisons between building cell means in the **M x B** interaction. On Scales 1 and 2, the ratings for Kenny and Scarfe are significantly different for the actual site and slide conditions, but are not significantly different for the written conditions. The same is true for Scale 3, although the **M x B** interaction is not significant. These results clearly show that Scales 1 to 3 can differentiate between actual building interiors, and slide simulations of those interiors. They do not differentiate between the nearly identical written descriptions for unfamiliar subjects -- another example of convergent validity.

On the distressing-relaxing scale (Scale 4), Kenny and Scarfe are rated the same on the actual site and slide media, but are rated as significantly different on the written description. That the two sites obtain similar ratings on this scale could be expected, but it is not clear what features of the written descriptions could have been responsible for the different ratings between Kenny and Scarfe in that medium.

On the whole, the pairwise contrasts among media cell means in **M x B x F** indicate that for the Kenny atrium, the slide simulation was quite adequate for producing valid ratings. The Scarfe ratings also suggest validity, but these findings must be interpreted with caution in light of the possibility that the ratings reflect prototypicality and not validity.

Subject familiarity effects were indicated, but they were not of consequence for rating the high quality slide simulations. The effects of familiarity became very evident when the viewers rated the information poor written descriptions.

In view of the effects of subject familiarity measured in this study, it is recommended that familiar subjects not be used in simulation validity studies. The compensating effects appear to be great enough to severely impair the ability of a study to determine whether or not valid ratings of actual sites are being obtained. Results of this study show that what look like valid ratings may be due solely to subject familiarity effects. Viewer familiarity would be even more detrimental to a validity study trying to assess poorer quality simulations, such as sketches. It is concluded that the most sensitive tests of simulation validity will be done using raters who are completely unfamiliar with the stimulus sites.

The results of this study also point to the choice of buildings as a factor in whether a simulation validity study will produce usable information. Further research is required on the role of prototypicality in ratings of simulations, and whether perhaps effort should be made to use only those stimulus sites that are not average.

It would be unfortunate if indices of simulation validity were due to prototypical buildings and not the good simulations, or, as in the case of Danford and Willems (1975) verbal rating techniques were dismissed as useless in a misattribution of cause to the rating scales instead of prototypical buildings. Another problem is that the more abstract a scale is -- that is, scales that measure less specific or concrete features or attributes -- the more of a problem prototypicality is likely to become (Ward, L., personal communication, April 4, 1986). This is because the more abstract the concept, the larger the group of things it will be applicable to.

Buildings should be rated in a pilot study so that it is known which of them are similar and different according to the rating scales used. This information can then be

compared at different media to determine whether the building ratings change as a function of the media by which they are represented.

In conclusion, color slides rated by the affective response instrument designed by Russell and Pratt (1980) produced valid ratings of the Kenny atrium. Because of the equivalent ratings of the Scarfe building interior on site, slide, and written description, and the suggestion that these ratings could be attributable to the prototypical nature of the interior, a conservative and cautious conclusion is that the validity of the use of the color slide for Scarfe was neither confirmed nor disconfirmed.

It also bears mentioning that I have encountered the same problem that was encountered in the other simulation validity studies; a lack of unequivocal evidence for the validity of photographic simulations. This study has, however, has produced evidence that weighs strongly in support of simulation validity. Ambiguity in validity study results can be expected to diminish as more becomes known about variables affecting viewer ratings.

In light of the above findings it seems prudent to design simulation validity studies so that convergent and divergent validity can be demonstrated, an approach that was recommended by Danford and Willems (1975).

As suggested previously, the complexity of simulation validity studies has been underestimated. This study indicates that viewer familiarity -- specifically the familiar-unfamiliar dichotomy -- and building prototypicality, require greater attention in future studies.

References

Brush, R. O., & Shafer, E. L. (1975). Application of a landscape preference model to land management. In E. H. Zube, R. O. Brush, & J. G. Fabos (Eds.), Landscape Assessment: Values, Perceptions and Resources. Pennsylvania: Dowden, Hutchinson & Ross, Inc.

Calvin, J. S., Dearing, J. A., & Curtin, M. E. (1972). An attempt at assessing preferences for natural landscapes. Environment and Behavior, 4, 447 - 470.

Carpman, J. R., Grant, M. A., & Simmons, D. A. (1985). Hospital design and wayfinding. A video simulation study. Environment and Behavior, 17, 296 - 314.

Chalmers, D. J., & Knight, R. G. (1985). The reliability of ratings of the familiarity of environmental stimuli: A generalizibility analysis. Environment and Behavior, 17, 223 - 238.

Coughlin, R. E., & Goldstein, K. A. (1970). The extent of agreement among observers on environmental attractiveness. Disc. Paper. 37. Philadelphia: Regional Science Research Institute.

Craik, K. H. (1972). Psychological factors in landscape appraisal. Environment and Behavior, 4, 260.

Danford, S., & Willems, E. P. (1975). Subjective responses to architectural displays. A question of validity. Environment and Behavior, 7, 486 - 516.

Daniel, T. C., & Boster, R. S. (1976). Measuring landscape esthetics: The Scenic Beauty Estimation method. Research Paper RM-167. USDA Rocky Mountain Forest and Range Experiment Station: Forest Service.

Daniel, T. C., & Ittelson, W. H. (1981). Conditions for environmental perception research: Comments on "The psychological representation of molar physical environments" by Ward and Russell. Journal of Experimental Psychology: General, 110, 153 - 168.

Dunn, M. C. (1976). Landscape with photographs: Testing the preference approach to landscape evaluation. Journal of Environmental Management, 4, 15 - 26.

Evans, G. W. (1980). Environmental cognition. Psychological Bulletin, 88, 259 - 287.

Evans, G. W., & Wood, K. W. (1980). Assessment of environmental aesthetics in scenic highway corridors. Environment and Behavior, 12, 255 - 273.

Glass, G. V., & Hopkins, K. D. (1984). Statistical Methods in Education and Psychology. Second Edition. New Jersey: Prentice-Hall, Inc.

Hershberger, R, G., & Cass, R. C. (1974). Predicting user responses to buildings. In D. H. Carson (Ed.), Man-Environment Interactions: Evaluations and Applications. Vol. 2. (pp. 117 - 134). Pennsylvania: Dowden, Hutchinson & Ross, Inc.

Herzog, T. R. (1985). A cognitive analysis of preference for waterscapes. Journal of Environmental Psychology, 5, 225 - 241.

Howard, R. B., Mlynarski, F. G., & Sauer, G. C. Jr. (1972). A comparative analysis of affective responses to real and represented environments. In W. J. Mitchell (Ed.), Proceedings of the Environmental Design: Research and Practice 3/ar 8 Conference. Vol. 1. (pp. 6-6-1 - 6-6-8). University of California, Los Angeles: Regents of the University of California.

Lau, J. J. H. (1969). Differences between full-size and scale-model rooms in the assessment of lighting quality. In D. V. Canter (Ed.), Architectural Psychology. (pp. 43 - 48). London: RIBA Publications.

Meckechnie, G. E. (1977). Simulation techniques in environmental psychology. In D. Stokols (Ed.), Perspectives on Environment and Behavior: Theory, Research and Applications. (pp. 169 - 189). New York: Plenum.

Moore, G. T. (1979). Knowing about environmental knowing: The current state of theory and research on environmental cognition. Environment and Behavior, 11, 33 - 71.

Nassauer, J. I. (1983). Framing the landscape in photographic simulation. Journal of Environmental Management, 17, 1 - 16.

Osgood, C. E., Suci, G., & Tannenbaum, P. H. (1957). The Measurement of Meaning. Urbana: University of Illinois Press.

Pomeroy, J. W., Green, M. B., & Fitzgibbon. (1983). Evaluation of urban riverscape aesthetics in the Canadian Praries. Journal of Environmental Management, 17, 263 - 276.

Rabinowitz, C. B., & Coughlin, R. E. (1971). Some experiments in quantative measurement of landscape quality. RSRI Discussion paper series: No. 43. Philadelphia: Regional Science Research Institute.

Rowe, R. D., & Chestnut, L. G. (Eds.). (1983). Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas. Colorado: Westview Press.

Russell, J. A., (1979). Affective space is bipolar. JPSP, 37, 345 - 356.

Russell, J. A. (1978). Evidence of convergent validity on the dimensions of affect. JPSP, 36, 1152 - 1168.

Russell, J. A., & Lanius, U. F. (1984). Adaptation level and the affective appraisal of environments. Journal of Environmental Psychology, 4, 000 - 000.

Russell, J. A., & Pratt, G. (1980). A description of the affective quality attributed to environments. JPSP, 38, 311 - 322.

Russell, J. A., Ward, L. M., & Pratt, G. (1981). Affective quality attributed to environments. A factor analytic study. Environment and Behavior, 13, 259 - 288.

Seaton, R. W., & Collins, J. B. (1972). Validity and reliability of simulated buildings. In W. J. Mitchell (Ed.), Proceedings of the Environmental Design: Research and Practice 3/ar 8 Conference. Vol 1. (pp. 6-10-1 - 6-10-12). University of California, Los Angeles: Reagents of the University of California.

Seung-Bin Im. (1984). Visual preferences in enclosed spaces: An exploration of a scientific approach to environmental design. Environment and Behavior, 16, 235 - 262.

Shafer, E. L. Jr., & Richards, T. A. (1974). A comparison of viewer reactions to outdoor scenes and photographs of those scenes. In D. Canter & T. Lee (Eds.), Psychology and the Built Environment. (pp. 71 - 79). England: Whitefriars Press Ltd.

Shelby, B., & Harris, R. (1985). Comparing methods for determining visitor evaluations of ecological impacts: Site visits, photographs, and written descriptions. Journal of Leisure Research, 17, 57 - 67.

Shuttleworth, S. (1980). The use of photographs as an environment presentation medium in landscape studies. Journal of Environmental Management, 11, (1), 61 - 76.

Tabachnick, B. G., & Fidell, L. S. (1983) Using Multivariate Statistics. New York: Harper & Row, Publishers.

Williams, S. (1985). How the familiarity of a landscape affects appreciation of it. Journal of Environmental Management, 21, (1), 63 - 67.

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

Zube, E. H., Pitt, D. G., & Anderson, T. W. (1975). Perception and prediction of scenic resource values of the northeast. In E. H. Zube, R. O. Brush, & J. G. Fabos (Eds.), Landscape Assessment: Values, Perceptions, and Resources. (pp. 151 - 167). Pennsylvania: Dowden, Hutchinson & Ross, Inc.

Appendix

Below are the names of 15 buildings on the UBC campus. I would like you to estimate how often you have ever been in each of these buildings. Indicate this with a number from 1 to 30. If the number of times you have been in a building listed is more than 30, indicate this with a +30.

Additionally, indicate with a checkmark how frequently you go to each of the buildings listed, choosing one of the following options: once a week or more; less than once a week; less than twice a month; hardly ever; never. If by chance you are not sure if you have been in one of the buildings listed (i.e., did not know its name when you visited it) you can express this with the option "don't know".

Building	0 to 30 or +30	once a week or more	less than once a week	less than twice a month	hardly ever	never	don't know
Museum of Anthropology							
Henry Angus							
Sedgewick							
Ponderosa							
Main Library							
Buchanan							
Memorial Gymnasium							
D. T. Kenny Psychology							
Graduate Student Centre							
Computer Science							
New Bookstore							
Van. School of Theology							
Freddy Wood Theatre							
Scarfe Education							
Home Economics							

I am recruiting volunteers to participate in a study which I will be conducting over the next few weeks. It involves descrimination of stimuli, and will take approximately 20 minutes. You may receive course credit for your participation. If you are interested in participating, complete the following section. You will be contacted by phone, and a suitable time for participation will be arranged.

NAME:
PHONE NUMBER(S):
COURSE:
SECTION:
COURSE PROFESSOR:

THANK YOU

The purpose of this study is to gather information on how people describe places. First you will approach and enter the D. T. Kenny Psychology building. You will go to a place inside the building where you will then be asked to rate how accurately a set of words describe the place that you will see. Then you will complete a short questionnaire.

I will assume that if you complete the following questionnaires, your voluntary consent to participate in this study has been granted.

Anyone who decides to participate in this study may at any time during the study withdraw. If you are from a class in which course credit is given for participating in experiments, you will still be entitled to credit even if you withdraw.

The study will take approximately 20 minutes of your time. Please do not turn the page in your booklet until requested by me.

Take a moment to get into the mood of the place that you now see. You will be asked to turn the page after a moment.

INSTRUCTIONS

Below is a list of words that can be used to describe places. We would like you to rate HOW ACCURATELY each word below describes this place. Use the following 1-8 rating scale for your answer. Please be sure that you have given as answer for EACH word.

IDENTIFICATION SEQUENCE 471

1	2	3	4	5
EXTREMLY INACCURATE	VERY INACCURATE	QUITE INACCURATE	SLIGHTLY INACCURATE	SLIGHTLY ACCURATE

6	7	8
QUITE ACCURATE	VERY ACCURATE	EXTREMLY ACCURATE

_ 01 active	__ 02 idle
_ 03 pleasing	___ 04 sensational
_ 05 tranquil	__ 06 pretty
_ 07 displeasing	__ 08 lazy
_ 09 nice	___ 10 dreary
_ 11 exhilarating	__ 12 stimulating
_ 13 boring	__ 14 calm
_ 15 dissatisfying	___ 16 slow
_ 17 alive	__ 18 exciting
_ 19 rushed	__ 20 frenzied
_ 21 dull	__ 22 monotonous
_ 23 uncomfortable	__ 24 repulsive
_ 25 interesting	___ 26 restful
_ 27 beautiful	__ 28 tense
_ 29 arousing	___ 30 unstimulating
_ 31 panicky	__ 32 hectic
_ 33 pleasant	___ 34 unpleasant
_ 35 inactive	__ 36 peaceful
_ 37 drowsy	__ 38 serene
_ 39 intense	__ 40 forceful

1) How old are you?_____

v) What is your area of study?_____

3) Are you a full-time student?_____Part-time?_____

4) Which year of studies are you in? Year 1 ( ) Year 2 ( ) Year 3 ( )
Year 4 ( ) Other____

5) Have you ever been in the D. T. Kenny Psychology building? YES ( ) NO ( )
If you have, estimate with a number from 1 to 30 how often you have been there._____More than 30 times?_____

6) If your answer to question 5 was YES, would you say that you visit the D. T. Kenny Psychology building at least:
   a) Once a week? ( )
   b) Less than once a week? ( )
   c) Less than twice a month? ( )
   d) Hardly ever? ( )
   If your answer to question 5 was NO, check not applicable. ( )

7) If you answered YES to question 5, put a checkmark beside the reasons for which you usually visit the D. T. Kenny Psychology building. If your answer to question 5 was no, check not applicable. ( )
   a) To see professors or graduate student teaching assistants. ( )
   b) To obtain test results. ( )
   c) To do homework. ( )
   d) To sign up for experiments. ( )
   e) To do research assistant or undergraduate marker duties. ( )
   f) Accompanying friends to the building. ( )
   g) To attend classes or lectures. ( )
   h) Other. ( ) Give an example. _____

8) Have you ever talked to friends about this building? YES ( ) NO ( )

9) Gender: Male ( ) Female ( )

10) Did you recognize the building from the outside during the experiment? YES ( ) NO ( )

The purpose of this study is to gather information on how people describe places. First you will approach and enter the Scarfe Education building. You will go to a place inside the building where you will then be asked to rate how accurately a set of words describe the place that you will see. Then you will complete a short questionnaire.

I will assume that if you complete the following questionnaires, your voluntary consent to participate in this study has been granted.

Anyone who decides to participate in this study may at any time during the study withdraw. If you are from a class in which course credit is given for participating in experiments, you will still be entitled to credit even if you withdraw.

The study will take approximately 20 minutes of your time. Please do not turn the page in your booklet until requested by me.

Take a moment to get into the mood of the place that you now see. You will be asked to turn the page after a moment.

INSTRUCTIONS

Below is a list of words that can be used to describe places. We would like you to rate HOW ACCURATELY each word below describes this place. Use the following 1-8 rating scale for your answer. Please be sure that you have given as answer for EACH word.

IDENTIFICATION SEQUENCE 471

1	2	3	4	5
EXTREMLY INACCURATE	VERY INACCURATE	QUITE INACCURATE	SLIGHTLY INACCURATE	SLIGHTLY ACCURATE

6	7	8
QUITE ACCURATE	VERY ACCURATE	EXTREMLY ACCURATE

_ 01 active	__ 02 idle
_ 03 pleasing	___ 04 sensational
_ 05 tranquil	__ 06 pretty
_ 07 displeasing	__ 08 lazy
_ 09 nice	___ 10 dreary
_ 11 exhilarating	__ 12 stimulating
_ 13 boring	__ 14 calm
_ 15 dissatisfying	___ 16 slow
_ 17 alive	__ 18 exciting
_ 19 rushed	__ 20 frenzied
_ 21 dull	__ 22 monotonous
_ 23 uncomfortable	__ 24 repulsive
_ 25 interesting	___ 26 restful
_ 27 beautiful	__ 28 tense
_ 29 arousing	___ 30 unstimulating
_ 31 panicky	__ 32 hectic
_ 33 pleasant	___ 34 unpleasant
_ 35 inactive	__ 36 peaceful
_ 37 drowsy	__ 38 serene
_ 39 intense	__ 40 forceful

1) How old are you?_____

2) What is your area of study?_____

3) Are you a full-time student?_____Part-time?_____

4) Which year of studies are you in? Year 1 ( ) Year 2 ( ) Year 3 ( )
Year 4 ( ) Other____

5) Have you ever been in the Scarfe Education building? YES ( ) NO ( )
If you have, estimate with a number from 1 to 30 how often you have been there._____More than 30 times?_____

6) If your answer to question 5 was YES, would you say that you visit the Scarfe Education building at least:
   a) Once a week? ( )
   b) Less than once a week? ( )
   c) Less than twice a month? ( )
   d) Hardly ever? ( )
   If your answer to question 5 was NO, check not applicable. ( )

7) If you answered YES to question 5, put a checkmark beside the reasons for which you usually visit the Scarfe Education building. If your answer to question 5 was no, check not applicable. ( )
   a) To see professors or graduate student teaching assistants. ( )
   b) To obtain test results. ( )
   c) To do homework. ( )
   d) To sign up for experiments. ( )
   e) To do research assistant or undergraduate marker duties. ( )
   f) Accompanying friends to the building. ( )
   g) To attend classes or lectures. ( )
   h) Other. ( ) Give an example. _____

8) Have you ever talked to friends about this building? YES ( ) NO ( )

9) Gender: Male ( ) Female ( )

10) Did you recognize the building from the outside during the experiment? YES ( ) NO ( )

The purpose of this study is to gather information on how people describe places. First you will see five slides that simulate the approach and entry into the D. T. Kenny Psychology building. You will then be asked to rate how accurately a set of words describe the place depicted in the last slide that you will see. Then you will complete a short questionnaire.

I will assume that if you complete the following questionnaires, your voluntary consent to participate in this study has been granted.

Anyone who decides to participate in this study may at any time during the study withdraw. If you are from a class in which course credit is given for participating in experiments, you will still be entitled to credit even if you withdraw.

The study will take approximately 20 minutes of your time. Please do not turn the page in your booklet until requested by me.

Take a moment to get into the mood of the place depicted in the slide that you now see. You will be asked to turn the page after a moment.

INSTRUCTIONS

Below is a list of words that can be used to describe places. We would like you to rate HOW ACCURATELY each word below describes this place. Use the following 1-8 rating scale for your answer. Please be sure that you have given as answer for EACH word.

IDENTIFICATION SEQUENCE 471

1	2	3	4	5
EXTREMLY INACCURATE	VERY INACCURATE	QUITE INACCURATE	SLIGHTLY INACCURATE	SLIGHTLY ACCURATE

6	7	8
QUITE ACCURATE	VERY ACCURATE	EXTREMLY ACCURATE

_ 01 active
_ 03 pleasing
_ 05 tranquil
_ 07 displeasing
_ 09 nice
_ 11 exhilarating
_ 13 boring
_ 15 dissatisfying
_ 17 alive
_ 19 rushed
_ 21 dull
_ 23 uncomfortable
_ 25 interesting
_ 27 beautiful
_ 29 arousing
_ 31 panicky
_ 33 pleasant
_ 35 inactive
_ 37 drowsy
_ 39 intense

__ 02 idle
___ 04 sensational
__ 06 pretty
__ 08 lazy
___ 10 dreary
__ 12 stimulating
__ 14 calm
___ 16 slow
__ 18 exciting
__ 20 frenzied
__ 22 monotonous
__ 24 repulsive
___ 26 restful
__ 28 tense
___ 30 unstimulating
__ 32 hectic
___ 34 unpleasant
__ 36 peaceful
__ 38 serene
__ 40 forceful

1) How old are you?_____

2) What is your area of study?_____

3) Are you a full-time student?_____Part-time?_____

4) Which year of studies are you in? Year 1 ( ) Year 2 ( ) Year 3 ( )
Year 4 ( ) Other____

5) Have you ever been in the D. T. Kenny Psychology building? YES ( ) NO ( )
If you have, estimate with a number from 1 to 30 how often you have been there._____More than 30 times?_____

6) If your answer to question 5 was YES, would you say that you visit the D. T. Kenny Psychology building at least:
   a) Once a week? ( )
   b) Less than once a week? ( )
   c) Less than twice a month? ( )
   d) Hardly ever? ( )
   If your answer to question 5 was NO, check not applicable. ( )

7) If you answered YES to question 5, put a checkmark beside the reasons for which you usually visit the D. T. Kenny Psychology building. If your answer to question 5 was no, check not applicable. ( )
   a) To see professors or graduate student teaching assistants. ( )
   b) To obtain test results. ( )
   c) To do homework. ( )
   d) To sign up for experiments. ( )
   e) To do research assistant or undergraduate marker duties. ( )
   f) Accompanying friends to the building. ( )
   g) To attend classes or lectures. ( )
   h) Other. ( ) Give an example. _____

8) Have you ever talked to friends about this building? YES ( ) NO ( )

9) Gender: Male ( ) Female ( )

10) Did you recognize the building from the outside during the experiment? YES ( ) NO ( )

The purpose of this study is to gather information on how people describe places. First you will see four slides that simulate the approach and entry into the Scarfe Education building. You will then be asked to rate how accurately a set of words describe the place depicted in the last slide that you will see. Then you will complete a short questionnaire.

I will assume that if you complete the following questionnaires, your voluntary consent to participate in this study has been granted.

Anyone who decides to participate in this study may at any time during the study withdraw. If you are from a class in which course credit is given for participating in experiments, you will still be entitled to credit even if you withdraw.

The study will take approximately 20 minutes of your time. Please do not turn the page in your booklet until requested by me.

Take a moment to get into the mood of the place depicted in the slide that you now see. You will be asked to turn the page after a moment.

INSTRUCTIONS

Below is a list of words that can be used to describe places. We would like you to rate HOW ACCURATELY each word below describes this place. Use the following 1-8 rating scale for your answer. Please be sure that you have given as answer for EACH word.

IDENTIFICATION SEQUENCE 471

1	2	3	4	5
EXTREMLY INACCURATE	VERY INACCURATE	QUITE INACCURATE	SLIGHTLY INACCURATE	SLIGHTLY ACCURATE

6	7	8
QUITE ACCURATE	VERY ACCURATE	EXTREMLY ACCURATE

_ 01 active	__ 02 idle
_ 03 pleasing	___ 04 sensational
_ 05 tranquil	__ 06 pretty
_ 07 displeasing	__ 08 lazy
_ 09 nice	___ 10 dreary
_ 11 exhilarating	__ 12 stimulating
_ 13 boring	__ 14 calm
_ 15 dissatisfying	___ 16 slow
_ 17 alive	__ 18 exciting
_ 19 rushed	__ 20 frenzied
_ 21 dull	__ 22 monotonous
_ 23 uncomfortable	__ 24 repulsive
_ 25 interesting	___ 26 restful
_ 27 beautiful	__ 28 tense
_ 29 arousing	___ 30 unstimulating
_ 31 panicky	__ 32 hectic
_ 33 pleasant	___ 34 unpleasant
_ 35 inactive	__ 36 peaceful
_ 37 drowsy	__ 38 serene
_ 39 intense	__ 40 forceful

1) How old are you?_____

2) What is your area of study?_____

3) Are you a full-time student?_____Part-time?_____

4) Which year of studies are you in? Year 1 ( ) Year 2 ( ) Year 3 ( )
Year 4 ( ) Other____

5) Have you ever been in the Scarfe Education building? YES ( ) NO ( )
If you have, estimate with a number from 1 to 30 how often you have been there._____More than 30 times?_____

6) If your answer to question 5 was YES, would you say that you visit the Scarfe Education building at least:
   a) Once a week? ( )
   b) Less than once a week? ( )
   c) Less than twice a month? ( )
   d) Hardly ever? ( )
   If your answer to question 5 was NO, check not applicable. ( )

7) If you answered YES to question 5, put a checkmark beside the reasons for which you usually visit the Scarfe Education building. If your answer to question 5 was no, check not applicable. ( )
   a) To see professors or graduate student teaching assistants. ( )
   b) To obtain test results. ( )
   c) To do homework. ( )
   d) To sign up for experiments. ( )
   e) To do research assistant or undergraduate marker duties. ( )
   f) Accompanying friends to the building. ( )
   g) To attend classes or lectures. ( )
   h) Other. ( ) Give an example. _____

8) Have you ever talked to friends about this building? YES ( ) NO ( )

9) Gender: Male ( ) Female ( )

10) Did you recognize the building from the outside during the experiment? YES ( ) NO ( )

The purpose of this study is to gather information on how people describe places. On the following page you will read a short description of a building. Then you will be asked to rate how accurately a set of words describe the place that you will read about. Then you will complete a short questionnaire.

I will assume that if you complete the following questionnaires, your voluntary consent to participate in this study has been granted.

Anyone who decides to participate in this study may at any time during the study withdraw. If you are from a class in which course credit is given for participating in experiments, you will still be entitled to credit even if you withdraw.

The study will take approximately 20 minutes of your time. Please do not turn the page in your booklet until requested by me.

Description. You approach and enter the D. T. Kenny Psychology building. You proceed directly to a point above the main entrance. From this point you can see out into the interior of the building.

Try to visualize the scene, taking a moment to get into the mood of the place. You will be asked to turn the page after a moment.

INSTRUCTIONS

Below is a list of words that can be used to describe places. We would like you to rate HOW ACCURATELY each word below describes this place. Use the following 1-8 rating scale for your answer. Please be sure that you have given as answer for EACH word.

IDENTIFICATION SEQUENCE 471

1	2	3	4	5
EXTREMLY INACCURATE	VERY INACCURATE	QUITE INACCURATE	SLIGHTLY INACCURATE	SLIGHTLY ACCURATE

6	7	8
QUITE ACCURATE	VERY ACCURATE	EXTREMLY ACCURATE

_ 01 active	__ 02 idle
_ 03 pleasing	___ 04 sensational
_ 05 tranquil	__ 06 pretty
_ 07 displeasing	__ 08 lazy
_ 09 nice	___ 10 dreary
_ 11 exhilarating	__ 12 stimulating
_ 13 boring	__ 14 calm
_ 15 dissatisfying	___ 16 slow
_ 17 alive	__ 18 exciting
_ 19 rushed	__ 20 frenzied
_ 21 dull	__ 22 monotonous
_ 23 uncomfortable	__ 24 repulsive
_ 25 interesting	___ 26 restful
_ 27 beautiful	__ 28 tense
_ 29 arousing	___ 30 unstimulating
_ 31 panicky	__ 32 hectic
_ 33 pleasant	___ 34 unpleasant
_ 35 inactive	__ 36 peaceful
_ 37 drowsy	__ 38 serene
_ 39 intense	__ 40 forceful

1) How old are you?_____

2) What is your area of study?_____

3) Are you a full-time student?_____Part-time?_____

4) Which year of studies are you in? Year 1 ( ) Year 2 ( ) Year 3 ( )
Year 4 ( ) Other____

5) Have you ever been in the D. T. Kenny Psychology building? YES ( ) NO ( )
If you have, estimate with a number from 1 to 30 how often you have been there._____More than 30 times?_____

6) If your answer to question 5 was YES, would you say that you visit the D. T. Kenny Psychology building at least:
   a) Once a week? ( )
   b) Less than once a week? ( )
   c) Less than twice a month? ( )
   d) Hardly ever? ( )
   If your answer to question 5 was NO, check not applicable. ( )

7) If you answered YES to question 5, put a checkmark beside the reasons for which you usually visit the D. T. Kenny Psychology building. If your answer to question 5 was no, check not applicable. ( )
   a) To see professors or graduate student teaching assistants. ( )
   b) To obtain test results. ( )
   c) To do homework. ( )
   d) To sign up for experiments. ( )
   e) To do research assistant or undergraduate marker duties. ( )
   f) Accompanying friends to the building. ( )
   g) To attend classes or lectures. ( )
   h) Other. ( ) Give an example. _____

8) Have you ever talked to friends about this building? YES ( ) NO ( )

9) Gender: Male ( ) Female ( )

10) Did you recognize the building from the outside during the experiment? YES ( ) NO ( )

The purpose of this study is to gather information on how people describe places. On the following page you will read a short description of a building. Then you will be asked to rate how accurately a set of words describe the place that you will read about. Then you will complete a short questionnaire.

I will assume that if you complete the following questionnaires, your voluntary consent to participate in this study has been granted.

Anyone who decides to participate in this study may at any time during the study withdraw. If you are from a class in which course credit is given for participating in experiments, you will still be entitled to credit even if you withdraw.

The study will take approximately 20 minutes of your time. Please do not turn the page in your booklet until requested by me.

Description. You approach and enter the Scarfe Education building through the main entrance. You proceed directly to the end of the lobby nearest you. From this point you can see out into the interior of the building.

Try to visualize the scene, taking a moment to get into the mood of the place. You will be asked to turn the page after a moment.

INSTRUCTIONS

Below is a list of words that can be used to describe places. We would like you to rate HOW ACCURATELY each word below describes this place. Use the following 1-8 rating scale for your answer. Please be sure that you have given as answer for EACH word.

IDENTIFICATION SEQUENCE 471

1	2	3	4	5
EXTREMLY INACCURATE	VERY INACCURATE	QUITE INACCURATE	SLIGHTLY INACCURATE	SLIGHTLY ACCURATE

6	7	8
QUITE ACCURATE	VERY ACCURATE	EXTREMLY ACCURATE

_ 01 active
__ 02 idle
_ 03 pleasing
___ 04 sensational
_ 05 tranquil
__ 06 pretty
_ 07 displeasing
__ 08 lazy
_ 09 nice
___ 10 dreary
_ 11 exhilarating
__ 12 stimulating
_ 13 boring
__ 14 calm
_ 15 dissatisfying
___ 16 slow
_ 17 alive
__ 18 exciting
_ 19 rushed
__ 20 frenzied
_ 21 dull
__ 22 monotonous
_ 23 uncomfortable
__ 24 repulsive
_ 25 interesting
___ 26 restful
_ 27 beautiful
__ 28 tense
_ 29 arousing
___ 30 unstimulating
_ 31 panicky
__ 32 hectic
_ 33 pleasant
___ 34 unpleasant
_ 35 inactive
__ 36 peaceful
_ 37 drowsy
__ 38 serene
_ 39 intense
__ 40 forceful

1) How old are you?_____

2) What is your area of study?_____

3) Are you a full-time student?_____Part-time?_____

4) Which year of studies are you in? Year 1 ( ) Year 2 ( ) Year 3 ( )
Year 4 ( ) Other____

5) Have you ever been in the Scarfe Education building? YES ( ) NO ( )
If you have, estimate with a number from 1 to 30 how often you have been there._____More than 30 times?_____

6) If your answer to question 5 was YES, would you say that you visit the Scarfe Education building at least:
   a) Once a week? ( )
   b) Less than once a week? ( )
   c) Less than twice a month? ( )
   d) Hardly ever? ( )

   If your answer to question 5 was NO, check not applicable. ( )

7) If you answered YES to question 5, put a checkmark beside the reasons for which you usually visit the Scarfe Education building. If your answer to question 5 was no, check not applicable. ( )
   a) To see professors or graduate student teaching assistants. ( )
   b) To obtain test results. ( )
   c) To do homework. ( )
   d) To sign up for experiments. ( )
   e) To do research assistant or undergraduate marker duties. ( )
   f) Accompanying friends to the building. ( )
   g) To attend classes or lectures. ( )
   h) Other. ( ) Give an example. _____

8) Have you ever talked to friends about this building? YES ( ) NO ( )

9) Gender: Male ( ) Female ( )

10) Did you recognize the building from the outside during the experiment? YES ( ) NO ( )