Deliverable # 6

Exploring Public Perceptions with Immersivity: an initial investigation of immersive displays

By: Paul Picard, Stephen Sheppard & Caitlin Akai

Project: Towards Best Practices for Public Processes in Sustainable Forest Management: Phase 2
R04-082

Submitted by:
Collaborative for Advanced Landscape Planning
CALP,
Forest Resources Management, UBC

March 31, 2004

Submitted To: Forestry Innovation Investment
Forestry Research Program

Please note: Some of this material is intended for future publication/technical review/practitioner testing and should not be made available publicly without first contacting Dr. S. Sheppard at CALP.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

significant effect is found between the two conditions, showing that context has a significant effect on Question 2 ratings (at least for the scale tested).21

1.0 BACKGROUND

New forms of immersive 3D visualization of realistic landscapes and other virtual reality displays promise a richer planning/decision-making environment and fuller communication of resource management issues. While everybody seems to like the idea of entering digital worlds, little evidence of or guidance on the benefits of landscape immersion have yet been passed on to practitioners. This paper reports on initial findings on the usefulness and validity of these immersive facilities relative to more conventional image display formats often used in planning processes and perceptual research. The exploratory study described here arises from early work at the Collaborative for Advanced Landscape Planning (CALP) at the University of British Columbia, Vancouver.

The objective of the study was to investigate the effect of immersivity (3 screens) on perceptions, as a first exploratory experiment. While perception experiments have been conducted for several decades using various photographic and visualisation images, there appears to be a very wide range of image sizes used, ranging from 4“ by 6” prints to slides projected onto large projection screens, with just about any possible size in between, and often no controls over where the observer sits in relation to the screen or image. This study investigates the effect of image scale or magnification, e.g. whether a bigger image yields different ratings than a smaller one.

At the same time, a single image of any size may not contain all the important contextual information about the place being represented, especially if that lateral (or peripheral) context frames the image in a particular setting or contains meaningful variations from the rest of the visible landscape (eg. buildings or cutblocks in an otherwise natural-appearing landscape). Therefore, the study described here also investigates the effect of context (what is immediately at the left and at the right of an image in the real-world scene) and image width as presented with large panoramic images.

Among potential variables that could influence respondents’ ratings are: the sheer size of the images presented; the presence/absence of peripheral vision (potentially affecting both the aspect ratio and the quantity/type of information given to the respondent); the setting itself, in this case a special immersion facility with 3 wrap-around screens (versus a flat projector screen) which may convey a hi-tech “wow” effect; and the viewing distance/angle creation for observers.

Underlying these potential variables to be tested are various theoretical constructs, two of which are central to the study presented here:

1) Image size matters: i.e. people’s responses to landscapes may change with image size. This is not necessarily a 3-screen issue, but the magnification of imagery relative to real-world image size projected on the

eye if the observer was actually standing in the landscape at the corresponding viewpoint. This translates to the solid (or in most cases horizontal) angle projected by an image on the retina, in either the real landscape or in an experimental setting with varying sizes of images. Facilities such as the Landscape Immersion Lab (LIL) maintained by CALP at UBC allow the researcher to recreate the image size and horizontal angle such that it is similar to what they would see in the real world, but only if the participant is sitting in the 'sweet spot'" at the right distance from the 3 screens. An underlying hypothesis for image validity is that using the correct horizontal angle leads to better or more reliable judgments/responses, more equivalent to these that would be received in the real world at that point. Bigger image size may provide an appropriate level of detail, and better sense of scale. Other important issues to control for are focal length and magnification in lenses. Possible implications of viewing smaller images include:

- Visual impacts are not noticeable, or seem less important,
- Scenic beauty is lower because the imagery is less vivid or intense.
- Less sense of being there, greater sense of detachment,
- Less emotional content - do they care as much for the place or environment?
- Information is lost and cognition reduced.

2) Image 'shape' matters: the horizontal field of view or image width may create effects, including the amount of important context visible. One hypothesis for image validity would suggest that the more 'real world' context the better, and that the viewers may lose information if they lose context. Here, we are only concerned with the horizontal angle of view, given that no stretching or distortion of images is allowed (the aspect ratio is maintained regardless of the image magnification/scale). Possible implications of providing a wide or panoramic view (relative to more conventional narrower images) include:

- Visual impact of projects within the landscape may vary, either being lowered because the dominance of the project is less relative to the larger landscape context, or that impact is increased because the wider context includes a project, which would be omitted form the narrower field of view
- Scenic Quality (Scenic Beauty) may be higher (due to a more interesting (novelty) effect
- Sense of being there increased
- More emotional content - greater concern for place or the environment due to greater engagement in the scene
- Cognition is increased due to more comprehensive information, although there is also the possibility of overkill through too much information

2.0 METHODS

2.1 Immersive facilities, experimental set-up and subject recruiting

2.1.1 IMMERSIVE FACILITY USED

The facility used to investigate immersitivity effects was the Landscape Immersion Lab (LIL) operated by CALP and located within the Forest Sciences Centre at the University of British Columbia. This facility is a semi-immersive environment, created by panoramic image projection onto three angled 7'x9' screens, using 3 LCD projectors. This set-up has the capacity to project a wide-angle image of various widths, although it is designed optimally for images that are approximately 120 degrees in width, approximating the majority of the human focal and peripheral field of view (given that the subject is located in the target area of the three screens). The advantage of the angled three-screen set-up is their ability to display "wrap-around" images that can approximate on a subject's retina the imprint that they would get in the real-world location represented in the image. The set-up also works as the equivalent of 3 standard 50mm camera photographs side by side. In addition, the three projectors are independent and have the ability to display images of various sizes and on any (or all) of the three screens, allowing for the creation of the various experimental conditions tested in the experiment (see below). Image 1 shows the LIL facility with a subject sitting in the optional target area (or "sweet spot").

Figure 1: Landscape Immersion Laboratory LIL facility with a subject sitting in the target area (or "sweet spot"), reproducing a 120-degree angle of view with a panoramic image.

The measurements for the location of the “target area” (where the chairs used by the subjects were located) are based on Sheppard (1989)[1], which provides clear guidelines on such experimental setups based on screen size and viewing distance to the screen, in order to attain image projections on each screen in such a way that the image edges form a 40 degree view field-of-view from the observer’s eyes (see Sheppard [1989] for more details). Given that the facility has a total of three adjacent (and angled) screens, this provides for a total of 120-degree field of view. Only 5 seats were used at a time, in order to ensure as much as possible that each subject is located as close as possible from the target area. The chairs were placed in a single row, parallel to the screen, located 12 feet 4 inches from the middle screen (based on Sheppard, 1989) and taped to the ground. This set-up was an extension of a 3-chair setup used in a previous experiment and shown in Figure 1 (without subjects). Each seat was also numbered, in order to be able to test whether the ratings were affected by the respondents seat number (e.g., whether the center seat, located exactly in the target area provides different ratings than the other seats located on either side of it).

[1] Sheppard, S. R. J. 1989. Visual simulation: A user’s guide for architects, engineers, and planners. New York, NY: Van Nostrand Reinhold.

Figure 2: Three-chair experimental set-up showing the essence of the experimental set-up used (5 chairs were used on a single row instead of 3 as shown here, with the extra chairs added at each end of the row shown here).

2.1.2 SUBJECT RECRUITING

Given that the scientific literature shows that students' ratings of landscape scenes are generally representative of those of the general population, the decision was made to recruit students from the UBC student community. Other advantages of recruiting students are that they are fairly readily available, and fairly inexpensive to recruit. Consequently, the experiment was advertised throughout the UBC campus, mainly via several student e-mail lists (securing a broad audience relative to the time/cost needed), and a Hotmail account was set up to receive/administer the potential subjects. The initial response was very good and the respondents were screened and booked into a time slot once they qualified for the experiment until enough subjects were obtained. A total of 125 subjects were recruited and paid $10 cash for their time.

Two categories of students were not allowed to participate in the experiment: landscape architecture students and forestry students. Landscape architecture students were not allowed to participate in the experiment because they have received (or are in the process of receiving) "specialized" training and education in landscapes, which includes training in how to "read" and "interpret" the landscape, which would affect their view of the landscape. Forestry students, they were excluded because they have received (or are receiving) a "specialized" training in harvest practices, and in landscape interpretations as well.

2.2 Experimental conditions tested

A total of four different experimental conditions were tested in an attempt to capture possible differences between size and or context/image width.

Condition A was the small, conventionally narrow image similar in aspect ratio to a standard photographic print or report page in landscape format. It included only the middle portion (one third) of the full 120 degree panoramic image, which was then reduced in scale in order to have its lateral edges form a horizontal angle of view of 11 degrees from the subject's perspective. The field of view contained in the image would be close to that in a 50mm lens for a 35 mm camera, i.e. approximately 40 degrees. At this size and viewing distance, Condition A approximates the image magnification obtained by viewing a small graphic in a report, watching TV, or sitting far from the projector screen at a public meeting.

Condition B also included only the middle portion (one third of the original 120 degree panoramic image) but it was left at the original scale, so it filled the central screen and its lateral edges formed an angle of view of 40 degrees from the subject's perspective at the 'sweet-spot'. This approximates a life-size image, as it would be seen on the retina by an observer standing at the real site and looking in the same direction, and the field of view of a conventional 50mm camera lens, but it fails to capture peripheral vision or adjacent landscape content. These same viewing angles (or image magnification) could be similar to a front seat close to a big screen in a public meeting, or a computer user typing right in front of a conventionally sized monitor.

Condition C included the entire panoramic image (field of view 120 degrees) but was reduced in scale in order to have its lateral edges form an angle of view of 40 degrees from the subject's perspective at the 'sweet spot', thus filling the width but not the height of the middle screen. Its image scale matches that of Condition A, but its width of view is three times larger.

Condition D included the entire panoramic image and was left at the original scale, so its lateral edges formed an angle of view of 120 degrees from the subject's perspective at the 'sweet spot' (and spanned across all three screens). This matches the actual view scale and much of the context/peripheral vision that an observer would receive at the actual site.

Figures 3 to 6 illustrate (in a generic manner) each condition used for the experiment. In essence, these conditions match each other in pairs either in size (scale) or in content: A and B are identical in content/field of view but differ in size, as do conditions C and D. Conditions A and C are identical in size (scale) but differ in content, as do conditions B and D. Conditions A and D represent the extremes of the display formats, from small and narrow to large and wide.

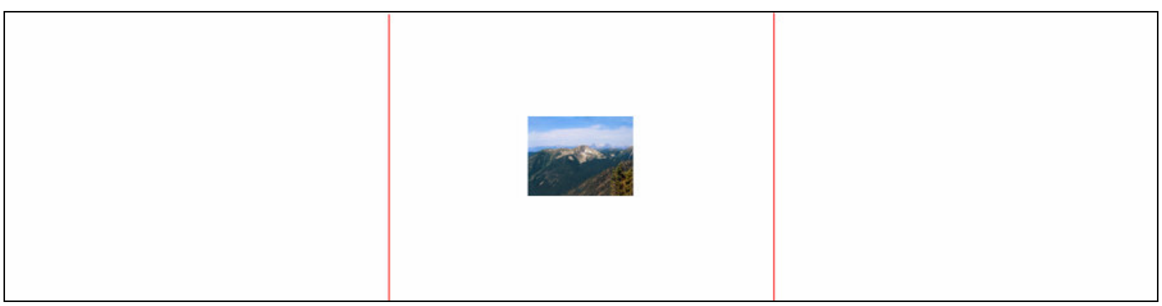

Figure 3: Condition A showing only the middle portion (one third of the original 120 degree panoramic image) on the middle screen and reduced in scale to have its lateral edges form an angle of view of 11 degrees from the subject's perspective. The red vertical lines indicate where the middle screen joins with the lateral screens.

Figure 4: Condition B showing only the middle portion (one third of the original 120 degree panoramic image) on the middle screen and left at the original scale to have its lateral edges form an angle of view of 40 degrees from the subject's perspective. The red vertical lines indicate where the middle screen joins with the lateral screens.

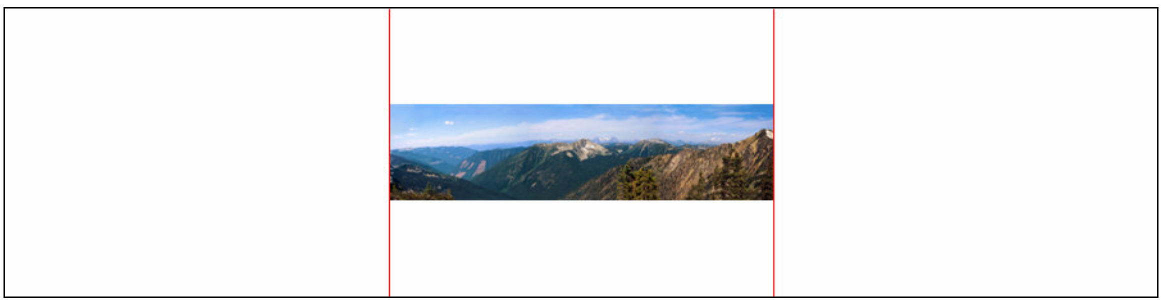

Figure 5: Condition C showing the entire panoramic image and reduced in scale in order to have its lateral edges form an angle of view of 40 degrees from the subject's perspective and shown on the middle screen. The red vertical lines indicate where the middle screen joins with the lateral screens.

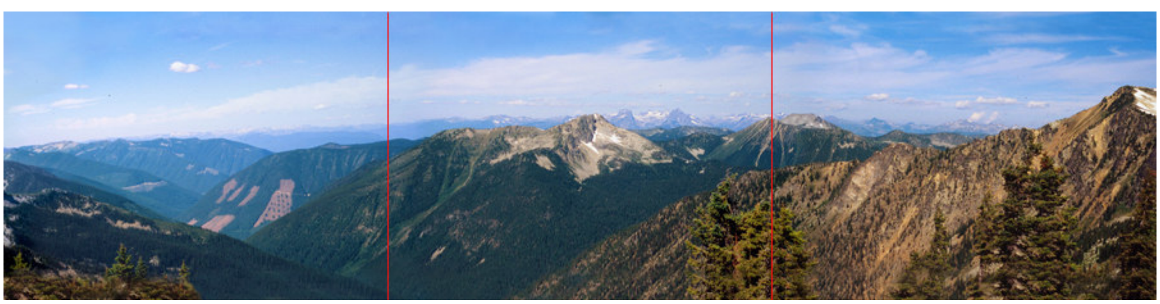

Figure 6: Condition D showing the entire panoramic image and left at the original scale to have its lateral edges form an angle of view of 120 degrees from the subject's perspective. The red vertical lines indicate where the middle screen joins with the lateral screens.

2.3 Image selection

The selection process used to assemble the images used in this experiment is described here. It should be kept in mind that no field photography was taken specifically for this experiment. Rather, a database of panoramic images that were photographed in previous years and intended for projection as panoramic images on the three screens in the LIL facility were used, from which the final image selection was made. These images largely represent forested landscapes from British Columbia, varying from steep to gently rolling terrain. The imagery used was taken using a tripod and mostly SLR camera with a 50mm lens (though some images were taken using a 35mm lens)[2], and then digitally stitched (with appropriate overlap of original images to minimise distortion) to create panoramic images of various width (up to 360 degrees). From an initial 62 possible images identified in the database, 25 images were retained for the experiment using the following criteria (in order of importance):

- Images wide enough (at least 120 degrees)
- Images tall enough (at least 85% of LIL screen height when properly located horizontally on LIL screens - see below for more details on this).
- No primarily foreground views.
- Clear images: not too dark, fuzzy, misty, over-exposed, heavily shaded or shadowed, etc.
- No distracting foreground: e.g. large road surface, foreground logging debris, dominant bridge railing, etc.
- No obvious distortion from wide-angle photography or apparent barrel distortion effects from splicing with linear features such as roads in the immediate foreground.
- No obvious duplication of the view in another image (some small overlap of adjoining pans was considered acceptable).
- No obviously recognizable images, e.g. no scene of the Vancouver skyline, given that UBC students were used as subjects.
- No images with specialized industrial or residential/urban land use dominating the scene: rural or natural/wildland context was required, with strong presence of trees/forests.
- No unusual viewpoints (e.g. no aerial views).

In addition to these criteria, a balanced mix of the following conditions was sought:

- Levels of human activity in the landscape, ranging from primarily natural, to moderate, to heavy resource management activity.

[2] Given that the images were edited to represent 120 degrees of field of view laterally, that their aspect ratio was preserved and that the subject's position to the screen was measured as to reproduced a 50mm lens projection, it was considered that all images represented what would be seen on the site or with a 50mm lens once the images were edited.

- Mix of weather from sunny/clear to grey/cloudy.
- Mix of scenic quality, from high to moderate to low (inherent and man-influenced).
- Mix of largely middleground landscape compositions: views across valley to hillside in middleground, axial views down valley and images with a mix of both.
- Mix of types of man-made activity: mostly forestry, but also farming, roads, bridges, scattered buildings, utilities, vehicles, etc. (but no largely urban or industrial images).
- Mix of panoramas with consistent landscape conditions across the entire width and those with marked landscape changes in one or more parts of the panoramic image.

2.4 Image preparation procedure

The end goal for the images to be used in the experiment described in this paper was to have a fairly consistent set of base images that would ideally be 120 degrees in width, with no distortion (initial horizontal/vertical ratio preserved from the original 50 mm photography) and that would be 3072 pixels in width by 768 pixels in height (as the target height, matching the height of the three screens). In order to achieve these specifications, the following procedure was applied to large panoramic images from the CALP image database:

a) Verification of image field of view and cropping (if necessary) of the images to 120 degrees in width.

The first step was to identify key landscape features from both a printed version of the digital image files (images were initially digitized either from scanned slide negatives or were taken directly with a 50mm digital cameras) and on a plan topographic map. Then, from the topographic map, we calculated the angle (from the previously recorded viewpoint where the photo was taken) between each of the key landscape features identified on the image. Once these angles were known, the image distance between each of the key landscape features identified in the image was measured (in cm) in order to obtain a list of paired angles (in degrees) and horizontal image distances (in cm). These were then averaged (with obvious outliers removed) in order to obtain the magnification or “scale” of the printed image in degrees per cm. Once this figure was obtained, it was multiplied by the width of the image (in cm) to obtain an assessment of how wide the image was in terms of degrees. Finally, once the width of the image was known in degrees, the images would be cropped down to a width of 120 degrees (based on the degree/cm figure just calculated) using image-editing software (Photoshop). It should be noted that this cropping operation on the width of the image did not affect at all the height of the image in any way.

For example, if the bearings for two mountain peaks from a common viewpoint are 34 degrees apart from each other, and on a corresponding photograph, which includes these two peaks, the horizontal distance between them (measured on the image) is 8.5 cm, then each cm on the image represents 4 degrees of horizontal angle. Then, by extension, if the image is 51 cm in width in total, the field of view captured by the image is approximately 204 degrees. Consequently, the image width needs to be cropped by 84 degrees (to get 120 degrees) or by 21 cm. This was done using image-editing software by cropping the image canvas down to 120 degrees (or down to 30 cm in the case of the example here).

b) Shrinking the 120 degree images down to 3072 pixels in width

Once the image is cropped to 120 degrees, the same image-editing software was used to shrink the image (it is important here to maintain the horizontal and vertical aspect ratios of the original image) down to 3072 pixels, in order for the 120-degree image to fit the three screens of the LIL facility described above.

c) Cropping down to 768 pixels in height

The next step, once a 3072 pixel-wide image (i.e. 120 degrees in terms of field of view covered) was obtained, was to crop the image canvas down to the needed height of 768 pixels, if necessary.

d) General image editing

A few general editing routines were then undertaken, in order to remove unwanted variability in the selected imagery, and to increase the image quality where necessary. These routines included treating any significant camera artefacts on the image such as obvious exposure changes across splices or major colour inconsistencies across splices, and obvious skyline inconsistencies at splices (which sometimes remain after the automated digital stitching operation). Some vehicles and other man-made equipment were tolerated on the images, unless the vehicle would be obviously associated with the camera crew (i.e., researchers). Vehicles on roads are an expected occurrence in many working landscapes, even though they may not get noticed in small images.

e) Some problems encountered

A problem that was encountered with the small number of images was when images were not “high” enough to obtain the 768 pixels in height needed for the experiment. This mostly occurred with images taken with a 50 mm lens, but also occurred (with less frequency) with the few images taken with a 35 mm lens.

An example is the image shown in Figure 7 below. View angle calculations reveal that the entire stitched image is well over 120 degrees in width, but the aspect ratio is less than 0.25 (768 pixels divided by 3072 pixels). This particular problem occurred when the 120-degree image was shrunk down to 3072 pixels in width to fit the screens of the Landscape Immersion Lab facility used, since the height did not always meet the minimum vertical requirement of 768 pixels. This required the image to be digitally modified (using photo-editing software) in order to “make up” for the missing pixels in height. An example is provided (see Figure 7 and Figure 8) showing an image that was digitally modified to increase its height to 768 pixels without changing the scale and/or ratio. In order to have an image of at least 768 pixels, it was necessary to borrow and expand the sky and in some cases “borrow” features such as trees and shrubs from the actual image to fill-in the blank areas.

Figure 7: Example of an image lacking the minimum height of 768 pixels (missing height highlighted in red) when the width is adjusted to 120 degrees and 3072 pixels.

Figure 8: Same image as in figure 7 above, after photo-editing in order to "fill-in" the missing height up to the required height of 768 pixels. Photo-editing credit: Paul Picard, CALP.

Once all the images were prepared at 120 degrees in width and had a size of 3072 pixels by 768 pixels (corresponding to condition D), the images were either shrunk down to a width of 1024 pixels (to create condition C), cropped down to the middle 1024 pixels (retaining the middle one third of the image, creating condition B), or cropped down to the middle 1024 pixels (as in Condition B) and then shrunk down to 282 pixels in width (creating condition A).

2.5 Experimental design and presentation format

The experimental design reflected the following four points:

- Scientific literature indicates that respondents' fatigue starts to be noticeable in the quality of the ratings after about 60 ratings, and that the data starts to be much less reliable beyond 100 ratings
- We needed to ask four questions involving the respondents rating images (see next section for more details on the questions asked)
- Four conditions (A, B, C, and D) need to be tested.
- The time an image is shown can affect perceptions, so the presentations need to be timed.

It was decided to show only one condition (randomly assigned) to each subject, requiring 4 separate treatment groups. In order to confirm that all four treatment groups were similar types of people, we included a common question with a common set of images at the beginning of each session (Question 1), as seen by all subjects. This initial question focused on scenic beauty, and used a mix of images showing landscapes that were largely natural, partially modified, and heavily modified by man-made activities. These were shown only on the middle screen (as in condition B) but at an intermediate size of 7 x 9 feet (with the image edges forming a 40 degree angle on the subjects' retina as in Condition B). Subjects were randomly assigned to one of the 4 treatments.

Also, in order to avoid any possible effect from the order in which the images are presented, three random images orders were used within each of the conditions tested. This was achieved by creating a total of twelve (12) presentations (one for each of the four conditions replicated in three different random image orders), and to assign each group of 5 subjects to one of these 12 experimental presentations on a random basis. The presentations were delivered by a researcher reading from a script, and each image shown was on a 15-second timer.

The design described above resulted in the subjects having to provide only 90 ratings (including 15 preview images), over a period of approximately 30 minutes.

2.6 Questions asked

Of the many possible response dimensions to be tested, three were identified as being highly relevant to the research project described: scenic beauty, naturalness and acceptability. These address a standard dimension for assessing affect (scenic beauty), a cognitive question addressing peoples ability to detect sings of naturalness or human activity, and an evaluative response on acceptability that mirrors people's opinions on planning issues in real world consultation and decision-making. This led to four actual questions for which perceptual ratings were collected, plus two open-ended questions (Questions 5 and 6 below), which explore some of the underlying meanings and opinions behind the ratings. These questions were phrased as follows:

2.6.1 QUESTION 1 (SCENIC BEAUTY –common to all participants)

In the context of forested landscapes in BC, how would you rate the scenic beauty of this scene on a scale of 1-10, with 1 meaning very low scenic beauty and 10 meaning very high scenic beauty? Please circle the appropriate value for each scene shown, and try to use the full spectrum of the scenic beauty scale.

2.6.2 QUESTION 2 (SAME AS QUESTION 1 BUT WITH DIFFERENT IMAGES)

Again, in the context of forested landscapes in BC, how would you rate the scenic beauty of this scene on a scale of 1-10, with 1 meaning very low scenic beauty and 10 meaning very high scenic beauty? Please circle the appropriate value for each scene shown, and try to use the full spectrum of the scenic beauty scale.

2.6.3 QUESTION 3 (NATURALNESS)

In the context of forested landscapes in BC and based on what you can see in the scene, how would you rate the naturalness of this scene on a scale of 1-10, from 1 (appearing highly modified by human activity) to 10 (appearing almost completely natural or largely unmodified by human activity)?

2.6.4 QUESTION 4 (ACCEPTABILITY)

In the context of forested landscapes in BC, how much do you agree with the following statement: "Land management activities that look like this are acceptable to you." Please circle the appropriate value for each scene shown, from 1 (Strongly Disagree) to 5 (Strongly Agree)

2.6.5 QUESTION 5 (OPEN-ENDED QUESTIONS)

a) What factors do you think contribute to higher scenic beauty in the scenes shown?
b) What factors do you think contribute to lower scenic beauty in the scenes shown?
c) What factors do you think make land management activities more acceptable in the scenes shown?
d) What factors do you think make land management activities less acceptable in the scenes shown?

2.6.6 QUESTION 6 (OPEN-ENDED QUESTIONS)

The subjects were then shown all four images conditions (A, B, C, and D) side by side, and then were asked to answer Question 6 on their evaluation of the different image formats, which was phrased as follows:

a) Which presentation format(s) would you prefer as a display technique, if you were being asked for your opinions on how to manage forestlands in these areas? Please write the format names (A, B, C, and D) in order from best to worst, and briefly indicate why you feel this way.

b) Which presentation format do you think makes you feel most like being in the real place? Please write the format names (A, B, C, D) in order from “Most like being there” to “Least like being there”, and briefly indicate why you feel this way.
c) Do the different presentation formats (A, B, C, D) change how you feel about the land management activities shown in the 3 scenes, and if so, how?

3.0 RESULTS

Several key results were obtained from the experiment described in this paper and are described below. These results are preliminary and further testing and analysis on the data obtained is needed in order to validate (or otherwise) the results presented below.

3.1 Image order and chair number

In order to assess whether the order in which the images were presented had any effect on the ratings, different random orders were used in the image presentation. The ratings for each image for each random image order were then compiled and analyzed using analysis of variance applied to the data from Question 1, because all subjects were shown the same image size/context (except for the image order) for this question. Results show that the order in which an image was shown had no significant effect on the rating obtained for this image. Figure 9 shows the analysis of variance output for the image order analysis.

ANOVA: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Image Random order 1	10	53.80588	5.380588	5.355168
Image Random order 2	10	59.725	5.9725	4.471514
Image Random order 3	10	61.76744	6.176744	4.26359

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Random orders	3.419802	2	1.709901	0.36406	0.698209	3.354131
Within Groups	126.8124	27	4.696757			
Total	130.2323	29				

Table 1: Analysis of variance output for the image order showing no significant effect of image order on subject ratings, using average respondent ratings for each of the 10 images shown for each of the three random orders used.

Similarly, an analysis of variance was undertaken between each of the five chairs used in order to assess whether there was a significant difference in the ratings obtained between any of the five chairs. Again, results show that the chair in which a subject was positioned had no significant effect on the rating obtained, based on the data from Question 1 using non-panoramic imagery. Figure 10 shows the analysis of variance output for the image order analysis.

ANOVA: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Chair 1:	10	58.09964	5.809964	5.202556
Chair 2:	10	57.84114	5.784114	4.837662
Chair 3:	10	57.9255	5.79255	4.784495
Chair 4:	10	58.42292	5.842292	4.790662
Chair 5:	10	58.48391	5.848391	4.637806

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups (between chairs)	0.033421	4	0.008355	0.001723	0.999994	2.578737
Within Groups (between images)	218.2786	45	4.850636			
Total	218.312	49				

Table 2: Analysis of variance output for the chair number showing no significant effect on subject ratings, using average respondents ratings for each of the 10 images shown for each of the five chairs used.

3.2 Experimental condition effect

One of the main goals of the study described in this paper was to test whether image size and image context had any effects on the ratings obtained. When all four conditions are considered together, and with data from all chairs and all random orders pooled together (since they had no significant effect on ratings), results from an analysis of variance on Questions 2, 3, and 4 show that the experimental condition represented (either condition A, B, C, or D) had no significant effect on the average ratings obtained. More specifically, as shown in Table 3, when the average rating of all respondents for each condition (A, B, C, or D) for each images are used as inputs in a one-way analysis of variance (experimental condition as the only factor), results show that there is no significant effect of context or scale for any of the three questions asked on the conditions (Questions 2 to 4). Figure 11 shows an example of the analysis of variance output for the experimental condition analysis for question 2.

	Image 1	Image 2	Image 3	Image 4	…	Image 17	Image 18	Image 19	Image 20
Condition A	5.91	8.24	6.70	7.09	…	2.52	6.94	5.52	4.85
Condition B	6.70	7.23	7.07	7.20	…	3.10	7.50	6.17	5.00
Condition C	6.61	6.71	6.87	6.42	…	3.00	6.97	6.00	4.77
Condition D	6.26	7.20	6.71	6.65	…	3.26	6.65	5.60	4.03

Table 3: Input data for the one-way analysis of variance (truncated here for simplicity sake) testing the effect of experimental condition on ratings for question

2. Note that this analysis uses the average ratings obtained for each image under each condition. Results of the analysis are shown in Table 4 below.

ANOVA: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Condition A	20	128.9091	6.445455	1.734677
Condition B	20	136.3	6.815	1.461927
Condition C	20	133.4516	6.672581	1.331209
Condition D	20	128.0258	6.40129	1.45564

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2.27569	3	0.758563	0.507107	0.678551	2.724946
Within Groups	113.6856	76	1.495863			
Total	115.9613	79				

Table 4: Analysis of variance output for the experimental condition showing no significant effect on subject ratings for the second question.

However, it is important to note that the above results are based on the average rating that an image obtained. Also, given that taking the average of a pool of numbers will result in less variance being retained (one number instead of several), it was hypothesized that this reduction in variance might have increased the probability to commit a type II error (accept the null hypothesis and conclude that there are no significant effect of either scale or context). Consequently, and as a test to this hypothesis, further analysis was undertaken for Question 2 (as a test), and two-way analyses of variance with replications were undertaken.

The two factors used for these analyses were the experimental condition (factor 1) and the image shown (factor 2) and each individual respondent was considered a replication of the experiment (resulting in 30 replications rather than a single averaged value as in the analysis from Tables 3 and 4 above). The results of these two-way analyses of variance for Question 2 (Scenic Beauty) are quite different from those presented above and are summarized in Tables 5 to 8 below.

	Image 6	Image 7	Image 8	Image 9	Image 10	Image 11	Image 12	Image 13	Image 14	...	Image 21	Image 22	Image 23	Image 24	Image 25	Total
Condition A																
Count	30	30	30	30	30	30	30	30	30	...	30	30	30	30	30	600
Sum	181	247	199	214	206	195	165	208	198	...	224	71	209	169	150	3921
Average	6.0333	8.2333	6.6333	7.1333	6.86667	6.5	5.5	6.93333	6.6	...	7.46667	2.36667	6.96667	5.63333	5	6.535
Variance	2.9989	1.7023	3.3437	2.0506	5.01609	3.63793	3.91379	3.37471	5.07586	...	3.49885	1.8954	2.86092	3.55057	3.10345	4.78675
Condition B																
Count	30	30	30	30	30	30	30	30	30	...	30	30	30	30	30	600
Sum	201	217	212	216	195	213	170	219	219	...	241	93	225	185	150	4089
Average	6.7	7.2333	7.0667	7.2	6.5	7.1	5.66667	7.3	7.3	...	8.03333	3.1	7.5	6.16667	5	6.815
Variance	4.769	3.6333	3.0299	2.9241	5.98276	2.64483	2.64368	3.45862	4.21724	...	2.86092	2.98966	2.32759	2.62644	3.44828	4.50495
Condition C																
Count	30	30	30	30	30	30	30	30	30	...	30	30	30	30	30	600
Sum	198	200	206	192	169	225	203	228	216	...	214	89	209	179	142	3983
Average	6.6	6.6667	6.8667	6.4	5.63333	7.5	6.76667	7.6	7.2	...	7.13333	2.96667	6.96667	5.96667	4.73333	6.63833
Variance	2.731	4.1609	2.0506	3.8345	6.37816	2.94828	3.42644	1.76552	2.57931	...	3.29195	2.65402	2.92989	3.96437	4.4092	4.34143
Condition D																
Count	30	30	30	30	30	30	30	30	30	...	30	30	30	30	30	600
Sum	190	216	201	197	142	231	193	231	203	...	191	92	203	172	123	3849
Average	6.3333	7.2	6.7	6.5667	4.73333	7.7	6.43333	7.7	6.76667	...	6.36667	3.06667	6.76667	5.73333	4.1	6.415
Variance	3.4023	3.6138	1.9414	3.4954	4.2023	1.80345	3.08161	2.7	4.11609	...	1.82644	3.78851	3.15057	2.13333	3.61034	4.40679
Total																
Count	120	120	120	120	120	120	120	120	120	...	120	120	120	120	120	
Sum	770	880	818	819	712	864	731	886	836	...	870	345	846	705	565	
Average	6.4167	7.3333	6.8167	6.825	5.93333	7.2	6.09167	7.38333	6.96667	...	7.25	2.875	7.05	5.875	4.70833	
Variance	3.4552	3.5182	2.5543	3.1204	5.9451	2.90084	3.46211	2.84342	3.98207	...	3.16387	2.84979	2.82101	3.03466	3.68732	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample (Condition)	51.685	3	17.228	5.474	0.00095	2.60874
Columns (Images)	3173.5	19	167.02	53.069	6E-166	1.591
Interaction	330.71	57	5.802	1.8435	0.00014	1.33261
Within	7301.7	2320	3.1473			
Total	10858	2399				

Table 5: Results of a two-way analysis of variance for Question 2 on Scenic Beauty (truncated for simplicity's sake), comparing all four experimental conditions (factor 1) and the images shown (factor 2) using the raw subject data (as 30 individual replications per condition) as opposed to the computed average ratings. Results show a significant effect of the experimental conditions shown, as well as a significant effect of the image shown and significant interaction between these two factors.

	Image 6	Image 7	Image 8	Image 9	Image 10	Image 11	Image 12	Image 13	Image 14	...	Image 21	Image 22	Image 23	Image 24	Image 25	Total
Condition A																
Count	30	30	30	30	30	30	30	30	30	...	30	30	30	30	30	600
Sum	181	247	199	214	206	195	165	208	198	...	224	71	209	169	150	3921
Average	6.0333	8.2333	6.6333	7.1333	6.86667	6.5	5.5	6.93333	6.6	...	7.46667	2.36667	6.96667	5.63333	5	6.535
Variance	2.9989	1.7023	3.3437	2.0506	5.01609	3.63793	3.91379	3.37471	5.07586	...	3.49885	1.8954	2.86092	3.55057	3.10345	4.78675
Condition B																
Count	30	30	30	30	30	30	30	30	30	...	30	30	30	30	30	600
Sum	201	217	212	216	195	213	170	219	219	...	241	93	225	185	150	4089
Average	6.7	7.2333	7.0667	7.2	6.5	7.1	5.66667	7.3	7.3	...	8.03333	3.1	7.5	6.16667	5	6.815
Variance	4.769	3.6333	3.0299	2.9241	5.98276	2.64483	2.64368	3.45862	4.21724	...	2.86092	2.98966	2.32759	2.62644	3.44828	4.50495
Total																
Count	60	60	60	60	60	60	60	60	60	...	60	60	60	60	60	
Sum	382	464	411	430	401	408	335	427	417	...	465	164	434	354	300	
Average	6.3667	7.7333	6.85	7.1667	6.68333	6.8	5.58333	7.11667	6.95	...	7.75	2.73333	7.23333	5.9	5	
Variance	3.9311	2.8768	3.1805	2.4463	5.4404	3.17966	3.23023	3.39294	4.69237	...	3.20763	2.53785	2.6226	3.10847	3.22034	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Condition	23.52	1	23.52	7.3785	0.0067	3.84949
Images	1807.4	19	95.125	29.842	1.3E-86	1.59549
Interaction	60.68	19	3.1937	1.0019	0.45559	1.59549
Within	3697.7	1160	3.1876			
Total	5589.3	1199				

Table 6: Results of a two-way analysis of variance for Question 2 on Scenic Beauty (truncated for simplicity sake), comparing experimental conditions A and B (factor 1) and the images shown (factor 2) using the raw subject data (as 30 individual replications per condition) as opposed to the computed average ratings. Results show that for the same content (deprived of lateral context), but at different scales, a significant effect is found between the two conditions, showing that scale has a significant effect on Question 2 ratings (at least for the scale tested) when no lateral context is provided.

	Image 6	Image 7	Image 8	Image 9	Image 10	Image 11	Image 12	Image 13	Image 14	...	Image 21	Image 22	Image 23	Image 24	Image 25	Total
Condition C																
Count	30	30	30	30	30	30	30	30	30	...	30	30	30	30	30	600
Sum	198	200	206	192	169	225	203	228	216	...	214	89	209	179	142	3983
Average	6.6	6.6667	6.8667	6.4	5.63333	7.5	6.76667	7.6	7.2	...	7.13333	2.96667	6.96667	5.96667	4.73333	6.63833
Variance	2.731	4.1609	2.0506	3.8345	6.37816	2.94828	3.42644	1.76552	2.57931	...	3.29195	2.65402	2.92989	3.96437	4.4092	4.34143
Condition D																
Count	30	30	30	30	30	30	30	30	30	...	30	30	30	30	30	600
Sum	190	216	201	197	142	231	193	231	203	...	191	92	203	172	123	3849
Average	6.3333	7.2	6.7	6.5667	4.73333	7.7	6.43333	7.7	6.76667	...	6.36667	3.06667	6.76667	5.73333	4.1	6.415
Variance	3.4023	3.6138	1.9414	3.4954	4.2023	1.80345	3.08161	2.7	4.11609	...	1.82644	3.78851	3.15057	2.13333	3.61034	4.40679
Total																
Count	60	60	60	60	60	60	60	60	60	...	60	60	60	60	60	
Sum	388	416	407	389	311	456	396	459	419	...	405	181	412	351	265	
Average	6.4667	6.9333	6.7833	6.4833	5.18333	7.6	6.6	7.65	6.98333	...	6.75	3.01667	6.86667	5.85	4.41667	
Variance	3.0328	3.8938	1.9692	3.6099	5.4065	2.34576	3.22712	2.19746	3.3387	...	2.66525	3.16921	2.99887	3.01102	4.04379	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Condition	14.963	1	14.963	4.8161	0.02839	3.84949
Images	1585.8	19	83.462	26.863	1.7E-78	1.59549
Interaction	50.337	19	2.6493	0.8527	0.64344	1.59549
Within	3604.1	1160	3.107			
Total	5255.1	1199				

Table 7: Results of a two-way analysis of variance (truncated for simplicity sake) comparing experimental conditions C and D (factor 1) and the images shown (factor 2) using the raw subject data (as 30 individual replications per condition) as opposed to the computed average ratings. Results show that for the same content (including lateral context), but at different scales, a significant effect is found between the two conditions, showing that scale has a significant effect on Question 2 ratings (at least for the scale tested) when lateral context is provided.

	Image 6	Image 7	Image 8	Image 9	Image 10	Image 11	Image 12	Image 13	Image 14	...	Image 21	Image 22	Image 23	Image 24	Image 25	Total
Condition B																
Count	30	30	30	30	30	30	30	30	30	...	30	30	30	30	30	600
Sum	201	217	212	216	195	213	170	219	219	...	241	93	225	185	150	4089
Average	6.7	7.2333	7.0667	7.2	6.5	7.1	5.66667	7.3	7.3	...	8.03333	3.1	7.5	6.16667	5	6.815
Variance	4.769	3.6333	3.0299	2.9241	5.98276	2.64483	2.64368	3.45862	4.21724	...	2.86092	2.98966	2.32759	2.62644	3.44828	4.50495
Condition D																
Count	30	30	30	30	30	30	30	30	30	...	30	30	30	30	30	600
Sum	190	216	201	197	142	231	193	231	203	...	191	92	203	172	123	3849
Average	6.3333	7.2	6.7	6.5667	4.73333	7.7	6.43333	7.7	6.76667	...	6.36667	3.06667	6.76667	5.73333	4.1	6.415
Variance	3.4023	3.6138	1.9414	3.4954	4.2023	1.80345	3.08161	2.7	4.11609	...	1.82644	3.78851	3.15057	2.13333	3.61034	4.40679
Total																
Count	60	60	60	60	60	60	60	60	60	...	60	60	60	60	60	
Sum	391	433	413	413	337	444	363	450	422	...	432	185	428	357	273	
Average	6.5167	7.2167	6.8833	6.8833	5.61667	7.4	6.05	7.5	7.03333	...	7.2	3.08333	7.13333	5.95	4.55	
Variance	4.0506	3.5624	2.4777	3.2573	5.79972	2.27797	2.96356	3.0678	4.16836	...	3.01017	3.33192	2.82938	2.38729	3.67542	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Condition	48	1	48	15.34	9.5E-05	3.84949
Images	1589.9	19	83.681	26.743	3.7E-78	1.59549
Interaction	118.47	19	6.2351	1.9926	0.00681	1.59549
Within	3629.7	1160	3.1291			
Total	5386.1	1199				

Table 8: Results of a two-way analysis of variance (truncated for simplicity sake) comparing experimental conditions B and D (factor 1) and the images shown (factor 2) using the raw subject data (as 30 individual replications per condition) as opposed to the computed average ratings. Results show that for the same scale, but with different content (condition B has no context), a significant effect is found between the two conditions, showing that context has a significant effect on Question 2 ratings (at least for the scale tested).

With these results, it can be seen that image size and shape do exert significant effects on at least the scenic beauty ratings. In the two-way analysis of variance for all four treatments (Table 5), the large single screen display (Condition B) was highest in scenic quality overall, while the large panoramic images (Condition D) tended to be the lowest. Variance in ratings seemed highest in the small single image (Condition A).

The results of the two-treatment comparisons show that when experimental conditions A and B are compared (same content deprived of lateral context, but at different scales), a significant effect is found, showing that image scale has a significant effect on scenic beauty ratings (at least for the image scales tested) when no lateral context is provided (see Table 6). In effect, the larger image without context (condition B) sustained consistently higher scenic beauty ratings, though the differences do not appear to be very large (overall average ratings differed by 6.8/10 for the large images and 6.5/10 for the small images). As for comparisons between experimental conditions C and D (same content including lateral context, but at different scales), results show that scale also had a significant effect here on scenic beauty ratings when lateral context is provided (see Table 7). In effect, results show that the full-size panoramic images sustained significantly lower scenic beauty ratings relative to the small panoramic images (6.6/10 for the small panoramic images and 6.4/10 for the large panoramic images).

Similarly, results show that when experimental conditions B and D are compared (same scale but different content, condition B being deprived of lateral context), a significant effect is again found, showing that context has a significant effect on ratings for scenic beauty (at least for the scale tested) (see Table 8). This appears to be a relatively strong effect, with the panoramic display having a somewhat negative effect overall on scenic beauty relative to the single large display, though inspection of results image by image reveals that this tendency is not particularly consistent across the images.

3.3 Open-ended question on scenic beauty and acceptability factors

In addition to these results, open-ended questions were asked about the factors contributing to scenic beauty and acceptability of management activities, providing more specific results as to which elements in the landscape affect perceptions of the landscape. Given that 125 subjects were used, every answer provided, while very valuable, cannot be presented here, but a few that are deemed representative are included:

Factors contributing to higher scenic beauty:

- Water, lush meadows, forests, or land, sunlight, point of view (on top of a mountain, from a valley), mountains, fresh snow.

- Not seeing exposed ground from clear cuts; not seeing roads on sides of mountains; sun shining, blue sky, cloudless; seeing mature trees compared to new growth
- Factors: leaving the environment as it is; leaving nature on its own.
- The replanting of new trees in the place of ones that have been cut down
- Forest scenes that do not show signs of clear-cutting are beautiful. Other secondary factors shown in the scenes that are beautiful include bodies of water and mountains.
- Natural, symmetric, colourful.

Factors contributing to lower scenic beauty:
- Obvious clearcutting, or clearing of land for roads or golf courses. Human habitation. Anything that looks like scars on the land. Houses, roads.
- Unhealthy brown colours, power lines, pebbly roads.
- Large square holes of no trees, larger square holes with hardly any wood material left in it.
- Destruction of natural environment.
- Obvious damage (e.g. clearcuts), uniformity (small, farmed trees).

Factors contributing to higher management acceptability:
- Anything in small amounts. A thinning of the forest rather than clearing the whole thing. Clearing small areas near a lake for cabins rather than clearing a huge area for a golf course.
- Replanting, thinning, spreading out the clear cuts so there aren't so many in one place.
- When patches are made, they are filled with grass to make it look natural, gravel roads where there must be traffic.
- Clearcuts that are divided into smaller blocks and clearcuts that stay away from the water. Reforestation.

Factors contributing to lower management acceptability:
- Bold clear cutting, murky/dirty water, as well as too much concrete.
- When the activities are noticeable so you don't see the scenery, but the activities.
- Large clearcuts.
- Areas with no growth since the logging (land management).
- When forests looked like they had been selectively logged it gave a greater appearance of naturalness - my criteria for acceptable forestry practices.
- Signs of deforestation such as wide spaces where trees were cut down.

3.4 Open-ended question on the experimental conditions

The open-ended questions evaluating the different display formats generally support the results described above, that context and scale have significant effects on respondent’s ratings. Results show that condition D (large panoramic image with context) is by far the most preferred condition while the small image without context (condition A) is the least preferred condition. Figure 9 shows these results in greater depth. The small panoramic format (Condition C) appears to be preferred over the larger single screen format (Condition B) by many.

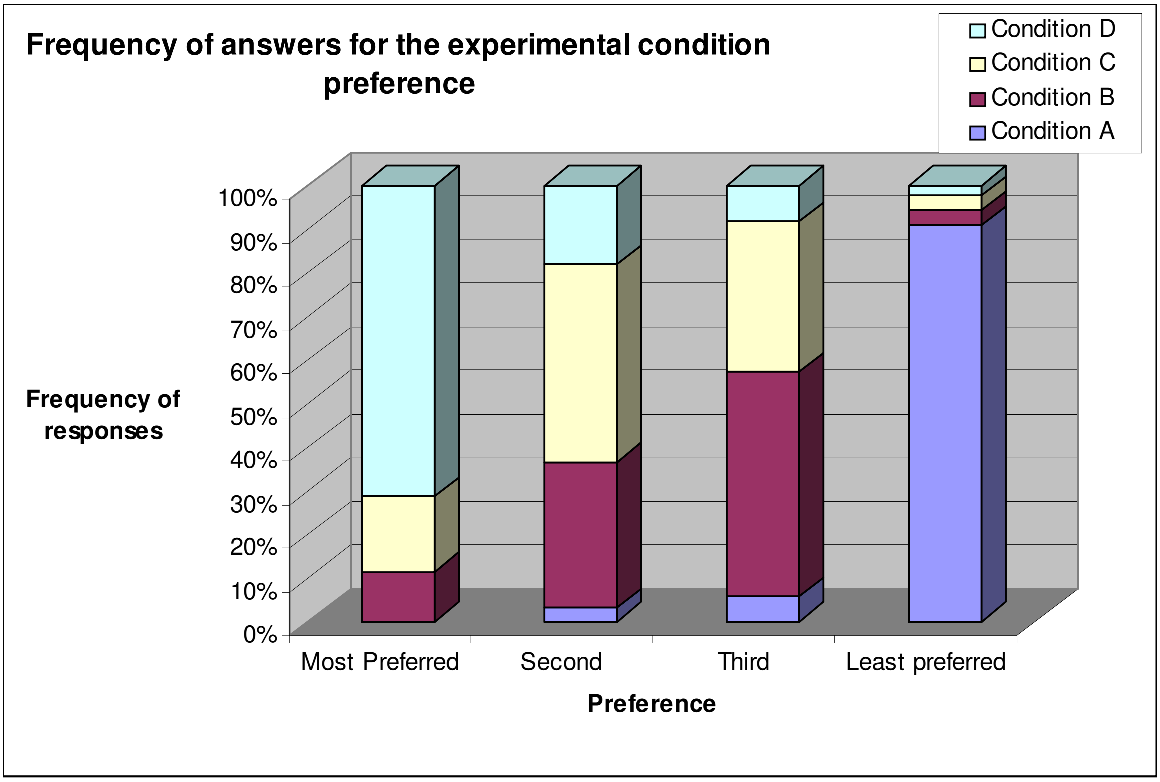

Figure 9: Frequency of answers for the preferences of the experimental conditions.

Similarly, when asked which experimental condition makes them feel most like being in the real place, respondents indicate condition D (large panoramic image with context) as being the most powerful condition and condition A (small image without lateral contextual information) as being the least effective in making them feel like they are in the real place. The results are presented in figure 10. The results appear almost identical to those in response to the preceding question, suggesting that people prefer immersive display that makes them feel more like they are in the real place.

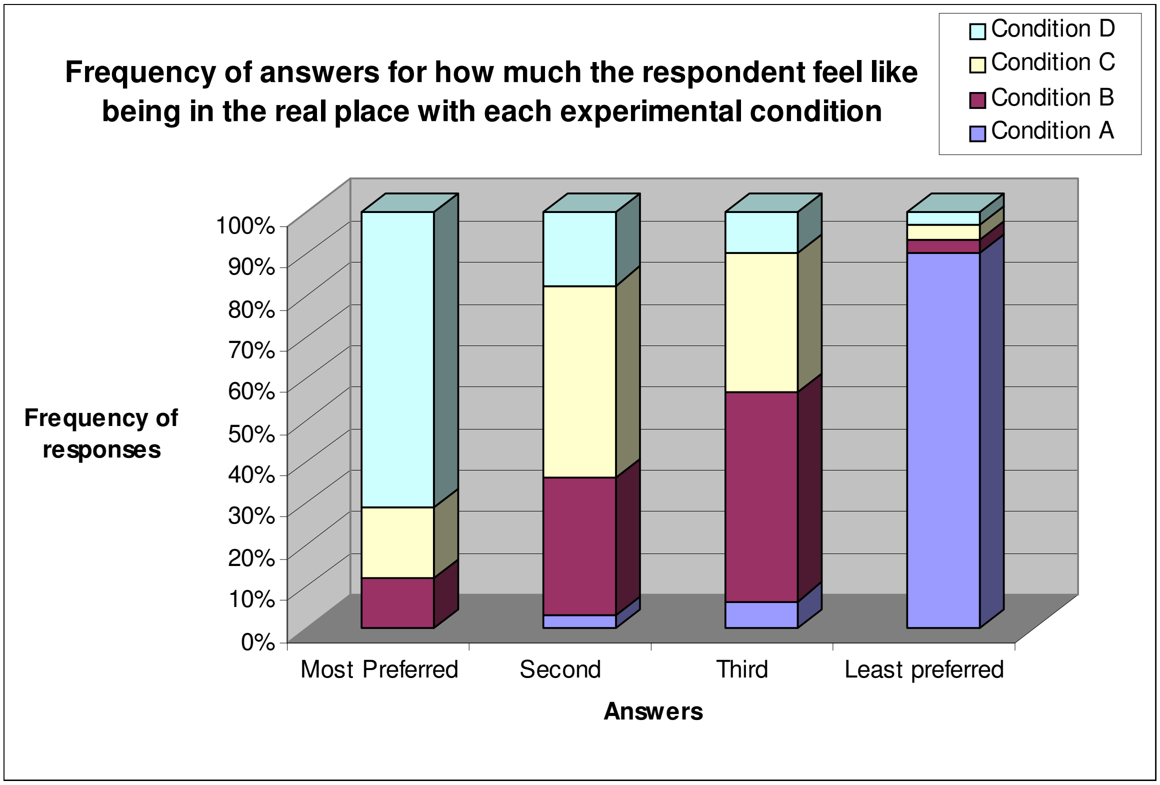

Figure 10: Frequency of answers for how much the experimental conditions make the subjects feel like they are in the real place.

In Question 6c, when asked whether the experimental conditions changed how they feel about the land management activities shown, 84% of the respondents indicated yes, showing again an effect of the image scale and of the context provided in an image as represented by the experimental conditions tested in this experiment. Another interesting finding is that of the 10 people that did not provide comments for the third part of question 6, 8 had answered “No”, indicating that the different presentation formats (A, B, C, D) did not change how they feel about the land management activities shown in the 3 scenes.

Comments from subjects which indicate that the different presentation formats (A, B, C, D) changed (they answered “Yes”) how they feel about the land management activities shown in the 3 scenes:

- I can see how each scene looks in the overall landscape and I feel more like I'm really seeing it when it is a larger image.
- Formats B & D show more detail and make me feel more involved and connected w/ the landscape. A & C are just too small and don't do the scene justice.
- Not only would I have rated the images differently if I had seen the panoramic images, the patchwork is so much more apparent that it would be the most important management problem.

- Formats D & C show the entire image, which show more evidence of logging/human activity than the cropped image A.
- While Formats A and B remind me of actual pictures, Format 'C' looks more professional and makes the scene look fake. Format 'D' made me feel like I was there.
- I feel so duped! In the small format (A) scene it looked like a pristine mountaintop, but in panorama (format D) you can see a variety of logging has gone on. Format 'A' is like looking at it with blinders on.
- Because in format A, the image is very small and therefore I can't notice all the details, and what is actually in the picture. Also, the large format (D) makes the image larger and therefore more interesting and captivates my attention versus a small image (format A).
- Definitely - A central shot of a completely logged off area looks terrible - the same logged areas look bad in the bigger/more detailed image - but hopefully the presence of unlogged areas or undisturbed shorelines gives a person hope that the damage is not extensive and therefore fixable.
- The bigger images brought some unnoticed details to my attention. Turned out some areas I perceived as humanly modified are natural, or vice versa.
- "A" didn't always show the "ugly" surrounding areas. However, "A" & "B" were usually more pleasing to look at.
- Change the feeling of space, small picture can give a better feeling of whole scene, but big picture gives more real feeling.

Comments from subjects which indicate that the different presentation formats (A, B, C, D) did not change (they answered “No”) how they feel about the land management activities shown in the 3 scenes:

- The bigger photos show more detail on the edges of cutblocks, which is important to me. A feels like a cheap version versus D, which seems expensive or better showing me the reality. I don't think it changes how I feel about the activities though.
- Although the format doesn't make me feel differently about the land management activities, the panoramic views let me see more of what's happening in the surrounding areas. This gives me a better perspective of what's actually going on in the area.
- Does not change my views but does place the locales into perspective.
- Not really, since they display the same image.
- Just have to look harder at certain image sizes.

4.0 DISCUSSION

4.1 Caveats

Despite some very interesting results obtained in the study, much more analysis is needed in order to obtain a clearer picture on a few issues. Among these issues, at least two caveats need special attention and are presented here.

First, the fact that the sum of squares from the error (labeled "within group" variation in the analysis of variance outputs above) is often much bigger than the treatment (or factor) sum of squares is thought to be a strong indication of a higher risk of committing of a type II error (accepting the null hypothesis and concluding that there is not significant effect when it is not true). In some cases, the sum of square from the error (labeled "within group" variation in the analysis of variance outputs above) is over 7,000 times bigger than the treatment (or factor) sum of square (as in Table 2).

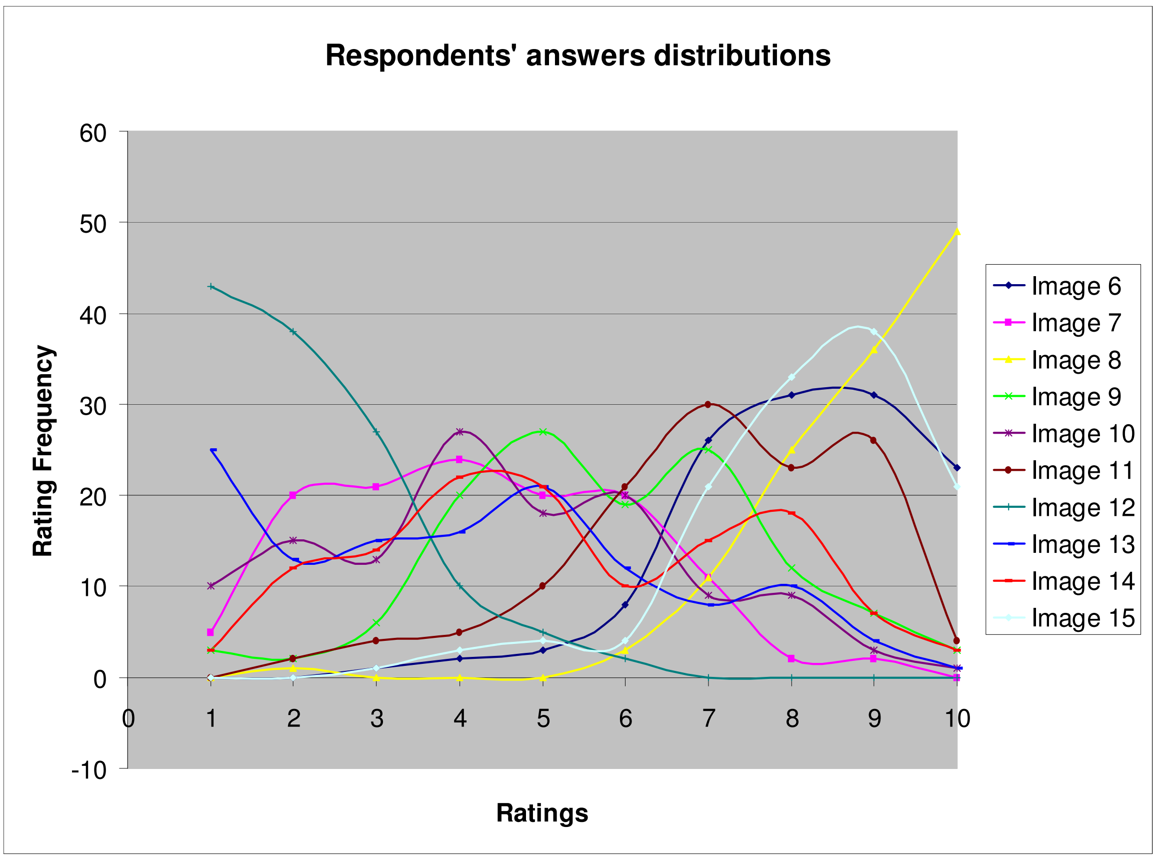

Figure 11: Distribution of the ratings for the images on the scenic beauty question (Question #1) showing some distributions that may not be normally distributed.

Another caveat to the results of the present study, is that it was assumed that the data was normally distributed and no tests of "Goodness of Fit" have yet been undertaken on the respondent's ratings. In fact, one of the assumptions of the analyses of variance is that the data is normally distributed, and the "Goodness of Fit" test is designed to test whether or not as dataset is normally distributed or not. However, after plotting the data, it appears that the use of a fixed scale (with a fixed upper limit and a fixed lower limit) may have caused the ratings of some images not to be normally distributed (as shown with images 8 and 12 on Figure 11). Also, it appears some images had a rather "flat" distribution (as opposed to a bell-shaped normal curve) perhaps reflecting the controversial nature of that image resulting in a split of opinion over it (see images 13 and 14 in Figure 11 for example).

4.2 Implications of results

A number of possible explanation appear likely for these results:

- Familiarity with the single image format/aspect ratio, even when large as in Condition B (which emulates a good standard slide projection screen set-up) may contribute to higher scenic beauty ratings of the scenes.
- The lower scenic beauty ratings of the large panoramic screens may reflect several factors, including possible overload of information under stressful conditions of answering repeated questions within a limited time when there is so much to take in; unfamiliarity with the format (not knowing how to extract the right information out of it, which may be more like the real world; and the negative effect of small details of man-made activities that may be less visible with smaller images and less encountered in more limited context-reduced images like Condition B; essentially there is a 3 times higher chance of containing a negative element in the panoramic images than the single screen images, yet the large scenic quality factors inherent in the larger natural landscape are likely to come through regardless of the scale of format.
- Ambiguity of information presented in small limited displays like Condition A, where the resolution may require participants to guess more at the content, perhaps contributing to higher variance and fewer extremes of scores.

The validity of the different formats cannot be ultimately confirmed without response equivalence testing against real world conditions. However, some initial implications for decision-makers and users of visualisation in public settings would suggest that larger screens and viewing distances do make a difference, and there appear to be no great advantages to smaller images; the great likelihood is that in most cases the viewing distances are too large and/or the image sizes are too small to get the most out of the imagery. The importance of adding context with panoramic imagery has been underscored, with the comparisons of Conditions B and D, and A and C it may be relatively easy to miss important context in non-panoramic imagery. The fact that people strongly prefer the large panoramic screen may be in part due to novelty, though it could also be due to identification with pleasurable entertainment displays (like Imax or movie theatres) or remind them more powerfully of the real world conditions.

4.3 *Future experiments to consider*

Several questions arise from the experiment described in this paper, each of which should be the focus of future research endeavours in order to validate and strengthen the results obtained. Key areas that should be prioritized include the following:

1. Have the subjects scroll between the different sized/formatted images and give more in-depth and direct feedback on what they prefer and why.
2. Take subjects to selected field locations and have them rate the location/place, then compare these ratings with ratings of the same places/panoramas from the LIL facility and test for any differences in response equivalence. This should be combined with new photography customised for this experiment, rather than using post-hoc conversion and verification of previously collected slides.
3. Do a more in-depth session similar to the experiment described in this paper but with more open-ended questions, more restricted landscape conditions, with clearer categories of image content (eg. altered states in peripheral images versus internally consistent panoramic images across the field of view); and perhaps with a smaller group of people, focusing the session more as a qualitative investigation of people's concerns and reactions.
4. Investigate and explore the theory that people may become more emotional in immersive settings, i.e. people may bond more with the landscape that they see, perhaps through physiological measurements or other emotional tests.
5. Replicate the experiment using various visualisation imagery to determine if the same relationships apply with rendered images, of if new issues arise.

5.0 CONCLUSIONS

In conclusion, significant results were obtained from the present study. Probably most important are the findings that scale and context have a significant effect on ratings, and more care should be taken in arranging the appropriate scale and image breadth in presentations of photography or visualizations. However, given the high variability of the respondent's answers (as expressed by the error sum of squares), such effects may be masked (or harder to detect) than initially thought. Further analysis and testing is recommended to isolate with more clarity these effects.

Also useful for further experimentation in the LIL facility at UBC is the finding that there are no significant perceptual differences due to chair placement with the five chairs used, at least with central screen displays; this will benefit research by requiring less sessions (by using five subjects at a time instead of three). Given the width of the space between the three screens, and unless a second row behind the first one is inserted, this figure (five subjects at a time) is probably optimal for the LIL facility.