Reliability and Validity of an Expert System for Landscape Visual Quality Assessment

by

Donald H. Schlagel

Thesis submitted to the Graduate Faculty of Virginia Polytechnic Institute and State University in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

Forestry

APPROVED:

Gregory J. Buhyoff, Chairman

Patrick A. Miller

R. Bruce Hull

April, 1994
Blacksburg, VA

Reliability and Validity of an Expert System for Landscape Visual Quality Assessment

(ABSTRACT)

A previously developed expert system for landscape visual quality assessment was evaluated for reliability and validity against four professional landscape architects, one of whom's rules upon which the system was based, and two lay subjects. Results indicated that the system has good initial reliability and internal validity, but the external validity assessments appeared poor. It is believed that this was due to the subjects' unfamiliarity with the system's format and logical construction. For this reason, it is recommended that a training program be developed and the reliability and validity be reassessed, with and without training the subjects.

Acknowledgements

I can not begin to adequately thank all of the people who have provided aid, encouragement, and insight to me in the completion of this project, but here are a few of many.

First of all, I would like to thank my advisor, Dr. Gregory J. Buhyoff. Without his frequent editorial advice and input this research would not have attained its present form. I learned much during my time at VPI and many of my newly-improved writing skills I owe to him.

Next, I would like to thank Dr. John Roach in the Computer Science department at VPI. His instruction in the PROLOG programming language was invaluable in EVA's integration into a unified system for landscape visual quality analysis.

Dr. Patrick Miller of VPI's Landscape Architecture department was the central figure in this research. Without many hours of his time and devotion EVA would not exist, nor would the reliability testing portion of this research.

I would like to thank Dr. Bruce Hull for his insight and probing questions relating to reliability and validity testing. Without his input, I would not have been ready to defend this research or my thesis.

Also, this research was supported by the United States Forest Service, North Central Forest Experiment station, Chicago, Illinois through Cooperative Research

Agreement 23-93-03. Their support is gratefully acknowledged. The opinions expressed herein are those of the author only.

I also would like to thank my parents, Jim and Emilie Schlagel. Without their love and support during my move to Virginia and my subsequent cultural adaptation, I would not have endured to this point.

Last, but not least, I would like to thank my wife Deborah for her love and encouragement during the final year of my studies at VPI. I could not have successfully completed my thesis without her understanding and the many sacrifices she has made in our sharing quality time together.

Table of Contents

List of Figures

Figure Page

List of Figures (Cont.)

List of Figures (Cont.)

List of Figures (Cont.)

List of Figures (Cont.)

List of Tables

Chapter 1

Introduction

Timothy O'Riordan said it was the assessment of intangible resources that represented, "the present challenge in resource management . . . but they are so much more difficult to define, to evaluate, and to allocate than the old ones" (O'Riordan 1971). Assessing or measuring visual quality, in particular, is a difficult task, at best, which faces public natural resource management.

Legislation such as the *Multiple Use and Sustained Yield Act* (U. S. Congress, 1960) and the *National Environmental Policy Act* (U. S. Congress, 1969) mandated that the national forests be managed with concern for previously immeasurable resources such as aesthetics and wildlife, as well as recreation. These legislative mandates confirmed the long-standing need for a method of managing these intangible resources.

Background

The difficulty in landscape visual quality assessment comes in knowing how to place a value on a visual setting because different people can find different settings attractive (Zube and Pitt 1981). Several studies have attempted to model and predict the reactions of observers evaluating the landscape, but results varied according to the backgrounds of the observers and the types of landscapes being evaluated. They found that the presence of man-made structures appears to be one of the factors associated

with the variability in perceived scenic quality of a landscape. It was unknown fromthis study, however, whether other cultural factors such as ethnicity, place of residence, or environmental experience were associated with the variability of the perceptions of natural landscapes.

Zube and Pitt (1979) suggest that there might be a relation between the evaluations given a landscape and the socioeconomic factors which governed the life of the observer. Inner city youths tended to find landscapes with minor man-made structures less intimidating than landscapes which were fully natural. In the 1980 study, natives of the Virgin Islands found no reduced scenic quality in developed beach scenes. The authors began to believe that not all international cultures shared the belief that man-made features reduced the scenic quality of a landscape. It was unknown whether this variation was due to cultural differences and living location (urban vs. rural), but the data seemed to suggest that there were significant differences between the cultures examined. Furthermore, it was noted that management practices developed for one culture may not be adequate for another culture, and the problem may be compounded in recreational areas frequented by cross cultural populations.

It is obvious then, that it can be extremely difficult to anticipate the reactions a viewer might have for any particular landscape. Managing landscapes for visual quality can be equally perplexing.

Another difficulty which has been considered in the area of predicting scenic quality is the ability of experts to accurately predict the reactions of certain groups to

landscapes. If a professional landscape architect could not effectively predict the desires of a client, the client would need other sources of input to obtain reliable information. One study has shown a relatively high agreement between non-experts and experts when landscape scenes are evaluated for visual quality (Zube, 1979). However, Zube demonstrated that the landscape architects could predict client reactions to an existing landscape -- not on proposed changes to a landscape.

Buhyoff et al. (1978) found that landscape architects can approximate the preferences of client groups when a general description of the groups' likes and dislikes was provided. Even though the personal preferences of the landscape architects may have been far different from the client group, the landscape architects were able to set aside personal biases and evaluate the landscapes accurately, with the motives of the client groups in mind. This study demonstrated that planners may be able to anticipate the desires of a client group, but only when they do not rely strictly on their own personal preferences.

Early attempts at managing visual resources by the United States Forest Service (USFS) began at the regional office level. In the early 1970's, the popular approach was to rely on the expert judgement of landscape architects. The USFS was the world's largest employer of landscape architects at that time (Daniel, 1990). Eventually, these regional office attempts at managing visual resources led the way to the early development of the USFS Visual Management System (VMS) (USFS, 1975; Smardon, 1986).

Visual Management System

The VMS is a straightforward system that uses easily discernible landscape features and formal constructs, which can be intuitively explained to lay-people, to arrive at an evaluation of the landscape. It is designed for use primarily by landscape architects since it uses terms developed in formal training of the field, but its "language" is nonmathematical and can be explained to a nonexpert. Three main components comprise the system : classification of the landscape, sensitivity analysis, and a management decision. The classification component of this system does attempt to yield a concise, verbal description of the landscape, but it is not a true evaluation of the landscape, rather it is an inventory of the natural features present, the public use of the land, and public concern for the land. These attributes, as well as a sensitivity analysis of the landscape, are then analyzed and a management suggestion is created for the land in question.

This management suggestion includes a landscape classification which is a broad overall description of the landscape being evaluated (USFS, 1975). Much detail is lost, however, when the all the factors that go into the final classification of the landscape are condensed into a single, final result. There is no way to recapture the thoughts and motivations of the expert doing the classifications of the landscape (Hull, 1989). The motivations of the expert are essential to understanding why a landscape was ultimately classified in the manner it was. The VMS is an example of a design based classification and assessment (Buhyoff et al., 1994). There are other methods,

however, of landscape evaluation that have been pursued over the past two decades: psychological predictors and psychophysical models.

Psychological Predictors

Psychological constructs can be used to predict the way people will react to certain landscapes through dimensional analysis of landscapes (Kaplan and Kaplan, 1982). The psychological approach assumes that people react primarily to the content, but also to the spatial organization of a landscape. Kaplan et al. (1989) stated that visual resource management depends on the recognition of these environmental attributes that seem to contribute to scenic quality. The Kaplans have identified several important environmental attributes which define spatial organization as being an important predictor of visual landscape preference. These are : "complexity", "mystery", "legibility", and "coherence". In several studies conducted between 1982 and 1989 the Kaplans found that coherence does not play a major role in predicting visual preference, but mystery consistently appears to be a very strong positive indicator for scenic preference (Kaplan et al., 1989). Although these studies have provided very useful and reliable content domains for predicting scenic quality, there has been no assimilation of the results into an applicable system for evaluating scenic quality (Dearden, 1981; R. Kaplan, 1985; Ruddell et al., 1989). Buhyoff et al. (1994) point out that in order to be understood and used by resource managers, previous

assessment approaches have had to be simple and intuitively understandable, thereby eliminating the inclusion of psychological variables.

Psychophysical Models

Psychophysical models were developed to find a mathematical relationship between physical features of a landscape and the perceptual judgements of human observers. Tangible landscape features such as land cover, forest stand structure, land use, and arrangement are measured and then related statistically to scenic quality opinions (Buhyoff et al., 1994). Most frequently, multiple linear regression has been used to determine these relationships. Studies (Daniel, 1990; Dearden, 1981; Brown and Daniel, 1986) have shown that psychophysical models are reliable and very sensitive to landscape variations. Furthermore, psychophysical models provide good predictors of public preferences for the scenic quality of a wide range of landscapes (Daniel and Vining, 1983). A problem with psychophysical models is that they tend to lack theoretical content in their structure (Ruddell et al., 1989). Furthermore, they are frequently viewed as "black box" models because the final result is usually an unexplained numerical value which comes from complex statistical calculations and manipulations that are not easily understood by the layman (Buhyoff et al., 1994).

A New Approach

These traditional approaches to visual quality assessment, have individual strengths and weaknesses. However, it is possible that artificial intelligence technology might be useful for designing a tool (in the form of an expert system) that would integrate consistant results stemming from this reasearch and use of these traditional methods (Coulson et al., 1987; Zhou, 1992).

Buhyoff et al., (1994) stated that one of the most important attributes of any new landscape evaluation system must be its ability to explain why a landscape is more or less visually appealing than another. Buhyoff and Fuller (1993) made a strong case for including explanations in any form of modeling system as well. An expert system would yield an explanation of why a landscape was given a particular evaluation (Buhyoff et al., 1994).

Expert Systems

A standard expert system contains a user interface, and inference engine, and a knowledge base specific to the realm of the system (Stock, 1987). Using these tools, the expert system receives information about a situation from a user (via the user interface), relates that information and makes inferences and deductions about it (via the inference engine), and formulate an answer to a problem (by searching through the knowledge base) (Zhou, 1992). Furthermore, an explanation of the resulting solution (describing the rationale of the system) should be generated (Swartout, 1987).

Expert systems, by definition, employ human knowledge to solve problems that ordinarily require human knowledge (Hayes-Roth, 1987). They are required to perform close to the standards expected of a human expert. The only way to confirm that a particular system does adhere to the quality standards set forth in the definition is to test the system for validity. Without such a test there is no way to know whether the system is adequately designed to model the results a human expert would give (O'Keefe et. al., 1988).

Explanation Visual Assessment (EVA)

Previous work (Zhou, 1992) at Virginia Polytechnic Institute and State University has led to the development of an artificial intelligence based expert system for the evaluation of landscape visual quality. Using criteria and decisions developed from the decision making process of a professional landscape architect, an expert system was constructed which ostensibly can make evaluations of landscape visual quality based on a number of landscape design attributes. The system contains four parts, each of which deal with a particular area of landscape evaluation: man-made features, natural features, spatial composition, and visual composition. These parts or modules were then unified into a single comprehensive expert system (Buhyoff et al., 1994).

The Explanation Visual Assessment system (EVA) contains over 7500 lines of code (Zhou, 1992) and was written in VPI PROLOG (Roach and Deigan, 1990).

There are five different components which comprise EVA, four of which deal with landscape attribute evaluations and a final visual quality rating component (Buhyoff et al., 1994). The four components which deal with landscape attributes are separate explanation systems capable of working independently or in conjunction with one another to form a single landscape evaluation system. The system assessments revolve around the identification of the content and the spatial organization of landscapes. The *man-made features* and *natural features* components evaluate the content of the landscape scene being considered, and the *spatial organization* and *visual composition* components evaluate the spatial distribution of the landscape (Buhyoff et al., 1994).

EVA uses a rule base that was developed using a graphical decision tree to ensure that the flow and relationship of the rules for each system were properly defined. This decision tree was tested by the authors for logical continuity and soundness using a wide range of photographic images of landscape scenes (Buhyoff et al., 1994).

The four main components of EVA are capable of functioning independently as explanation systems of man-made features, natural features, spatial organization, and visual composition of each landscape scene. The fifth explanation system analyzes the output from each of the other four components and derives a final overall assessment of the landscape visual quality (Buhyoff et al., 1994).

A standard stage in developing any software is validation and verification of the final result and reliability testing of the system (O'Keefe et al., 1988; Rettig,

1991). Validation is concerned with both the accuracy of a measurement and the precision of the measurement. Verification or internal validation refers to testing a system and determining that it correctly implements its specifications (i.e. does the system correctly model the expert on whom it was based?) (O'Keefe et al., 1988; Gilstrap, 1991). Many of the psychophysical and psychological models previously mentioned have been tested for validity and replicability (reliability) over the course of their development (Dearden, 1981; Kaplan, 1985; Brown and Daniel, 1986; Daniel and Vining, 1983; Daniel, 1990; Ribe, 1990; Buhyoff et al., 1994). It is obvious then, that any attempt at creating a new system for evaluating landscape quality must also be tested for validity and reliability.

This research tests the reliability, and validity of EVA's ability to generate visual quality assessments of landscapes. Several methodologial stages are used to test the internal and external validity of the system as well as its reliability.

Chapter 2

Methods

Introduction:

This research was designed to preliminarily test EVA's reliability and validity. Reliability and validity cannot be determined based on a single test. Only repeated testing can provide the support needed to pronounce something valid and reliable (Nunnally, 1978; Guilford, 1954; Pedhazur and Schmelkin, 1991).

O'Keefe et al. (1988) recommended that validation of an expert system be undertaken at some early point in the development of the system. They suggest that the system should have some level of acceptable performance before such testing begins. The authors further point out that the level of acceptable performance will differ for every system being tested, but:

> *. . . with a research prototype, medium performance plus indications that the basic approach is correct may be acceptable.*

EVA existed as a prototype after its initial construction and promising initial reliability evaluations (Zhou, 1992). The next logical step then, was to further test the system for reliability and validity.

Reliability:

Nunnally (1978) states that a measure is reliable if its error is slight. He further states that reliability concerns the extent to which a measure is repeatable by

different persons, in slightly different conditions, which are not intended to influence the results of measurement. Reliability is also defined as the tendency of an instrument to yield the same results when testing identical situations in repeated trials (Carmines and Zeller, 1989). It is further mentioned by Carmines and Zeller that the subsequent measurement trials of a reliable instrument will most likely not yield identical results, but the results will tend to be consistent. Efforts to capture indications of consistency are illustrated in the data analysis section.

By comparing the ability of EVA to assess the visual quality of different landscape scenes at two different times (test-retest), an estimate of EVA's reliability was obtained. Another estimate of EVA's reliability was obtained by comparing EVA's ability to assess different landscape scenes with different operators running EVA.

In order to control bias the literature recommends that another researcher, other than the system developers, perform the validation test (O'Keefe et al., 1988). However, Rettig (1991) states that although the literature recommends an outside researcher perform the testing, it may be more desirable to have original system developers performing the "glass box" testing because they understand system requirements and structure. This study was conducted by a researcher outside of the original design group, but somewhat familiar with internal workings of the system coding. This research was, however, supervised and partially designed by system developers.

Daniel and Vining (1983) suggest that a comparison of the agreement between independent experts (landscape architects) assessing different landscape scenes would be a straightforward approach for determining the reliability of a method of landscape quality assessment. They point out, however, that no such test has yet been performed to any large scale.

Validity:

A measurement can be consistently accurate (reliable), but without precision (validity), the resulting measurement could be far from the actual value being measured. Likewise, an instrument could yield remarkably precise measurements, but have little accuracy (reliability) over time, or across users.

O'Keefe, et. al. (1988) indicate that the validity of an expert system is determined by the degree to which it effectively models the expert on whose decision-making processes it was modeled. By comparing visual quality assessments made by the landscape architect on whose knowledge EVA was based, to those received from EVA with him as EVA's operator, an estimate can be made of EVA's internal validity. That is, does EVA competently model the decision-making processes of the expert on whose knowledge it was based?

External validity refers to the generalizability of findings to or across populations (Pedhazur and Schmelkin, 1991) (i.e. Does EVA actually measure landscape scenic quality). To test EVA's external validity, a comparison of landscape

assessments made by EVA to those made by landscape architects other than the expert on whose knowledge the system was based and a group of lay subjects was performed. If EVA accurately measures landscape visual quality, a high degree of agreement on the visual quality of a series of landscapes should have been present between the different judges.

Test Stimuli:

A study conducted by Shafer and Richards in 1973 found that photographs of outdoor scenes were sufficiently effective in representing the actual scenes while attempting to accurately predict the responses of viewers to the actual outdoor scenes (Schafer and Richards, 1974). The same results appeared in later studies as well (Brown and Daniel, 1986; Daniel and Boster, 1976). Although the validity of these results was questioned by Hull and Stewart (1992), photographic slides of various landscape scenes were used while testing EVA. It would have been impractical to transport subjects to sites for each scene to be evaluated.

Seventy-five 35mm color landscape slides were non-randomly selected from a photographic stimulus library as representative samples of landscape scenes from across the United States. A sample of landscapes was selected which contained natural scenes, scenes where man had impacted natural settings and urban scenes. Both vista-panoramic and on-ground/near-view scenes were included. A variety of landscape scenes were selected in an effort to test the "boundaries" of EVA's effectiveness as a

measuring device of landscape quality. These scenes included in-stand scenes as well as urban, wetlands, coastal, range, and mountainous landscapes.

The landscape color slides were viewed by all subjects in the study on a Telex CARAMATE 4000 slide projector. This projector has a self-contained screen and was placed on the desktop near the computer. Subjects advanced the slides at their own pace and were allowed as much time as needed to complete their evaluations of the landscape scenes.

Stage I: Test of EVA's Internal Validity

The first stage of testing was aimed at determining if the system could replicate the professional judgement of the landscape architect (LA1) on whose knowledge EVA was based. Comparing EVA-generated assessments to the LA1's professional judgements would permit an evaluation of how good a model, EVA is, of the LA1's decision processes relative to landscape visual quality assessment. This comparison provides a preliminary indication of EVA's internal validity. Landscape architects' evaluations of landscape visual quality are the only professional standard for estimating landscape visual quality (Hull, 1989; USFS, 1975).

The seventy-five slides were randomly ordered and LA1 observed and evaluated each of the slides using his best professional judgement for visual quality based on the following scale:

A. Low Visual Quality

B. Moderately Low Visual Quality

C. Moderate Visual Quality

D. Moderately High Visual Quality

E. High Visual Quality

F. Outstanding Visual Quality

This is the identical scale that EVA uses to present its final evaluations of landscape quality. The subject was then asked to use EVA to evaluate the same landscape scenes in the same order. Since EVA is an extensive artificial intelligence system it takes several minutes to complete a single landscape evaluation. It was observed that the subject could generally evaluate the visual quality of a scene, based on his professional standards, in about thirty seconds. However, an EVA generated evaluation for an individual landscape took between five and fifteen minutes. For this reason the evaluations for all seventy-five slides could not be completed in one session. Rather, in order to avoid fatiguing the subject, several sessions of two to three hours each were used to complete the evaluations for all seventy-five landscapes.

Stage II: Test of EVA's Reliability

The reliability of EVA was tested using a "test-retest" method (as suggested by Nunally, 1978) by comparing the final evaluations generated by EVA for the seventy-

five landscape scenes in stage I to the evaluations (using the same seventy-five landscape scenes and the same test subject (LA1)) generated by EVA at a second time.

This test design permitted EVA's reliability (and that of LA1) to be tested. Identical scenes were evaluated under identical conditions at two different times. The reliability of LA1 and the reliability of EVA were tested by comparing the results of the landscape visual quality evaluations at time 1 to the results of the landscape visual quality evaluations at time 2. Agreement between LA1 at time 1 and time 2 would indicate a measure of his reliability. Likewise, a partial measure of EVA's reliability could be determined. Assuming EVA was an effective model of LA1's decision-making processes, the reliability estimate of EVA should be similar to that of LA1.

In an effort to further illustrate the reliability of EVA, an analysis of EVA's internal paths taken by LA1 in the course of his evaluations of the seventy-five slides was performed. The paths taken by LA1 and EVA would be analyzed for differences between times 1 and 2. This analysis would illustrate how consistent LA1 was in the operation of EVA between times 1 and 2. The use of different internal paths with consistent evaluation results between times 1 and 2 would support EVA's reliability.

Stage III: Test of EVA's External Validity

In order to test the external validity of the rules used in EVA, three professional landscape architect volunteers from the Jefferson National Forest, USDA

Forest Service were used as test subjects to provide both professional judgements of some of the same landscape scenes used in stage I.

Each of these subjects evaluated twenty-five randomly selected landscape slides from the original set of seventy-five used in stages I and II. The subjects were asked to provide a professional judgement of landscape visual quality for each landscape slide using the scale from stage I of the testing. Each subject then used EVA to evaluate the same landscapes. There was no pre-training of the subjects in the operation or format of EVA.

Based on the time requirements demanded by the testing, a single two-to-three-hour session was required for each of these subjects. Subjects were allowed to complete the testing at their own pace and advanced the slides themselves, as needed.

If it can be assumed that a landscape architects' evaluations of landscape visual quality are valid in the professional arena, then EVA's evaluations could be compared to an accepted standard.

To further test EVA's reliability, then three landscape architect subjects were asked to use EVA to perform landscape visual quality assessments on twenty-five of the same landscapes used in stage I. These assessments were compared to LA1 (stage I subject) EVA-generated evaluations of the same scenes. High agreement between LA1 and the three landscape architect subjects in this stage would provide further evidence of EVA's reliability.

Stage IV: Further Test of EVA's External Validity

To further test the external validity of EVA's landscape evaluations, two graduate students from the College of Forestry and Wildlife Resources at Virginia Tech volunteered as laysubjects. Twenty-five landscape slides were randomly selected from the seventy-five landscapes in stage I. The lay subjects were asked to evaluate the twenty-five scenes using the scale used in the previous stages. They then used EVA to perform evaluations of the landscape visual quality for the twenty-five scenes.

Time requirements for this portion of the testing were similar to those needed in stage III. As before, no preliminary training regarding EVA's format or operation was given to the lay subjects.

The resulting evaluations (human and EVA) were compared to those of the landscape architect in stage I and his EVA-generated evaluations. If The EVA-generated evaluations in stage I could predict lay subject responses to landscape scenes it would partially validate EVA as a system for evaluating landscape visual quality. In addition, EVA's reliability could further be tested by comparing the EVA-generated landscape evaluations of the lay subjects to those of the landscape architect subject in stage I.

Data Analysis:

The data from the testing were evaluated using both the intra-class correlation coefficient (Guilford, 1954) as a test of reliability and a nonparametric binomial test

for agreement between judges. The intra-class correlation coefficient was calculated using the following formula:

$$\bar{r}_{11} = \frac{V_p - V_e}{V_p + (k-1)\,V_e}$$

where $\bar{r}_{11}$ = the correlation of ratings for a single rater,

V_p = the variance for slide ratings time 1 vs. time 2,

V_e = the variance for error, and

k = the number of raters[1].

Likewise, the binomial χ^2 score was calculated using the following formula:

$$\chi^2 = \frac{(|N_1 - NP| - .5)^2}{NP(1-P)}$$

where N_1 = number of agreements between judges,

N = total number of slides viewed,

P = probability of chance agreement between judges, ... and

.5 = a correction for continuity.

In addition to the χ^2 score, a measure of agreement was used, as suggested by Tinsley and Weiss (1975):

[1] When k = 2 the intra-class correlation coefficient is approximately the same as Pearson's product moment correlation coefficient.

$$T = \frac{N_1 - NP}{N - NP}$$

where N1, N, and P are defined as in equation 2. The value of T is 0 when the observed agreement is equal to chance agreement, and T is 1 when perfect interrater agreement is observed.

Both the landscape architect and EVA-generated landscape evaluations were converted from the scale noted in stage I to a six-point scale, with a score of 1 corresponding to low visual quality and a score of 6 corresponding to outstanding visual quality. Due to unanticipated limitations of EVA's rule base at the time of testing two other numerical score conversions were performed: 2.5 is equivalent to an evaluation of "slightly negative effect on visual quality", while 3.5 compared to an evaluation of "slightly positive effect on visual quality".

High correlations from the intra-class correlation coefficient would indicate high agreement between the "judges" being compared. Likewise, high χ^2 scores would indicate that the agreement between "judges" was not random chance. For the purposes of testing EVA, evaluations were said to be consistent (or in agreement) if the transformed numerical scores from each evaluation were within + or - 1 of each other.

Experimental Design Concerns:

Ideally, it would have been preferable to train several subjects in the use of EVA and determine if the training program would improve or impair their ability to use EVA effectively. This new data would enhance the estimation of EVA's validity, by eliminating any possible confounding between LA1 and EVA.

Unfortunately, no such training process was developed. Implementation of such a procedure would require prior knowledge of potentially confusion of ambiguous terms in EVA, as well as idiosyncrasies of the system. Prior to the testing of the subjects as listed in Stages I - IV, no one outside the original development group had used EVA extensively. For this reason there was no evidence as to where confusing terms existed within EVA, and therefore there was no support for developing a training method that might be beneficial to new users of EVA.

Chapter 3

Results and Discussion

The results which proceed in this chapter are summarized in Table 1.

Stage I

The objective of Stage I of the research was to compare EVA's final results for seventy-five landscape scenes against the professional judgements made by the landscape architect (LA1) upon whom's rules EVA's algorithms are based. Appendix C-1 to C-3 and C-4 to C-6 show the number of disagreements relative to the agreements between EVA and the landscape architect at times 1 and 2 respectively. LA1 and EVA generally disagreed upon the visual quality of urban scenes, denoted by a 'U' in Appendix C. This might indicate a weakness in EVA's ability to effectively evaluate the visual quality of urban landscape scenes. Therefore, for the purposes of this study, statistics were calculated with and without the urban scenes included in the data, in order to provide a clearer picture of how representative EVA is of the landscape architect's decision making processes.

Table 1 : Summary of Statistical Test Results

Judges	chi-square	p	T	r	p <=
LA1-1 vs. LA1's EVA-1	49.1400	<= 0.002	0.736	0.71	0.0000
LA1-1 vs. LA1's EVA-1 *	60.6900	<= 0.002	0.846	0.77	0.0000
LA1-2 vs. LA1's EVA-2	26.5300	<= 0.002	0.544	0.79	0.0000
LA1-2 vs. LA1's EVA-2 *	31.6500	<= 0.002	0.614	0.83	0.0000
LA1's EVA-1 vs. LA1's EVA-2	39.8500	<= 0.002	0.664	0.63	0.0000
LA1's EVA-1 vs. LA1's EVA-2 *	34.4100	<= 0.002	0.640	0.63	0.0000
LA1-1 vs. LA1-2	66.7820	<= 0.002	66.782	0.83	0.0000
LA2 vs. LA1-1	24.8645	<= 0.002	0.928	0.81	0.0000
LA2 vs. LA2's EVA	6.6125	0.02 < p < 0.05	0.496	0.53	0.0001
LA2 vs. LA1's EVA-1	8.8445	0.002 < p < 0.01	0.568	0.71	0.0000
LA2 vs. LA1-2	17.4845	<= 0.002	0.784	0.71	0.0000
LA2 vs. LA1's EVA-2	6.6125	0.02 < p < 0.05	0.496	0.49	0.0005
LA2's EVA vs. LA1's EVA-1	6.6125	0.02 < p < 0.05	0.496	0.70	0.0000
LA2's EVA vs. LA1's EVA-2	0.9245	> 0.20	N/A	0.13	0.2615
LA2's EVA vs. LA1-1	3.1205	0.2 > p > 0.1	N/A	0.74	0.0000
LA2's EVA vs. LA1-2	0.3125	> 0.20	N/A	0.50	0.0003
LA3 vs. LA1-1	8.8445	0.002 < p < 0.01	0.568	0.73	0.0000
LA3 vs. LA3's EVA	0.3125	> 0.20	N/A	0.46	0.0012
LA3 vs. LA1's EVA-1	0.6050	> 0.20	N/A	0.63	0.0000
LA3 vs. LA1-2	17.4845	<= 0.002	0.784	0.84	0.0000
LA3 vs. LA1's EVA-2	0.0245	> 0.20	N/A	0.68	0.0000
LA3's EVA vs. LA1's EVA-1	0.0245	> 0.20	N/A	0.60	0.0000
LA3's EVA vs. LA1's EVA-2	0.3125	> 0.20	N/A	0.55	0.0001
LA3's EVA vs. LA1-1	0.9245	> 0.20	N/A	0.60	0.0000
LA3's EVA vs. LA1-2	0.3125	> 0.20	N/A	0.66	0.0000
LA4 vs. LA1-1	11.4010	<= 0.002	0.640	0.76	0.0000
LA4 vs. LA4's EVA	1.8605	> 0.20	N/A	0.48	0.0007
LA4 vs. LA1's EVA-1	0.0245	> 0.20	N/A	0.80	0.0000
LA4 vs. LA1-2	14.2805	<= 0.002	0.712	0.76	0.0000
LA4 vs. LA1's EVA-2	0.9245	> 0.20	N/A	0.75	0.0000
LA4's EVA vs. LA1's EVA-1	1.1045	> 0.20	N/A	0.36	0.0133
LA4's EVA vs. LA1's EVA-2	0.4205	> 0.20	N/A	0.19	0.1614
LA4's EVA vs. LA1-1	0.0245	> 0.20	N/A	0.46	0.0004
LA4's EVA vs. LA1-2	0.3125	> 0.20	N/A	0.34	0.0198
LAY1 vs. LA1-1	21.0125	<= 0.002	0.856	0.76	0.0000
LAY1 vs. LAY1's EVA	1.8605	> 0.20	N/A	0.52	0.0002
LAY1 vs. LA1's EVA-1	14.2805	<= 0.002	0.712	0.45	0.0002
LAY1 vs. LA1-2	17.4845	<= 0.002	0.784	0.75	0.0000
LAY1 vs. LA1's EVA-2	3.1205	0.2 > p > 0.1	N/A	0.49	0.0005
LAY1's EVA vs. LA1's EVA-1	0.9245	> 0.20	N/A	0.33	0.0236
LAY1's EVA vs. LA1's EVA-2	3.1205	0.2 > p > 0.1	N/A	0.35	0.0161
LAY1's EVA vs. LA1-1	1.8605	> 0.20	N/A	0.54	0.0001
LAY1's EVA vs. LA1-2	0.9245	> 0.20	N/A	0.50	0.0003
LAY2 vs. LA1-1	3.1205	0.2 > p > 0.1	N/A	0.42	0.0034
LAY2 vs. LAY1's EVA	0.9053	> 0.20	N/A	0.19	0.1614
LAY2 vs. LA1's EVA-1	0.0245	> 0.20	N/A	0.23	0.1044
LAY2 vs. LA1-2	11.4005	<= 0.002	0.640	0.62	0.0000
LAY2 vs. LA1's EVA-2	4.7045	0.1 > p > 0.05	0.424	0.49	0.0005
LAY2's EVA vs. LA1's EVA-1	0.2625	> 0.20	N/A	0.23	0.1310
LAY2's EVA vs. LA1's EVA-2	0.9053	> 0.20	N/A	0.24	0.1181
LAY2's EVA vs. LA1-1	0.2625	> 0.20	N/A	0.29	0.0651
LAY2's EVA vs. LA1-2	0.9053	> 0.20	N/A	0.35	0.0274

At time 1 (urban scenes included) EVA and LA1's evaluations tended to agree (χ^2=49.14, $p \leq 0.002$; T = 0.736[2]; r = 0.71, $p \leq 0.00001$). At time 2 this evidence of agreement was supported, but not as strongly (χ^2=26.53, $p \leq 0.002$; T = 0.544; r = 0.79, $p \leq 0.00001$). The intraclass correlation results do increase between times 1 and 2, but this is not intuitively consistent because the number of agreements between EVA and LA1 decreases.

With the urban scenes removed the agreement between LA1 and EVA appeared to improve (Time 1: χ^2=60.69, $p \leq 0.002$; T = 0.846; r = 0.77, $p \leq 0.00001$. Time 2: χ^2=31.65, $p \leq 0.002$; T = 0.614; r = 0.83, $p \leq 0.00001$).

Given these statistical results it would appear then that EVA does replicate the landscape architect subject's decisions to a good degree. This indicates that EVA has a strong foundation upon which it was constructed and the rules used by EVA seem to have substantial internal validity. It is important to realize, however, that EVA relies on human observations of landscape scenes and responses to the questions EVA asks to arrive at an evaluation of the visual quality of a landscape. Human errors, such as failing to notice a particular landscape feature, could influence the results of this test. Furthermore, with untrained users operating EVA these errors could become much worse. For these reasons these measures of agreement may be clouded by subtle errors. Since this was a test-retest design, the personal development and attitudes of

[2]Note: P-values were not calculated for the T statistic because it relates only the degree of agreement between judges.

LA1 could not be held constant, thereby introducing an unmeasurable bias into the results of this study.

Another potential source of error in this study is the possible confounding of EVA's evaluations while being used by LA1. It is possible that LA1 could influence EVA's results by altering his inputs in response to EVA's questions, so that EVA's final results would agree with his evaluations of the landscape being viewed. Since EVA was designed to be a model of LA1's decision-making processes this is entirely possible, but due to the complexity of EVA it is not likely.

The decision tree EVA uses to arrive at its evaluations of a landscape scene contains 367 possible paths which can be traversed by answering the questions posed to the user of EVA. The responses EVA uses to arrive at an evaluation of landscape visual quality are shrouded in the many questions the user is asked about the landscape scene under scrutiny. Each module in EVA has a separate number of questions that are asked, and some questions have precedence over others when EVA evaluates landscape visual quality.

Although LA1 was responsible for creating the decision tree EVA uses, he was not responsible for the programming of EVA. EVA's programming is such that all questions are asked, even though there is a hierarchy to them. Certain responses to some questions will supercede the responses to others. As a result, questions are frequently asked that have no actual bearing on the final decision path taken by EVA. When this factor is taken into consideration along with the number of landscape scenes

evaluated by EVA and LA1, and the amount of time necessary for the testing, the probablility of any confounding between EVA and LA1 appears to be fairly low.

Stage II

The objective of Stage II of this research was to compare the reliability of EVA to that of the landscape architect subject (LA1) from Stage I. Appendix C-7 to C-9 and C-10 to C-12 represent the agreement between LA1 at times 1 and 2 and the agreement between EVA at times 1 and 2, respectively. This was accomplished by comparing the consistency of LA1's landscape evaluations at times 1 and 2 to obtain an estimate of his reliability. Likewise, the constancy of EVA's landscape evaluations at times 1 and 2 were compared to obtain a reliability estimate for EVA.

LA1 was highly consistent between times 1 and 2, indicating a high degree of reliability (χ^2=66.782, $p \leq 0.002$; T = 0.856; r = 0.83, $p \leq 0.00001$). EVA was not as consistent as LA1 (χ^2=39.85, $p \leq 0.002$; T = 0.664; r = 0.63, $p \leq 0.00001$). Once again, however, these reliability estimates may be affected by errors made in using the system.

Table 2 lists the evaluations of landscape visual quality obtained from EVA and the landscape architect subject at times 1 and 2. Also this table indicates whether or not the landscape architect subject (LA1) responded consistently to EVA's questions (i.e. took the same path through EVA) about the landscapes at times 1 and 2. A '0' in the "path" column indicates inconsistent responses, while a '1' indicates consistency.

Table 2: Path Analysis of LA1 and EVA Times 1 and 2.

Slide	Evaluation	Module	Path	EVA
1	M/H	3,4	0	H
1	H	3,4		H
2	H	2	0	O
2	O	2		O
3	M/L	3,4	0	M
3	M/L	3,4		M
4	M	3,4	0	M
4	M/H	2		M/H
5	M/H	3,4	0	H
5	M	3,4		H
6	M/L	3,4	0	M
6	M/H	3,4		M/H
7	O	2	0	O
7	O	2		H
8	M/H	3,4	0	M
8	L	3,4		M
9	M/L	3,4	0	M/L
9	L	3,4		St P
10	M/H	2	0	O
10	H	2		H
11	H	3,4	0	H
11	O	3,4		H
12	M/H	2	0	H
12	M/H	2		O
13	M/L	3,4	0	M/L
13	L	3,4		L
14	H	1	0	H
14	H	1		M/H
15	H	2	0	O
15	H	2		O
16	L	1	1	L
16	L	1		L
17	M	1	0	H
17	M	3,4		H
18	M/L	1	0	H
18	L	3,4		H
19	M/L	3,4	0	M
19	M/L	3,4		M/H
20	O	2	0	O
20	O	2		O
21	M/L	3,4	0	M/L
21	M/L	3,4		M
22	O	2	0	O
22	H	2		O
23	M/H	3,4	0	H
23	M/H	3,4		H
24	H	2	0	O
24	H	2		O
25	O	2	1	O
25	H	2		O

Slide	Evaluation	Module	Path	EVA
26	M/H	3,4	0	M/H
26	H	3,4		H
27	M/L	3,4	0	M/L
27	M	3,4		H
28	M/L	3,4	0	L
28	M/L	3,4		M/L
29	H	2	0	O
29	M/H	2		H
30	H	2	0	H
30	H	3,4		H
31	H	2	0	H
31	H	2		H
32	M/L	3,4	0	M
32	M/L	3,4		M/L
33	O	2	0	O
33	O	2		O
34	M/H	3,4	0	L
34	M/H	3,4		H
35	M/H	2	0	M/H
35	M/H	2		H
36	M/L	3,4	0	M
36	L	3,4		L
37	O	2	0	O
37	O	2		O
38	M/H	3,4	0	M/H
38	M/L	3,4		M
39	H	2	1	O
39	O	2		O
40	H	3,4	0	H
40	H	3,4		H
41	L	3,4	1	M
41	L	3,4		M
42	O	2	0	M/L
42	O	2		O
43	H	2	0	M/H
43	H	2		H
44	M/L	3,4	0	L
44	M	1		M/H
45	L	3,4	0	L
45	L	3,4		M/L
46	M	3,4	0	St Pos.
46	M/L	3,4		St Pos.
47	M/H	2	0	O
47	M/H	1		M/H
48	H	3,4	0	H
48	H	3,4		H
49	L	3,4	0	M
49	M/L	1		St. Neg.
50	M/H	2	0	H
50	M	3,4		H

Slide	Evaluation	Module	Path	EVA
51	M/L	3,4	0	M
51	L	3,4		L
52	M/H	3,4	0	H
52	M/H	3,4		H
53	H	2	0	O
53	O	2		O
54	M/H	3,4	0	H
54	M/H	3,4		M/L
55	M/L	3,4	0	H
55	M/L	3,4		H
56	O	2	0	O
56	O	2		O
57	M/H	3,4	0	M
57	M/L	3,4		M
58	H	2	1	O
58	O	2		O
59	M/L	3,4	0	H
59	M/H	3,4		H
60	L	1	0	L
60	M/L	1		M/L
61	M/L	3,4	0	M/L
61	M/L	3,4		M
62	M/L	3,4	0	H
62	M	1		M/H
63	O	2	0	O
63	O	2		O
64	M/H	2	0	M/H
64	M/H	2		O
65	M/L	3,4	0	M
65	L	3,4		M
66	H	2	1	O
66	O	2		O
67	M	3,4	0	M/L
67	M/H	2		H
68	L	3,4	0	M/L
68	L	3,4		L
69	M	3,4	0	M/H
69	M/L	3,4		L
70	M/H	3,4	0	St Po
70	M/H	3,4		St Po
71	M/H	3,4	0	H
71	M	3,4		H
72	M/H	3,4	0	H
72	M	3,4		H
73	H	2	0	M/H
73	M/H	3,4		M/L
74	M/H	3,4	0	H
74	M	3,4		H
75	M/H	3,4	0	M/H
75	L	3,4		L

The majority of the consistent responses occurred in module 2 -- natural features (the simplest EVA module). However, it can be clearly seen that inconsistent responses were much more frequently the case, but EVA's evaluations remained in agreement from time 1 to time 2. In sixty-two of the seventy-five landscape scenes EVA did yield the same evaluation (or did agree as per the definition of agreement used in this study) of the landscape visual quality between times 1 and 2, while LA1 accomplished the same feat in sixty-nine of the seventy-five scenes.

This could be a preliminary indication that EVA is a fairly robust system, and that small errors made by human operators do not overly affect the final evaluations of landscape visual quality. However, the 'paths' column of Table 1 indicates **any** differences between the recorded responses of LA1 at times 1 and 2. This may not reflect thetrue decision path EVA used to arrive at a landscape evaluation. As stated in the discussion of Stage I, EVA's questions have a definite hierarchy in their "importance" to the decision-making process. It is quite possible that the differences noted in the paths taken at times 1 and 2 have no actual bearing on EVA's evaluations of landscape visual quality. Further investigation of these data while considering which questions have precedence over others would shed more light on this.

Stage III

The objectives of Stage III of this research were to test the external validity of the rules used in EVA's algorithms and further test EVA's reliability. EVA's external

validity was tested by comparing the professional evaluations of the visual quality of twenty-five landscape scenes made by three professional landscape architect subjects (LA2, LA3, and LA4) to EVA's evaluations of the same landscape scenes. Appendix C-13 to C-15, C-16 to C-18, and C-19 to C-21 illustrate the agreements and disagreements of LA2, LA3, and LA4's professional evaluations of landscape visual quality and their evaluations of visual quality using EVA, respectively. It is obvious that the subjects did not agree with EVA as often as LA1 did in Stage II.

LA2's professional evaluations of the visual quality of the slides did not generally appear to agree with the evaluations given by EVA under the operation of LA2 (χ^2=6.6125, $0.05 \geq p \geq 0.02$; T = 0.496; r = 0.53, $p \leq 0.0001$). This lack of agreement may be due to a misunderstanding of the terms used by EVA when asking the user questions about the landscape being evaluated. LA2 asked several times for an explanation of specific terms used by EVA. Due to the nature of this test, however, those definitions were not provided.

Likewise, LA3's professional evaluations often differed with the EVA-generated evaluations (χ^2=0.0605[3], $p \geq 0.1$; r = 0.46, $p \leq 0.0012$). Here again, a lack of understanding of the terms EVA uses may have played an important role in this lack of agreement. Like LA2, LA3 asked for definitions of some of the terms used by EVA.

[3]Note: T not calculated due to an insignificant χ^2 (Tinsley and Weiss, 1975).

There was little agreement between EVA and LA4 (χ^2=0.0245, $p \geq 0.1$; $r = 0.48$, $p \leq 0.0007$). Once again, the definitions of some of the terms used by EVA presented a problem for the subject.

The failure of the subjects in this stage of testing to agree with EVA may be due to one of two possible reasons. First, the rules used by EVA may not be externally valid and only one person (LA1) may be able to use it effectively. Second, EVA's rules may be valid, but a training session for the use of EVA may need to be constructed. The subjects received no training in the use of EVA and some of the terms used within EVA were not adequately defined, judging from the subjects' requests for definitions.

A training program for the use of EVA was not developed as part of this research. Development of such a procedure would require prior knowledge of potentially confusing or ambiguous terms in EVA, as well as idiosyncrasies of the system. Prior to the testing of the subjects in this research, no one outside the original development group had used EVA extensively. For this reason there was no evidence as to where confusing terms existed within EVA, and therefore there was no support for developing a training method that might be beneficial to new users of EVA.

To further test EVA's reliability, the "professional judgement" evaluations made by LA2, LA3, and LA4 were compared to those produced by EVA with LA1 as the user at times 1 and 2.

LA2 agreed the most strongly with the EVA evaluations at time 1 (χ^2= 8.8445, $0.002 \leq p \leq 0.01$; T = 0.568; r = 0.71, $p \leq 0.00001$). When comparing LA2's evaluations to LA1's EVA-generated evaluations at time 2, the number of disagreements increased slightly, but the agreement was fairly consistent with time 1 (χ^2= 6.6125, $0.02 \leq p \leq 0.05$; T = 0.496; r = 0.49, $p \leq 0.0005$). Appendix C-22 to C-24 & C-25 to C-27 illustrate these data.

In addition, LA2 agreed the most strongly with LA1's personal evaluations of the landscape scenes (Appendix C-28 to C-30 and C-31 to C-33). At time 1 the agreement was almost perfect between the two LA subjects (χ^2= 24.8645, $p \leq 0.002$; T = 0.928; r = 0.81, $p \leq 0.00001$). At time 2 the agreement was still high, but not as highly correlated (χ^2= 17.4845, $p \leq 0.002$; T = 0.784; r = 0.71, $p \leq 0.00001$).

LA3's professional judgements did not strongly agree with the EVA-generated evaluations from LA1 at times 1 (χ^2= 0.0605, $0.1 \leq p \leq 0.9$; r = 0.63, $p \leq 0.00001$) and 2 (χ^2= 0.0245, $0.1 < p < 0.9$; r = 0.68, $p \leq 0.00001$). Here again, the intra-class correlation estimate is not intuitively consistent. The number of disagreements indicate that reliability should be low. Appendix C-34 to C-36 & C-37 to C-39 illustrate these data.

When comparing LA3's personal evaluations to LA1's personal evaluations of the landscape scenes at time 1 (Appendix C-40 to C-42), the agreement is higher (χ^2= 8.8445, $0.002 \leq p \leq 0.01$; T = 0.568; r = 0.73, $p \leq 0.00001$). At time 2 (Appendix C-

43 to C-45) the degree of agreement is even higher still (χ^2= 17.4845, $p \leq 0.002$; T = 0.784; r = 0.84, $p \leq 0.00001$).

LA4 also did not agree with EVA-generated assessments (LA1 as user). At time 1 (Appendix C-46 to C-48) (χ^2= 0.0245, $0.1 \leq p \leq 0.9$; r = 0.80, $p \leq 0.00001$) the agreement is just as poor as the agreement observed in the results of LA3. At time 2 (Appendix C-49 to C-51) this agreement improves, but it is still not as high as the agreement observed of LA2 (χ^2= 0.9245, $0.1 \leq p \leq 0.9$; r = 0.76, $p \leq 0.00001$).

When LA4 and LA1 (time 1) are compared, the results are consistent with the previous comparisons of the personal evaluations made by the LA subjects (χ^2= 11.401, $p \leq 0.002$; T = 0.64; r = 0.76, $p \leq 0.00001$). Comparing LA4 and LA1 (time 2) resulted in a similar amount of agreement (χ^2= 14.2805, $p \leq 0.002$; T = 0.712; r = 0.76, $p \leq 0.00001$). These data are detailed in Appendix C-52 to C-54 and C-55 to C-57.

If LA1's EVA-generated evaluations are compared to the personal evaluations of LA2, LA3, and LA4, it would seem to indicate that EVA does not have good reliability across users. This might not be the case, however, if the results of comparing the pesonal evaluations of LA1 at times 1 and 2 to those of LA2, LA3, and LA4 are examined. Note that LA2 agreed with LA1's personal evaluations of landscape visual quality to a much higher degree than LA3 and LA4. Likewise, LA2's agreement with LA1's EVA-generated results is much higher than that of LA3 and LA4. This would seem to indicate one of two things. First, that LA2 either had a

better understanding of the terms used within EVA than LA3 and LA4. Or second, LA2's decision-making processes are more closely related to those of LA1 than LA3 and LA4 are. This apparent similarity between the decison-making processes of LA1 and LA2 may give support to EVA's reliability.

Stage IV:

The objectives of Stage IV of this research were to further test EVA's external validity by comparing the EVA-generated evaluations of LA1 at times 1 and 2 to the user preference evaluations of two graduate student lay subjects from the College of Forestry and Wildlife Resources at Virginia Tech. Appendix C-58 to C-60, C-64 to C-66 and C-61 to C-63, C-67 to C-69 represent the degree of agreement between Lay subjects 1 and 2 and LA1's EVA-generated evaluations at times 1 and 2. Also, a further estimate of EVA's reliability was obtained by comparing both Lay subjects' EVA-generated evaluations of the landscape slides to LA1's EVA-generated evaluations at times 1 and 2.

Lay subject 1 generally agreed with LA1's EVA-generated evaluations at time 1 (χ^2= 14.2805, $p \leq 0.002$; T = 0.712; r = 0.45, $p \leq 0.0002$). Likewise, Lay subject 1 agreed with LA1's EVA evaluations at time 2, though not as strongly (χ^2= 3.1205, $0.2 \geq p \geq 0.1$; r = 0.49, $p \leq 0.0005$).

Lay subject 2 did not agree with LA1's EVA evaluations at time 1 (χ^2= 0.0245, 0.9 $\geq$ p; r = 0.23, p $\leq$ 0.1044). At time 2, however, the agreement increased substantially (χ^2= 4.7045, 0.1 $\geq$ p $\geq$ 0.05; r = 0.49, p $\leq$ 0.0005).

These preliminary results indicate that EVA, when in the control of an experienced user (LA1), may have the ability to predict the visual quality responses of a Lay subject (in the case of Lay subject 1), but not always (in the case of Lay subject 2 vs. LA1's EVA evaluations at time 1). Comparing Lay subject 1's & 2's personal judgement evaluations to those of LA1 might serve to explain this result.

Lay subject 1's personal judgement evaluations agreed highly with LA1's professional judgement evaluations at time 1 (Appendix C-70 to C-72) (χ^2= 21.0125, p $\leq$ 0.002; T = 0.856; r = 0.76, p $\leq$ 0.00001). Lay subject 1 also agreed with LA1's professional evaluations at time 2 (Appendix C-73 to C-75) , though not to the same degree (χ^2= 17.4845, p $\leq$ 0.002; T = 0.784; r = 0.75, p $\leq$ 0.00001).

Lay subject 2 did not generally agree with LA1's professional evaluations from time 1 (Appendix C-76 to C-78) (χ^2= 3.1205, 0.2 $\geq$ p $\geq$ 0.1; r = 0.42, p $\leq$ 0.0034). At time 2, however, the agreement between the two judges (Appendix C-79 to C-81) increased dramatically (χ^2= 11.4005, p $\leq$ 0.002; T = 0.64; r = 0.62, p $\leq$ 0.00001).

These results demonstrate that perhaps the definition of agreement used in this study may not have been broad enough. LA1 demonstrated agreement with himself at times 1 and 2 in Stage II, and there is agreement indicated between Lay subject 2 and LA1 only at time 2 in this stage of the study. In this study, agreement was defined as

being present when a transformed numerical score was within + or - one of another score. If this definition of agreement was expanded to + or - two, instead of the original agreement range, these results would most likely become more consistent with one another, but the test would not be as accurate in assessing EVA's sensitivity.

To further test EVA's reliability, a comparison of Lay subject 1's & 2's EVA-generated evaluations was compared to the EVA-generated evaluations of LA1. Lay subject 1's EVA evaluations did not agree with those of LA1 at times 1 and 2 (Appendix C-82 to C-87) (Time 1: $\chi^2 = 0.9245$, $0.9 \geq p \geq 0.1$; $r = 0.33$, $p \leq 0.0236$; Time 2: $\chi^2 = 3.1205$, $0.2 \geq p \geq 0.1$; $r = 0.35$, $p \leq 0.0161$).

Likewise, Lay subject 2's EVA evaluations did not agree with those of LA1 at times 1 and 2 (Appendix C-88 to C-93) (Time 1: $\chi^2 = 0.2625$, $0.9 \geq p \geq 0.1$; $r = 0.23$, $p \leq 0.1310$; Time 2: $\chi^2 = 0.9053$, $0.9 \geq p \geq 0.1$; $r = 0.24$, $p \leq 0.1181$).

Once again, as in Stage III of the testing, this lack of agreement may be due to the inexperience of lay users -- particularly in understanding some of the concepts used as the basis of many EVA rules. No training in the use of EVA was provided to Lay subjects 1 and 2, and although they asked for clarification of some of EVA's questions, none was provided.

Chapter 4

Conclusions

Based on the results obtained in Chapter 3, it is obvious that EVA requires some improvement and retesting before we can be confident in its validity and reliability.

The estimates of EVA's internal validity in Stage I indicate that EVA (as a prototype) is a surprisingly effective model of LA1's decision making processes. Based on the results from this stage of testing and the figures depicting the agreements of EVA and LA1, it appears that EVA is not capable of effectively evaluating the landscape visual quality of urban scenes. The addition of several new rules for dealing with urban scenes might improve this condition, but in its current state such scenes should not be evaluated using EVA.

The reliability test in Stage II demonstrates that the LA1 ($T = 0.856$; $r = 0.832$) is more reliabile than EVA ($T = 0.664$; $r = 0.662$), but considering the fact that EVA is a prototype system, this estimate of its reliability indicates promising performance. Nunnally (1978) states that in early stages of research one saves time by working with instruments of modest reliability, which corresponds to a reliability estimate of 0.7 or higher. As the research progresses this standard can be increased.

Feimer et. al. (1981) were unsuccessful in obtaining a reliability estimate greater than 0.7 when using individual reliability estimates received from the intraclass

correlation coefficient. The average individual reliability estimates they report for scenic beauty were around 0.18 to 0.20, though they mention that increasing the number of judges and calculating a composite score will improve the reliability estimates.

Hull and Buhyoff (1984) found reliabilities of 0.767 to 0.799 when they asked subjects to order eleven photographs in order of scenic quality. This was accomplished in a test-retest scenario with a year between trials. They reported these reliabilities as quite good, but perhaps not suitable for long-term planning.

These two studies were performed in different manners, however. Hull and Buhyoff allowed subjects to sort the photographs according to visual quality, while Feimer et. al. had the subjects rate the photos according to a pre-determined scale. It is possible that the difference in the type of rating technique used in these two studies may explain the differences in the reliabilities obtained. In comparison, EVA performs well above the reliability estimates experieced by Feimer et. al. The rating format used by the subjects is most similar between these studies, and if there is some effect due to the choice of rating method, it is appropriate to compare similar studies.

The subsequent reliability tests in Stages III and IV indicate that EVA's reliability for different users is relatively poor. This is due, however, to the fact that EVA relies on human responses to the questions it poses to arrive at the landscape quality evaluations it generates. For this reason, users without the same knowledge of EVA as LA1 cannot hope to generate reliable estimates of landscape visual quality

using EVA. Furthermore, the best estimate of EVA's reliability from this research is the testing in Stage II, since LA1 is the most familiar with the constructs used in EVA, and therefore the most standardized of all possible users of EVA. However, the possibility of improving the reliability estimates in Stages III and IV by developing a training method for new users of EVA is obvious. If new users could be trained to have the same understanding of EVA as LA1 the reliability estimates obtained from their use of EVA should increase to levels near those of LA1. For this reason, a training program for EVA should be developed and EVA's reliability should be retested.

Likewise, the external validity tests in Stages III and IV indicate that EVA would perhaps become more valid if further explanations of problematic definitions were given to new users. To aid in the construction of a training system for EVA, these problematic areas could be identified by analyzing the recorded paths taken by LA1,2,3, and 4, and determining where the paths diverged from one another.

After the development of a training program for EVA the external validity should be retested with trained and untrained users to determine if the training has any discernible effect on the validity estimates. This would not only serve to determine if EVA was improved by the training session, it would allow the researcher to ascertain whether or not LA1 is the only person able to effectively use EVA to evaluate landscape visual quality. Since EVA was designed to model LA1's intuitive processes it is possible that he was able to influence the outcome of the EVA-generated

evaluations obtained from his trials using EVA. This issue was not entirely resolved in this study because there was no pre-existing evidence to indicate how a training procedure should be developed. As a result, there was no way to statistically remove the potentially confounding effects between LA1 and EVA.

It is important to realize that there can be no single absolute answer as to whether an expert system is valid or not. We can not expect perfect performance for an expert system since it is an attempt to model a human expert. Human experts are capable of errors, therefore, the system may make errors as well. In addition, a range of acceptable results from the system must be determined. Gilstrap (1991) states that it is not reasonable to require 100% success for all the tests a system is given, because a human expert will not be correct 100% of the time. It is therefore necessary to determine an acceptable percentage of "pass/fail" results to evaluate system performance and determine the validity of the system.

To further test EVA's validity, EVA's paths and resulting evaluations (obtained from the different landscape architects in this study) should be analyzed and compared for effectiveness relative to other methods of landscape visual assessment for the landscape scenes in this study. In this manner EVA's convergent validity (the validity of an instrument compared to another instrument) could be determined.

Future research should also attempt to determine the robustness of EVA. In this study Table 2 listed whether or not the paths taken at two different times (under the operation of LA1) were identical. It was obvious that the same path was only

followed rarely. However, as stated in the results, the actual path used to obtain an evaluation may only rely upon two or three responses to EVA's questions, and therefore the path analysis undertaken in this study is superficial, at best. The flow charts from which EVA's rules bases were constructed should be compared to the recorded paths, and the path analysis should be redone to obtain a more accurate estimate of EVA's robustness.

In conclusion, EVA does appear to be a promising system for landscape visual quality assessment. With further work it may one day prove to be as valid as a professional landscape architect.

References

Brown, T. C. and T. C. Daniel. 1986. Predicting scenic beauty of timber stands. *Forest Science* 32(2): 471-487.

Buhyoff, G. J., J. D. Wellman, H. Harvey, and R. A. Fraser. 1978. Landscape Architects' Interpretations of People's Landscape Preferences. *Journal of Environmental Management*. 6: 255-262.

Buhyoff, G. J., and L. G. Fuller. 1993. Explanation of statistical models: A need and an example. *Artificial Intelligence Applications*. 7(1):37-43.

Buhyoff, G. J., P. A. Miller, J. W. Roach, D. Zhou, and L. G. Fuller. (1994). EVA: Explanation Visual Assessment. *Artificial Intelligence Applications*.

Carmines, Edward G., and R. A. Zeller. 1989. *Reliability and Validity Assessment*. Sage Publications. London. 71pp.

Coulson R. N., L. J. Folse, D. K. Loh. 1987. Artificial intelligence and natural resource management. *Science*. 237: 262-267.

Craik, K. H. 1975. Individual Variations in Landscape Description. *Landscape Assessment*. (E. H. Zube, R. O. Brush, J. G. Fabas eds.). Stroudsburg, Dowden, Hutchinson and Ross. pp. 130-150.

Daniel, T. C., and R. S. Boster. 1976. *Measuring landscape aesthetics: The scenic beauty estimation method.* (U.S.D.A. Forest Service Research Paper 167). Ft. Collins, CO. Rocky Mountain Forest and Range Experiment Station.

Daniel, T. C. and J. Vining. 1983. Methodological issues in the assessment of landscape quality. In *Human behavior and environment*. (I. Altman and J. Wohlwill eds.). Vol. 4. pp. 39-84. New York: Plenum.

Daniel, T. C. 1990. Measuring the quality of the natural environment: A psychophysical approach. *American Psychologist*. 45(5): 633-637.

Dearden, Philip. 1981. Landscape Evaluation: The Case for a Multi-dimensional Approach. *Journal of Environmental Management*. 13: 95-105

Feimer, Nickolaus R., R. C. Smardon, and K. H. Craik. 1981. Evaluating the effectiveness of observer based visual resiource and impact assessment methods. *Landscape Research*. 6(1): 12-16.

Gilstrap, Lewey. 1991. Validation and verification of expert systems. *Telematics and Informatics*. 8(4): 439-448.

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

Hayes-Roth, F. 1987. Expert Systems. in *Encyclopedia of Artificial Intelligence*. Shapiro, S.C. (ed.). pp. 287-298.

Hull, R. B. IV, and G. J. Buhyoff 1984. Individual and group reliability of landscape assessments. *Landscape Planning*. 11: 67-71.

Hull, R. B. IV. 1989. Forest visual quality management and research. *U.S.D.A. Forest Service General Technical Report,* No. Se-52. Outdoor Recreation Benchmark 1988: Proceedings From the National Outdoor Recreation Forum. U.S.D.A. Forest Service. U.S. Government Printing Office, Washington, D.C. pp. 485-498.

Hull, R. B. IV. and W. P. Stewart. 1992. Validity Of Photo-Based Scenic Beauty Judgements. *Journal of Environmental Psychology*. 12: 101-114.

Kaplan, R. 1985. The analysis of perception via preference: A strategy for studying how the environment is experienced. *Landscape Planning*. 12: 161-176.

Kaplan, S. and R. Kaplan. 1982. *Cognition and Environment: Functioning in an Uncertain World.* Praeger Publishers, New York, New York. 287 pp.

Kaplan, R., S. Kaplan, and T. Brown. 1989. Environmental preference a comparison of four domains of predictors. *Environment and Behavior*. 21(5): 509-530.

Nunnally, Jum C. 1978. *Psychometric Theory*. McGraw-Hill. New York, New York. 701 pp.

O'Keefe, Robert M., O. Balci, and E. P. Smith. 1988. Validating Expert System Performance. *AI Applications*. 2: 35-43.

O'Riordan, Timothy. 1971. *Perspectives on Resource Management*. Pion Press. London. p.71

Pedhazur, Elazar J., and L. P. Schmelkin. 1991. *Measurement, Design, and Analysis: An Integrated Approach*. Lawrence Erlbaum Associates. Hillsdale, New Jersey. 891 pp.

Rettig, M. 1991. Testing made palatable. *Communications of the ACM*. 34(5): 25-29.

Ribe, R. G. 1990. A general model for understanding the perception of scenic beauty in northern hardwood forests. *Landscape Journal*. 9(2): 86-101.

Roach, J. W., and J. Deighan. 1990. *VPI Prolog User's Manual.* Virginia Polytechnic Institute and State University, Department of Computer Science, Blacksburg, VA.

Ruddell, E. J., J. H. Gramann, V. A. Rudis, and J. M. Westphal. 1989. The psychological utility of visual penetration in near-view forest scenic beauty models. *Environment and Behavior.* 21(4): 393-412.

Schafer, E. L. Jr., and T. A. Richards. 1974. *A Comparison of Viewer Reactions to Outdoor Scenes and Photographs of Those Scenes.* USDA Forest Service Research Paper NE-302. 26 p.

Smardon, R. C. 1986. Historical evolution of visual resource management within three federal agencies. *Journal of Environmental Management.* 22: 301-317.

Stock, M. 1987. Ai and Expert Systems: An Overview. *AI Applications in Natural Resource Management.* 1(1): 9-17.

Swartout, W. R. 1987. Explanation. in *Encyclopedia of Artificial Intelligence.* S. C. Shapiro (ed.). pp. 298-300.

Tinsley, Howard E. A., and D. J. Weiss. 1975. Interrater reliability and agreement of subjective judgements. *Journal of Counseling Psychology.* (22)4: 358-376.

U. S. Congress. 1960. *The Multiple Use and Sustained Yield Act.* Washington D. C.

U. S. Congress. 1969. *The National Environmental Policy Act.* Washington D. C.

USDA Forest Service. 1975. *National Forest Landscape Management System. Vol. 2. Ch. 1. The Visual Management System. Agricultural Handbook 462.* 47 pp.

Zhou, D. 1992. Artificial intelligence based visual assessment: An explanation system for landscape aesthetics. M. S. Thesis. Department of Forestry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 24061. 278 pp.

Zube, E. H. and D. G. Pitt. 1981. Cross Cultural Perceptions of Scenic and Heritage Landscapes. *Landscape Planning*. 8: 69-87.

Appendix A

	LA1 Time 1			LA1 Time 2			LA2			LA3			LA4			Laysubject 1			Laysubject 2		
Slide		Rat.	Prog.	Order	Rat.	Prog.	Order	Rat.	Prog.	Order	Rat.	Prog.	Order	Rat.	Prog.	Order	Rat.	Prog.	Order	Rat.	Prog.
1		4	5	33	5	5	20	3	4				17	3	2	16	4	1			
2		5	6	36	6	6	21	5	6												
3		2	3	18	2	3	1	3	2				1	3	5	20	3	4			
4		3	3	40	4	4															
5		4	5	43	3	5	4	3	3.5												
6		2	3	73	4	4										1	4	6	18	2	*
7		6	6	54	6	5	25	5	6				14	4	2				19	3	5
8		4	3	24	1	3				9	1	1	13	1	1						
9		2	2	42	1	3.5	7	3	2	4	1	3.5				3	3	3.5			
10		4	6	45	5	5	8	5	6										-4	6	6
11		5	5	10	6	5							25	3	2.5	5	5	6			
12		4	5	32	4	6													21	2	6
13		2	2	31	1	1				5	1	1							3	2	*
14		5	5	2	5	4	12	5	5							22	4	5			
15		5	6	21	5	6	2	4	6				15	4	6						
16		1	1	13	1	1	17	1	1	19	2	1	23	2	1						
17	U	3	5	1	3	5	13	2	1				9	2	1						
18	U	2	5	72	1	5	14	2	4							4	1	2.5			
19		2	3	38	2	4															
20		6	6	5	6	6							3	5	6						
21		2	2	4	2	3													7	2	*
22		6	6	60	5	6										13	6	6	2	5	6
23		4	5	6	4	5				24	3	1									
24		5	6	57	5	6	16	5	3.5	12	4	2	24	4	2				6	6	4
25		6	6	61	5	6				10	3	6							1	4	6
26		4	4	70	5	5							21	3	2				10	5	4
27		2	2	44	3	5													9	4	1
28		2	1	64	2	2							11	1	1						
29		5	6	15	4	5	10	5	6	18	4	6	2	5	4				24	4	4
30		5	5	75	5	5	11	5	2												
31		5	5	25	5	5							20	3	2				11	5	1
32		2	3	20	2	2										25	3	2			
33		6	6	67	6	6															
34		4	1	53	4	5										6	3	5	5	5	5
35		4	4	59	4	5				23	4	1	10	4	2						
36		2	3	37	1	1				17	1	1	4	1	1	15	3	1			
37		6	6	28	6	6	22	5	4	15	5	2.5				9	6	2.5			
38		4	4	46	2	3							12	3	6	8	3	3	13	1	*
39		5	6	7	6	6				8	3	6				17	5	6			
40		5	5	55	5	5													12	2	2.5
41		1	3	71	1	3	16	2	2	6	1	1	19	1	1						
42		6	2	66	6	6															
43		5	4	47	5	5													14	5	4
44		2	1	26	3	4	5	2	1	25	3	1									
45		1	1	62	1	2	9	2	2												
46		3	3.5	56	2	3.5							6	2	2	2	3	5			
47		4	6	68	4	4															
48		5	5	30	5	5															
49		1	3	22	2	2.5							16	2	2						
50		4	5	69	3	5				20	3	2							20	2	1
51		2	3	29	1	1				11	1	2							23	1	5
52		4	5	12	4	5	15	5	2	21	3	1				19	4	1			
53		5	6	19	6	6				14	5	6	18	5	6	14	5	6			
54		4	5	23	4	2	3	3	6				22	3	6						
55	U	2	5	34	2	5										12	2	1			
56		6	6	41	6	6				3	3	6				24	6	6			
57		4	3	65	2	3															
58		5	6	8	6	6				2	5	6									
59		2	5	14	4	5	24	4	2												
60		1	1	46	2	2										18	1	1			
61		2	2	74	2	3	6	3	2												
62	U	2	5	51	3	4															
63		6	6	27	6	6							8	5	6	21	5	6	8	3	6
64		4	4	63	4	6													15	5	5
65		2	3	17	1	3	23	3	3							11	2	3.5			
66		5	6	50	6	6															
67		3	2	3	4	5				16	3	2									
68		1	2	49	1	1				22	1	1				7	2	1			
69		3	4	52	2	1	19	3	5												
70		4	3.5	9	4	3.5							5	2	5						
71	U	4	5	16	3	5				13	2	1	7	3	1	23	2	1			
72		4	5	11	3	5													17	3	2
73		5	4	58	4	2				1	4	2							25	5	2
74		4	5	39	3	5				7	2	1							22	2	1
75		4	4	35	1	1										10	4	6	16	2	3

Appendix B

B-1

ANOVA for LA1 Time 1 vs. LA1's EVA Time 1

Source	DF	SS	MS	F	P <=
Judge	1	6.8267	6.8267	9.640	0.003
Slide	74	310.0900	4.1904	5.920	0.00001
Error	74	52.4233	0.7084		
Total	149	369.3400			

r = 0.71
t = 12.266
p <= 0.00001

ANOVA for LA1 Time 2 vs. LA1's EVA Time 2

Source	DF	SS	MS	F	P <=
Judge	1	15.3600	15.3600	26.350	0.00001
Slide	74	365.7933	4.9432	8.480	0.00001
Error	74	43.1400	0.5830		
Total	149	424.2933			

r = 0.79
t = 15.68
p <= 0.00001

ANOVA for LA1's EVA Time 1 vs. LA1's EVA Time 2

Source	DF	SS	MS	F	P <=
Judge	1	0.5400	0.5400	0.580	0.448
Slide	74	307.7233	4.1584	4.480	0.00001
Error	74	68.7100	0.9285		
Total	149	376.9733			

r = 0.63
t = 9.87
p <= 0.00001

B-2

ANOVA for LA1 Time 1 vs. LA1 Time 2

Source	DF	SS	MS	F	P <=
Judge	1	0.3267	0.3267	0.730	0.396
Slide	74	361.8400	4.8897	10.910	0.00001
Error	74	33.1733	0.4483		
Total	149	395.3400			

r = 0.83
t = 18.1
p <= 0.00001

ANOVA for LA1 Time 1 vs. LA1 Time 2 (Urban Scenes Removed)

Source	DF	SS	MS	F	P <=
Judge	1	0.2571	0.2571	0.560	0.457
Slide	69	341.4000	4.9768	10.820	0.00001
Error	69	31.7429	0.4600		
Total	139	375.4000			

r = 0.83
t = 17.48
p <= 0.00001

ANOVA for LA1's EVA Time 1 vs LA1's EVA Time 2 (Urban Scenes Removed)

Source	DF	SS	MS	F	P <=
Judge	1	0.7143	0.7143	0.720	0.398
Slide	69	301.3500	4.3674	4.430	0.00001
Error	69	68.0357	0.9860		
Total	139	370.1000			

r = 0.63
t = 9.53
p <= 0.00001

B-3

ANOVA for LA1 Time 1 vs. LA1's EVA Time 1 (Urban Scenes Removed)

Source	DF	SS	MS	F	P <=
Judge	1	2.8571	2.8571	4.880	0.030
Slide	69	308.4214	4.4699	7.640	0.00001
Error	69	40.3929	0.5854		
Total	139	351.6714			

r = 0.77
t = 14.18
p <= 0.00001

ANOVA for LA1 Time 2 vs. LA1's EVA Time 2 (Urban Scenes Removed)

Source	DF	SS	MS	F	P <=
Judge	1	9.2571	9.2571	19.810	0.00001
Slide	69	363.4714	5.2677	11.270	0.00001
Error	69	32.2429	0.4673		
Total	139	404.9714			

r = 0.83
t = 17.38
p <= 0.00001

LA1 Time 1 vs. LA1 Time 2
(Urban Included)

Agreements	69
Disagreements	6
# Slides	75
Chi-square	66.782
T =	0.856
P <=	0.002

B-4

LA1 Time 1 vs. LA1's EVA Time 1

	Urban Included	Urban Removed
Agreements	64	64
Disagreements	11	6
# Slides	75	70
Chi-square	49.14	60.69
T =	0.736	0.846
P <=	0.002	0.002

LA1 Time 2 vs. LA1's EVA Time 2

	Urban Included	Urban Removed
Agreements	56	55
Disagreements	19	15
# Slides	75	70
Chi-square	26.53	31.65
T =	0.544	0.614
P <=	0.002	0.002

LA1's EVA Time 1 vs. LA1's EVA Time 2

	Urban Included	Urban Removed
Agreements	61	56
Disagreements	14	14
# Slides	75	70
Chi-square	39.85	34.41
T =	0.664	0.64
P <=	0.002	0.002

B-5

ANOVA for LA2 vs. LA1 Time 1

Source	DF	SS	MS	F	P <=
Judge	1	0.18	0.18	0.46	0.503
Slide	24	92.92	3.8717	9.97	0.00001
Error	24	9.32	0.3883		
Total	49	102.42			

r = 0.81
t = 9.57
p <= 0.00001

ANOVA for LA2 vs. LA2's EVA

Source	DF	SS	MS	F	P <=
Judge	1	0.02	0.02	0.02	0.897
Slide	24	91.75	3.823	3.25	0.003
Error	24	28.23	1.176		
Total	49	120			

r = 0.53
t = 4.33
p <= 0.0001

ANOVA for LA2 vs. LA1's EVA Time 1

Source	DF	SS	MS	F	P <=
Judge	1	7.22	7.22	10.64	0.003
Slide	24	97	4.0417	5.96	0.00001
Error	24	16.28	0.6783		
Total	49	120.5			

r = 0.71
t = 6.98
p <= 0.00001

B-6

ANOVA for LA2 vs. LA1 Time 2

Source	DF	SS	MS	F	P <=
Judge	1	1.62	1.62	2.61	0.119
Slide	24	88.72	3.6967	5.96	0.00001
Error	24	14.88	0.62		
Total	49	105.22			

r = 0.71
t = 6.98
p <= 0.00001

ANOVA for LA2 vs. LA1's EVA Time 2

Source	DF	SS	MS	F	P <=
Judge	1	3.92	3.92	3.91	0.06
Slide	24	70	2.917	2.91	0.006
Error	24	24.08	1.003		
Total	49	98			

r = 0.49
t = 3.89
p <= 0.0005

ANOVA for LA2's EVA vs. LA1's EVA Time 1

Source	DF	SS	MS	F	P <=
Judge	1	8	8	8.26	0.008
Slide	24	129.53	5.3971	5.57	0.00001
Error	24	23.25	0.9687		
Total	49	160.78			

r = 0.7
t = 6.79
p <= 0.00001

B-7

ANOVA for LA2's EVA vs. LA1's EVA Time 2

Source	DF	SS	MS	F	P <=
Judge	1	4.5	4.5	1.85	0.186
Slide	24	75.33	3.139	1.29	0.267
Error	24	58.25	2.427		
Total	49	138.08			

r = 0.13
t = 0.908
p <= 0.2615

ANOVA for LA2's EVA vs. LA1 Time 1

Source	DF	SS	MS	F	P <=
Judge	1	0.08	0.08	0.08	0.776
Slide	24	118.57	4.9404	5.12	0.00001
Error	24	23.17	0.9654		
Total	49	141.82			

r = 0.74
t = 7.62
p <= 0.00001

ANOVA for LA2's EVA vs. LA1 Time 2

Source	DF	SS	MS	F	P <=
Judge	1	1.28	1.28	0.87	0.361
Slide	24	107.63	4.485	3.03	0.004
Error	24	35.47	1.478		
Total	49	144.38			

r = 0.5
t = 4
p <= 0.0003

B-8

LA2 vs. LA1 Time 1

Agreements	24
Disagreements	1
# Slides	25
Chi-square	24.8645
T =	0.928
P <=	0.002

LA2 vs. LA2's EVA

Agreements	18
Disagreements	7
# Slides	25
Chi-square	6.6125
T =	0.496
	$0.02 < p < 0.05$

LA2 vs LA1's EVA Time 1

Agreements	19
Disagreements	6
# Slides	25
Chi-square	8.8445
T =	0.568
	$0.002 < p < 0.01$

B-9

LA2 vs. LA1 Time 2

Agreements	22
Disagreements	3
# Slides	25
Chi-square	17.4845
T =	0.784
P <=	0.002

LA2 vs. LA1's EVA Time 2

Agreements	18
Disagreements	7
# Slides	25
Chi-square	6.6125
T =	0.496
	$0.02 < p < 0.05$

LA2's EVA vs. LA1's EVA Time 1

Agreements	18
Disagreements	7
# Slides	25
Chi-square	6.6125
T =	0.496
	$0.02 < p < 0.05$

B-10

LA2's EVA vs. LA1's EVA Time 2

Agreements	14
Disagreements	11
# Slides	25
Chi-square	0.9245
T =	N/A
P >=	0.1

LA2's EVA vs. LA1 Time 1

Agreements	16
Disagreements	9
# Slides	25
Chi-square	3.1205
T =	0.352
	0.1 > p > 0.05

LA2's EVA vs. LA1 Time 2

Agreements	13
Disagreements	12
# Slides	25
Chi-square	0.3125
T =	N/A
P >=	0.1

B-11

ANOVA for LA3 vs. LA1 Time 1

Source	DF	SS	MS	F	P <=
Judge	1	11.52	11.52	19.09	0.00001
Slide	24	94	3.9167	6.49	0.00001
Error	24	14.48	0.6033		
Total	49	120			

r = 0.73
t = 7.4
p <= 0.00001

ANOVA for LA3 vs. LA3's EVA

Source	DF	SS	MS	F	P <=
Judge	1	0.32	0.32	0.19	0.665
Slide	24	107.77	4.49	2.7	0.009
Error	24	39.93	1.664		
Total	49	148.02			

r = 0.46
t = 3.589
p <= 0.0012

ANOVA for LA3 vs LA1's EVA Time 1

Source	DF	SS	MS	F	P <=
Judge	1	24.5	24.5	26.73	0.00001
Slide	24	97.68	4.07	4.44	0.00001
Error	24	22	0.9167		
Total	49	144.18			

r = 0.63
t = 5.62
p <= 0.00001

B-12

ANOVA for LA3 vs. LA1 Time 2

Source	DF	SS	MS	F	P <=
Judge	1	5.12	5.12	11.29	0.003
Slide	24	121.92	5.08	11.21	0.00001
Error	24	10.88	0.4533		
Total	49	137.92			

r = 0.84
t = 10.73
p <= 0.00001

ANOVA for LA3 vs. LA1's EVA Time 2

Source	DF	SS	MS	F	P <=
Judge	1	23.805	23.805	26.8	0.00001
Slide	24	110.72	4.6133	5.19	0.00001
Error	24	21.32	0.8883		
Total	49	155.845			

r = 0.68
t = 6.43
p <= 0.00001

ANOVA for LA3's EVA vs. LA1's EVA Time 1

Source	DF	SS	MS	F	P <=
Judge	1	30.42	30.42	20.66	0.00001
Slide	24	141.97	5.915	4.02	0.001
Error	24	35.33	1.472		
Total	49	207.72			

r = 0.60
t = 5.196
p <= 0.00001

B-13

ANOVA for LA3's EVA vs. LA1's EVA Time 2

Source	DF	SS	MS	F	P <=
Judge	1	29.645	29.645	16.55	0.00001
Slide	24	146.68	6.112	3.41	0.002
Error	24	42.98	1.791		
Total	49	219.305			

r = 0.55
t = 4.56
p <= 0.0001

ANOVA for LA3's EVA vs. LA1 Time 1

Source	DF	SS	MS	F	P <=
Judge	1	15.68	15.68	11.21	0.003
Slide	24	132.53	5.522	3.95	0.001
Error	24	33.57	1.399		
Total	49	181.78			

r = 0.60
t = 5.196
p <= 0.00001

ANOVA for LA3's EVA vs. LA1 Time 1

Source	DF	SS	MS	F	P <=
Judge	1	8	8	5.86	0.023
Slide	24	157.67	6.57	4.81	0.00001
Error	24	32.75	1.365		
Total	49	198.42			

r = 0.66
t = 6.09
p <= 0.00001

LA3 vs. LA1 Time 1

Agreements	19
Disagreements	6
# Slides	25
Chi-Square	8.8445
T =	0.568
	$0.002 < p < 0.01$

LA3 vs. LA3's EVA

Agreements	13
Disagreements	12
# Slides	25
Chi-Square	0.3125
T =	N/A
P >=	0.1

LA3 vs. LA1's EVA Time 1

Agreements	10
Disagreements	15
# Slides	25
Chi-Square	0.605
T =	N/A
P >=	0.1

B-15

LA3 vs. LA1 Time 2

Agreements	22
Disagreements	3
# Slides	25
Chi-Square	17.4845
T =	0.784
P <=	0.002

LA3 vs. LA1's EVA Time 2

Agreements	11
Disagreements	14
# Slides	25
Chi-Square	0.0245
T =	N/A
P >=	0.1

LA3's EVA vs. LA1's EVA Time 1

Agreements	12
Disagreements	13
# Slides	25
Chi-Square	0.0245
T =	N/A
P >=	0.1

B-16

LA3's EVA vs. LA1's EVA Time 2

Agreements	13
Disagreements	12
# Slides	25
Chi-Square	0.3125
T =	N/A
P >=	0.1

LA3's EVA vs. LA1 Time 1

Agreements	14
Disagreements	11
# Slides	25
Chi-Square	0.9245
T =	N/A
P >=	0.1

LA3's EVA vs. LA1 Time 2

Agreements	13
Disagreements	12
# Slides	25
Chi-Square	0.3125
T =	N/A
P >=	0.1

B-17

ANOVA for LA4 vs. LA1 Time 1

Source	DF	SS	MS	F	P <=
Judge	1	8.82	8.82	18.12	0.00001
Slide	24	87.28	3.6367	7.47	0.00001
Error	24	11.68	0.4867		
Total	49	107.78			

r = 0.76
t = 8.10
p <= 0.00001

ANOVA for LA4 vs. LA4's EVA

Source	DF	SS	MS	F	P <=
Judge	1	0.045	0.045	0.03	0.858
Slide	24	95.12	3.963	2.88	0.006
Error	24	33.08	1.378		
Total	49	128.245			

r = 0.48
t = 3.79
p <= 0.0007

ANOVA for LA4 vs. LA1's EVA Time 1

Source	DF	SS	MS	F	P <=
Judge	1	23.12	23.12	57.62	0.00001
Slide	24	86.27	3.5946	8.96	0.00001
Error	24	9.63	0.4013		
Total	49	119.02			

r = 0.80
t = 9.24
p <= 0.00001

B-18

ANOVA for LA4 vs. LA1 Time 2

Source	DF	SS	MS	F	P <=
Judge	1	5.78	5.78	9.42	0.005
Slide	24	106	4.4167	7.2	0.00001
Error	24	14.72	0.6133		
Total	49	126.5			

r = 0.76
t = 8.10
p <= 0.00001

ANOVA for LA4 vs. LA1's EVA Time 2

Source	DF	SS	MS	F	P <=
Judge	1	16.245	16.245	30.87	0.00001
Slide	24	89.83	3.7429	7.11	0.00001
Error	24	12.63	0.5263		
Total	49	118.705			

r = 0.75
t = 7.86
p <= 0.00001

ANOVA for LA4's EVA vs. LA1's EVA Time 1

Source	DF	SS	MS	F	P <=
Judge	1	21.125	21.125	11.2	0.003
Slide	24	96.93	4.039	2.14	0.034
Error	24	45.25	1.885		
Total	49	163.305			

r = 0.36
t = 2.67
p <= 0.0133

B-19

ANOVA for LA4's EVA vs. LA1's EVA Time 2

Source	DF	SS	MS	F	P <=
Judge	1	14.58	14.58	5.84	0.024
Slide	24	88.82	3.701	1.48	0.171
Error	24	59.92	2.497		
Total	49	163.32			

r = 0.19
t = 1.34
p <= 0.1614

ANOVA for LA4's EVA vs. LA1 Time 1

Source	DF	SS	MS	F	P <=
Judge	1	7.605	7.605	4.68	0.041
Slide	24	106.22	4.426	2.72	0.009
Error	24	39.02	1.626		
Total	49	152.845			

r = 0.46
t = 3.98
p <= 0.0004

ANOVA for LA4's EVA vs. LA1 Time 2

Source	DF	SS	MS	F	P <=
Judge	1	4.805	4.805	2.08	0.162
Slide	24	111.68	4.653	2.02	0.046
Error	24	55.32	2.305		
Total	49	171.805			

r = 0.34
t = 2.5
p <= 0.0198

B-20

LA4 vs. LA1 Time 1

Agreements	20
Disagreements	5
# Slides	25
Chi-Square	11.401
T =	0.64
P <=	0.002

LA4 vs. LA4's EVA

Agreements	15
Disagreements	10
# Slides	25
Chi-Square	1.8605
T =	N/A
P >=	0.2

LA4 vs. LA1's EVA Time 1

Agreements	11
Disagreements	14
# Slides	25
Chi-Square	0.0245
T =	N/A
P >=	0.2

B-21

LA4 vs. LA1 Time 2

Agreements	21
Disagreements	4
# Slides	25
Chi-Square	14.2805
T =	0.712
P <=	0.002

LA4 vs. LA1's EVA Time 2

Agreements	14
Disagreements	11
# Slides	25
Chi-Square	0.9245
T =	N/A
P >=	0.2

LA4's EVA vs. LA1's EVA Time 1

Agreements	8
Disagreements	17
# Slides	25
Chi-Square	1.1045
T =	N/A
P >=	0.2

LA4's EVA vs. LA1's EVA Time 2

Agreements	9
Disagreements	16
# Slides	25
Chi-Square	0.4205
T =	N/A
P >=	0.2

LA4's EVA vs. LA1 Time 1

Agreements	12
Disagreements	13
# Slides	25
Chi-Square	0.0245
T =	N/A
P >=	0.2

LA4's EVA vs. LA1 Time 2

Agreements	13
Disagreements	12
# Slides	25
Chi-Square	0.3125
T =	N/A
P >=	0.2

B-23

ANOVA for Layman 1 vs. LA1 Time 1

Source	DF	SS	MS	F	P <=
Judge	1	0.18	0.18	0.32	0.574
Slide	24	99	4.125	7.43	0.00001
Error	24	13.32	0.555		
Total	49	112.5			

r = 0.76
t = 8.10
p <= 0.00001

ANOVA for Layman 1 vs. Layman 1's EVA

Source	DF	SS	MS	F	P <=
Judge	1	0.5	0.5	0.33	0.572
Slide	24	114.42	4.767	3.13	0.003
Error	24	36.5	1.521		
Total	49	151.42			

r = 0.52
t = 4.22
p <= 0.0002

ANOVA for Layman 1 vs. LA1's EVA Time 1

Source	DF	SS	MS	F	P <=
Judge	1	4.205	4.205	3.49	0.074
Slide	24	76.48	3.187	2.64	0.01
Error	24	28.92	1.205		
Total	49	109.605			

r = 0.45
t = 4.12
p <= 0.0002

ANOVA for Layman 1 vs. LA1 Time 2

Source	DF	SS	MS	F	P <=
Judge	1	0.08	0.08	0.11	0.746
Slide	24	124	5.1667	6.92	0.00001
Error	24	17.92	0.4767		
Total	49	142			

r = 0.75
t = 7.86
p <= 0.00001

ANOVA for Layman 1 vs. LA1's EVA Time 2

Source	DF	SS	MS	F	P <=
Judge	1	5.12	5.12	4.15	0.053
Slide	24	85.87	3.578	2.9	0.006
Error	24	29.63	1.235		
Total	49	120.62			

r = 0.49
t = 3.89
p <= 0.0005

ANOVA for Layman 1's EVA vs. LA1's EVA Time 1

Source	DF	SS	MS	F	P <=
Judge	1	1.805	1.805	0.79	0.383
Slide	24	109.18	4.549	1.99	0.049
Error	24	54.82	2.284		
Total	49	165.805			

r = 0.33
t = 2.42
p <= 0.0236

ANOVA for Layman 1's EVA vs. LA1's EVA Time 2

Source	DF	SS	MS	F	P <=
Judge	1	2.42	2.42	1.03	0.321
Slide	24	117.52	4.897	2.08	0.04
Error	24	56.58	2.358		
Total	49	176.52			

r = .35
t = 2.59
p <= 0.0161

ANOVA for Layman 1's EVA vs. LA1 Time 1

Source	DF	SS	MS	F	P <=
Judge	1	0.08	0.08	0.05	0.827
Slide	24	131.5	5.479	3.34	0.002
Error	24	39.42	1.643		
Total	49	171			

r = 0.54
t = 4.45
p <= 0.0001

ANOVA for Layman 1's EVA vs. LA1 Time 2

Source	DF	SS	MS	F	P <=
Judge	1	0.98	0.98	0.47	0.502
Slide	24	150	6.25	2.97	0.005
Error	24	50.52	2.105		
Total	49	201.5			

r = 0.50
t = 4
p <= 0.0003

B-26

Layman 1 vs. LA1 Time 1

Agreements	23
Disagreements	2
# Slides	25
Chi-Square	21.0125
T =	0.856
P <=	0.002

Layman 1 vs. Layman 1's EVA

Agreements	15
Disagreements	10
# Slides	25
Chi-Square	1.8605
T =	N/A
P >=	0.2

Layman 1 vs. LA1's EVA Time 1

Agreements	21
Disagreements	4
# Slides	25
Chi-Square	14.2805
T =	0.712
P <=	0.002

B-27

Layman 1 vs. LA1 Time 2

Agreements	22
Disagreements	3
# Slides	25
Chi-Square	17.4845
T =	0.784
P <=	0.002

Layman 1 vs. LA1's EVA Time 2

Agreements	16
Disagreements	9
# Slides	25
Chi-Square	3.1205
T =	N/A
	0.2 > p > 0.1

Layman 1's EVA vs. LA1's EVA Time 1

Agreements	14
Disagreements	11
# Slides	25
Chi-Square	0.9245
T =	N/A
P >=	0.2

B-28

Layman 1's EVA vs. LA1's EVA Time 2

Agreements	16
Disagreements	9
# Slides	25
Chi-Square	3.1205
T =	N/A
	0.2 > p > 0.1

Layman 1's EVA vs. LA1 Time 1

Agreements	15
Disagreements	10
# Slides	25
Chi-Square	1.8605
T =	N/A
P >=	0.2

Layman 1's EVA vs. LA1 Time 2

Agreements	14
Disagreements	11
# Slides	25
Chi-Square	0.9245
T =	N/A
P >=	0.2

B-29

ANOVA for Layman 2 vs. LA1 Time 1

Source	DF	SS	MS	F	P <=
Judge	1	6.48	6.48	5.27	0.031
Slide	24	72	3	2.44	0.017
Error	24	29.52	1.23		
Total	49	108			

r = 0.42
t = 3.21
p <= 0.0034

ANOVA for Layman 2 vs. Layman 2's EVA

Source	DF	SS	MS	F	P <=
Judge	1	0.006	0.006	0	0.96
Slide	24	68.976	3.449	1.46	0.201
Error	24	47.119	2.356		
Total	49	116.101			

r = 0.19
t = 1.34
p <= 0.1614

ANOVA for Layman 2 vs. LA1's EVA Time 1

Source	DF	SS	MS	F	P <=
Judge	1	10.58	10.58	5.78	0.024
Slide	24	70	2.917	1.59	0.13
Error	24	43.92	1.83		
Total	49	124.5			

r = 0.23
t = 1.64
p <= 0.1044

B-30

ANOVA for Layman 2 vs. LA1 Time 2

Source	DF	SS	MS	F	P <=
Judge	1	1.62	1.62	1.78	0.195
Slide	24	92.28	3.845	4.22	0.00001
Error	24	21.88	0.9117		
Total	49	115.78			

r = 0.62
t = 5.47
p <= 0.00001

ANOVA for Layman 2 vs. LA1's EVA Time 2

Source	DF	SS	MS	F	P <=
Judge	1	12.5	12.5	9.37	0.005
Slide	24	92.32	3.847	2.88	0.006
Error	24	32	1.333		
Total	49	136.82			

r = 0.49
t = 3.89
p <= 0.0005

ANOVA for Layman 2's EVA vs. LA1's EVA Time 1

Source	DF	SS	MS	F	P <=
Judge	1	8.149	8.149	3.88	0.063
Slide	20	66.976	3.349	1.6	0.152
Error	20	41.976	2.099		
Total	41	117.101			

r = 0.23
t = 1.49
p <= 0.1310

B-31

ANOVA for Layman 2's EVA vs. LA1's EVA Time 2

Source	DF	SS	MS	F	P <=
Judge	1	10.006	10.006	4.54	0.046
Slide	20	71.976	3.599	1.63	0.141
Error	20	44.119	2.206		
Total	41	126.101			

r = 0.24
t = 1.56
p <= 0.1181

ANOVA for Layman 2's EVA vs. LA1 Time 1

Source	DF	SS	MS	F	P <=
Judge	1	5.006	5.006	2.93	0.102
Slide	20	61.405	3.07	1.8	0.099
Error	20	34.119	1.706		
Total	41	100.53			

r = 0.29
t = 1.92
p <= 0.0651

ANOVA for Layman 2's EVA vs. LA1 Time 2

Source	DF	SS	MS	F	P <=
Judge	1	1.006	1.006	0.57	0.458
Slide	20	72.976	3.649	2.08	0.055
Error	20	35.119	1.756		
Total	41	109.101			

r = 0.35
t = 2.36
p <= 0.0274

Layman 2 vs. LA1 Time 1

Agreements	16
Disagreements	9
# Slides	25
Chi-Square	3.1205
T =	N/A
	0.2 > p > 0.1

Layman 2 vs. Layman 2's EVA

Agreements	12
Disagreements	9
# Slides	21
Chi-Square	0.9053
T =	N/A
P >=	0.2

Layman 2 vs. LA1's EVA Time 1

Agreements	11
Disagreements	14
# Slides	25
Chi-Square	0.0245
T =	N/A
P >=	0.2

B-33

Layman 2 vs. LA1 Time 2

Agreements	20
Disagreements	5
# Slides	25
Chi-Square	11.4005
T =	0.64
P <=	0.002

Layman 2 vs. LA1's EVA Time 2

Agreements	17
Disagreements	8
# Slides	25
Chi-Square	4.7045
T =	0.424
	0.1 > p > 0.05

Layman 2's EVA vs. LA1's EVA Time 1

Agreements	11
Disagreements	10
# Slides	21
Chi-Square	0.2625
T =	N/A
P >=	0.2

B-34

Layman 2's EVA vs. LA1's EVA Time 2

Agreements	12
Disagreements	9
# Slides	21
Chi-Square	0.9053
T =	N/A
P >=	0.2

Layman 2's EVA vs. LA1 Time 1

Agreements	11
Disagreements	10
# Slides	21
Chi-Square	0.2625
T =	N/A
P >=	0.1

Layman 2's EVA vs. LA1 Time 2

Agreements	12
Disagreements	9
# Slides	21
Chi-Square	0.9053
T =	N/A
P >=	0.2

Appendix C

The figures refered to in the preceding work follow. A "U" inside a figure denotes an urban scene.

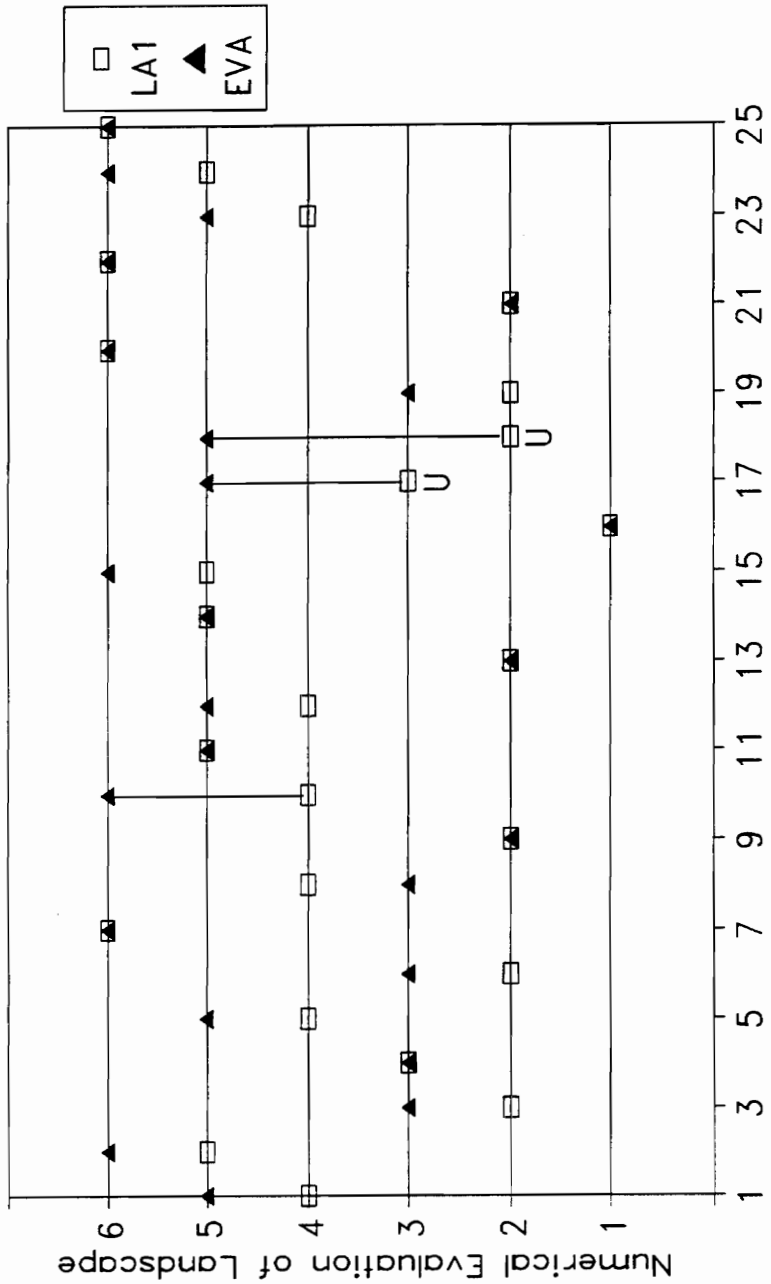

Figure 1A: Agreement of LA1 and EVA
Slides 1–25, Time 1

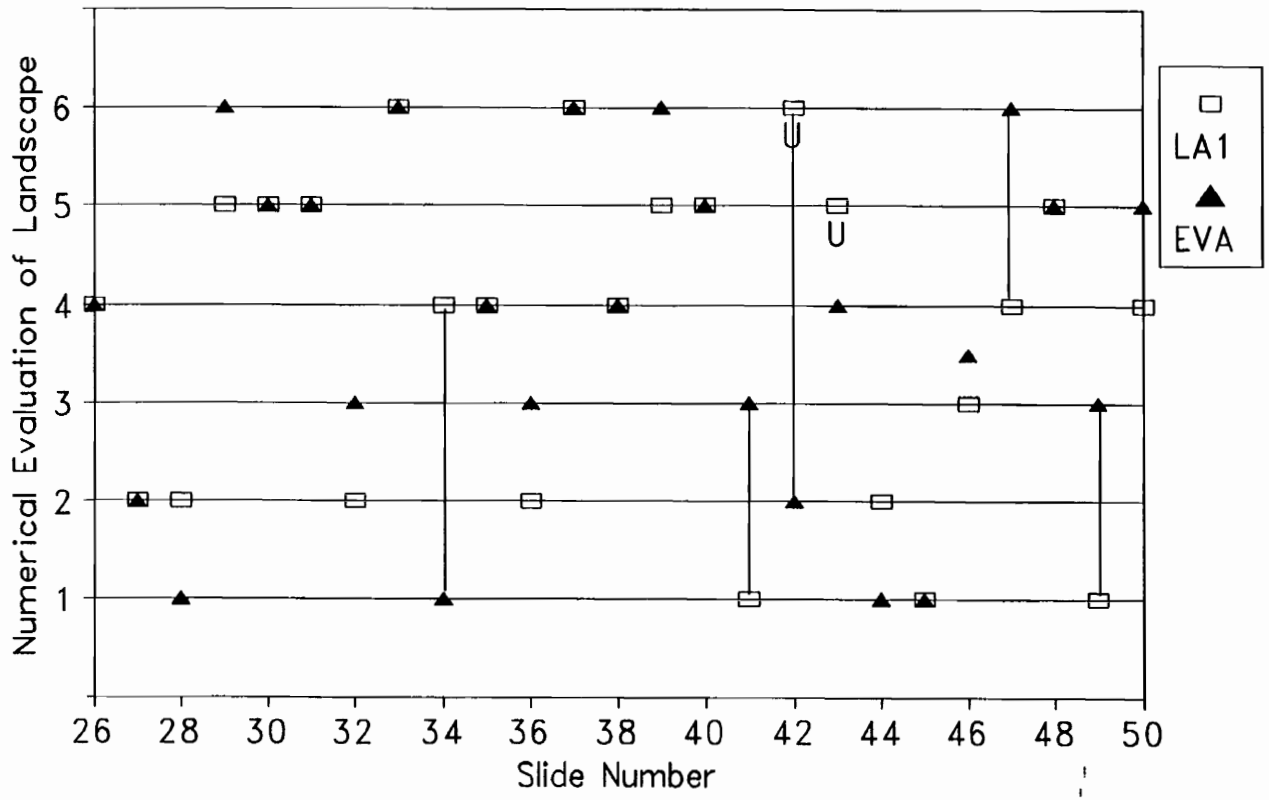

Figure 1B: Agreement of LA1 and EVA
Slides 26–50, Time 1

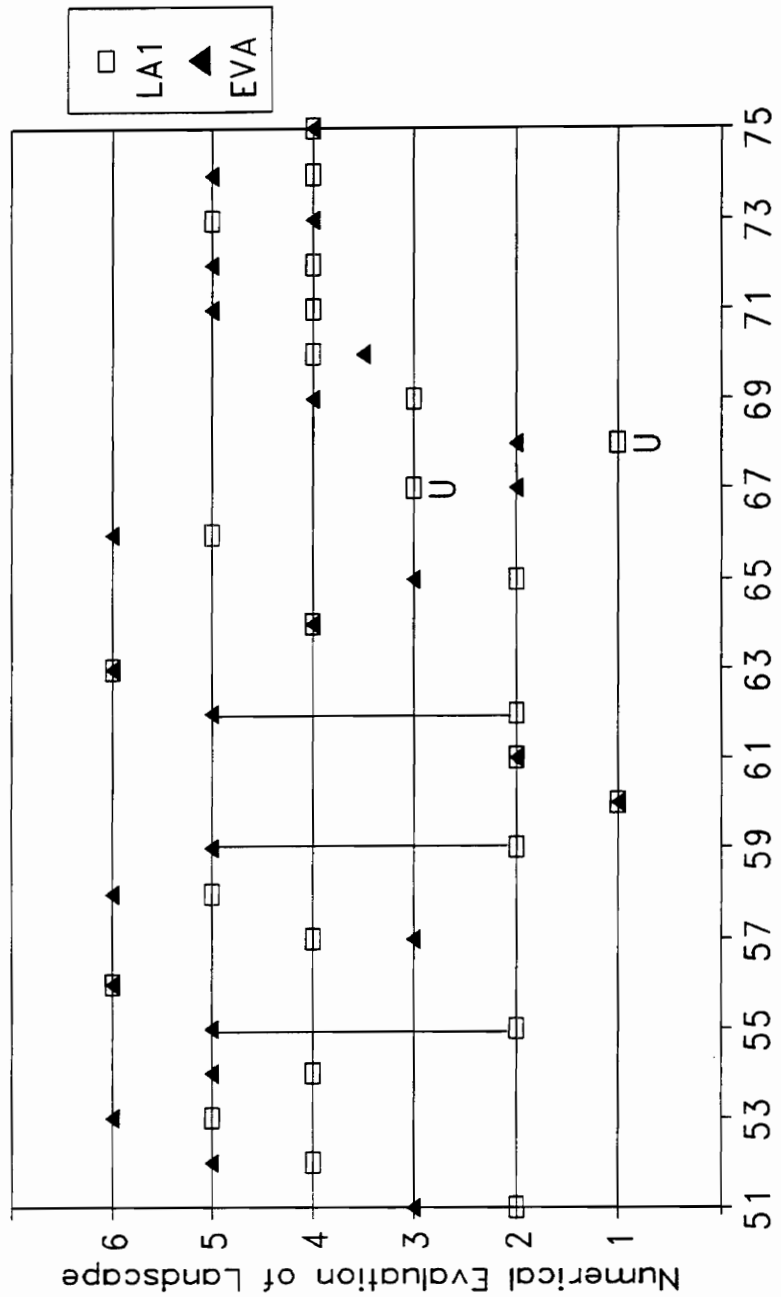

Figure 1C: Agreement of LA1 and EVA
Slides 51–75, Time 1

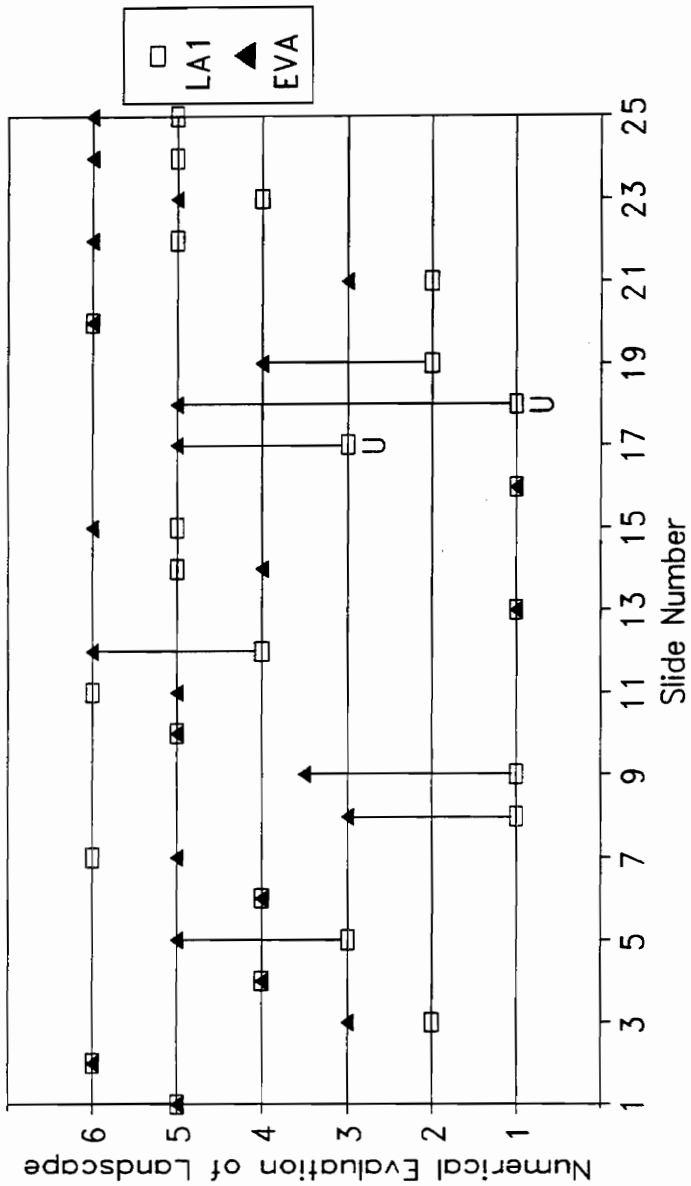

Figure 2A: Agreement of LA1 and EVA
Slides # 1–25, Time 2

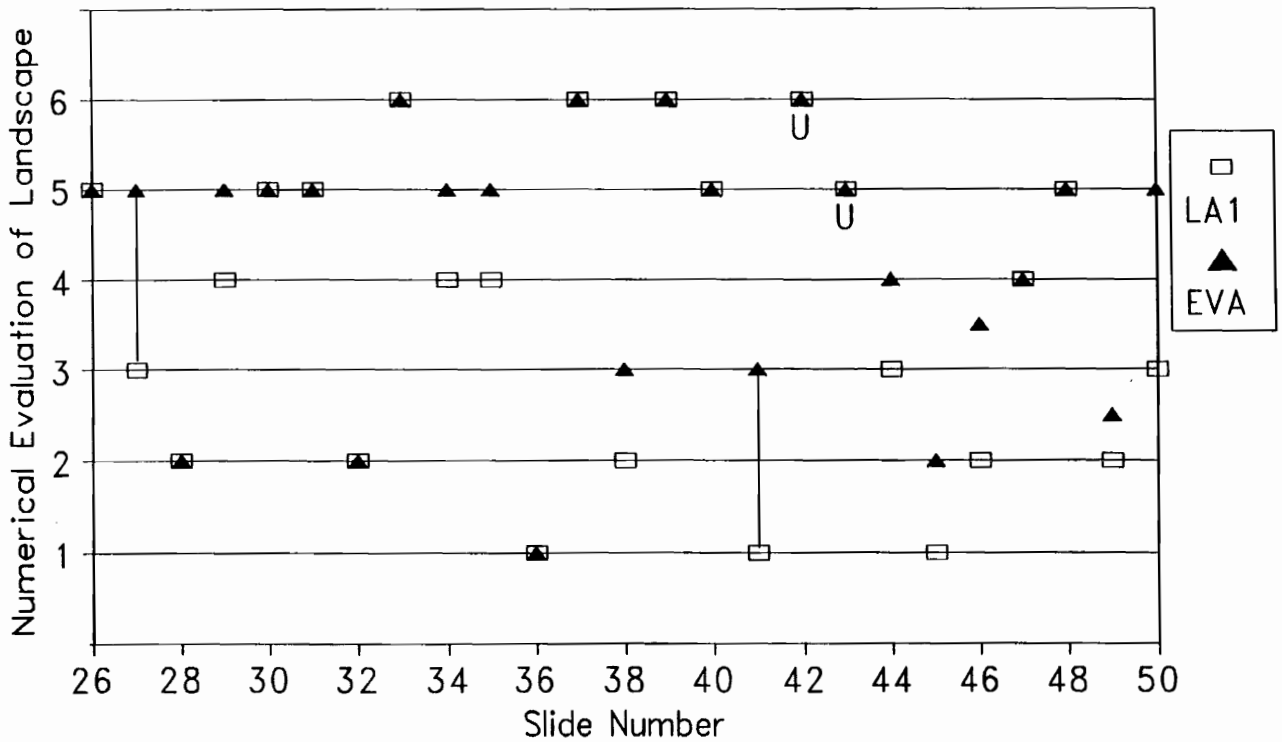

Figure 2B: Agreement of LA1 and EVA
Slides # 26–50, Time 2

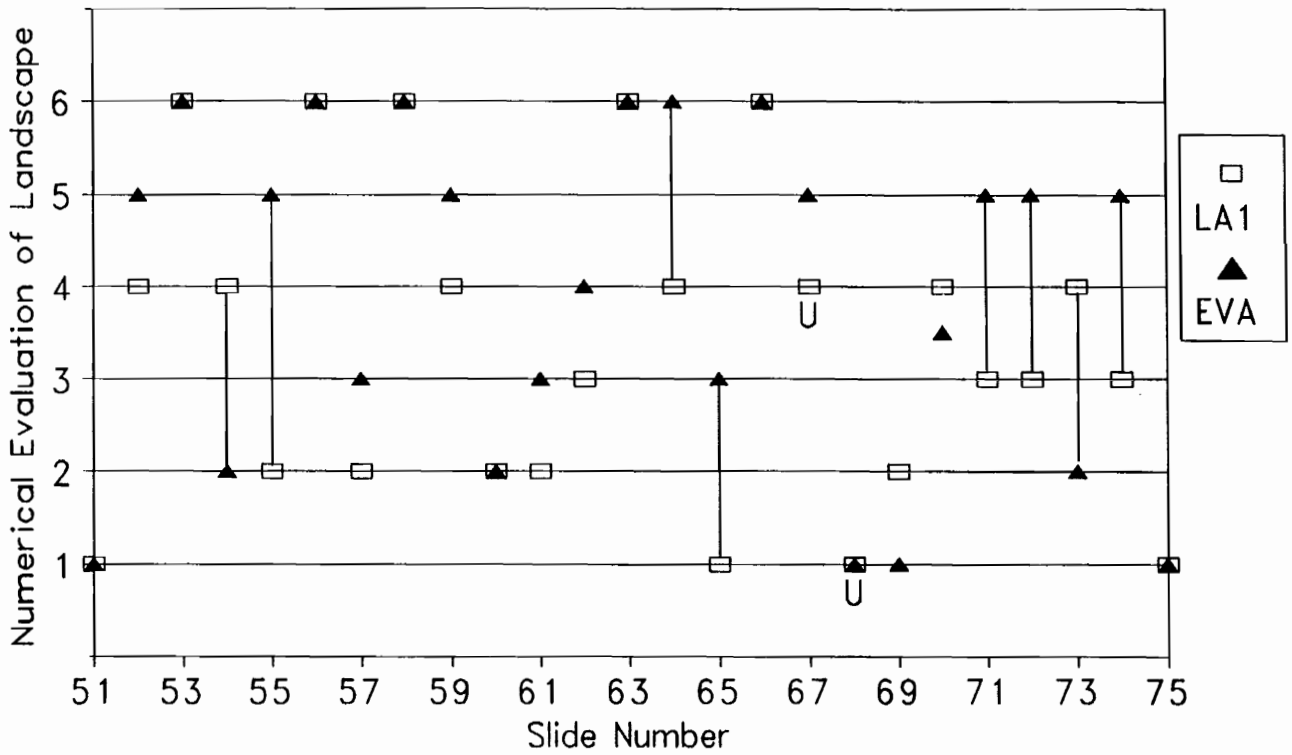

Figure 2C: Agreement of LA1 and EVA
Slides # 51–57 Time 2

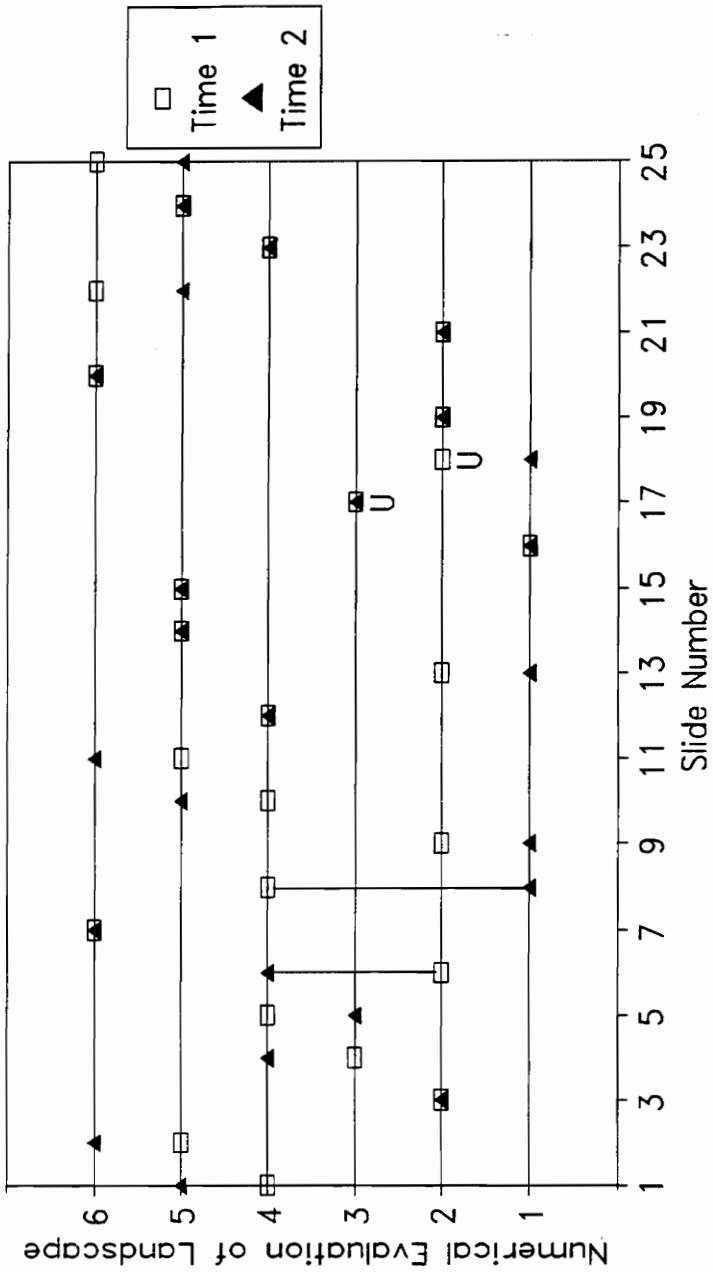

Figure 3A: Agreement:LA1 Time 1, Time 2
Slides # 1-25

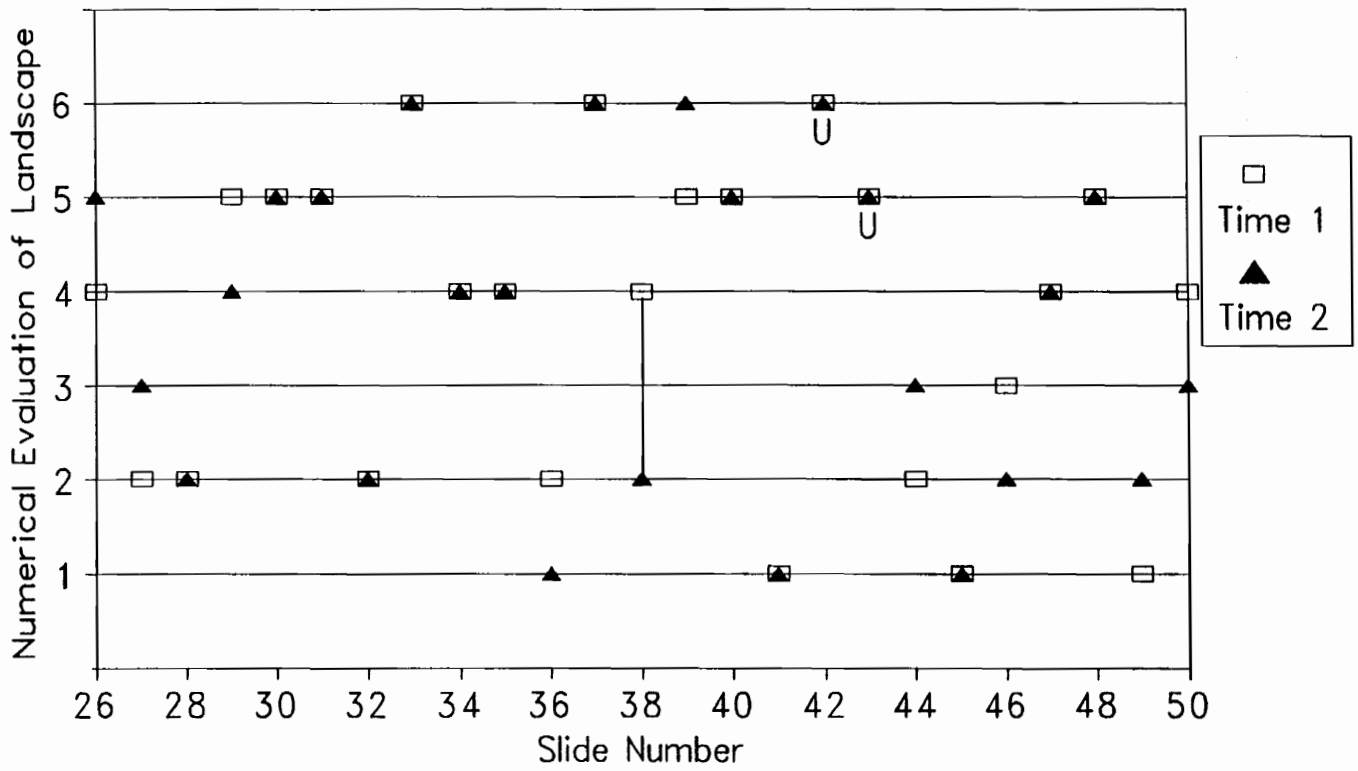

Figure 3B: Agreement:LA1 Time 1, Time 2
Slides # 26-50

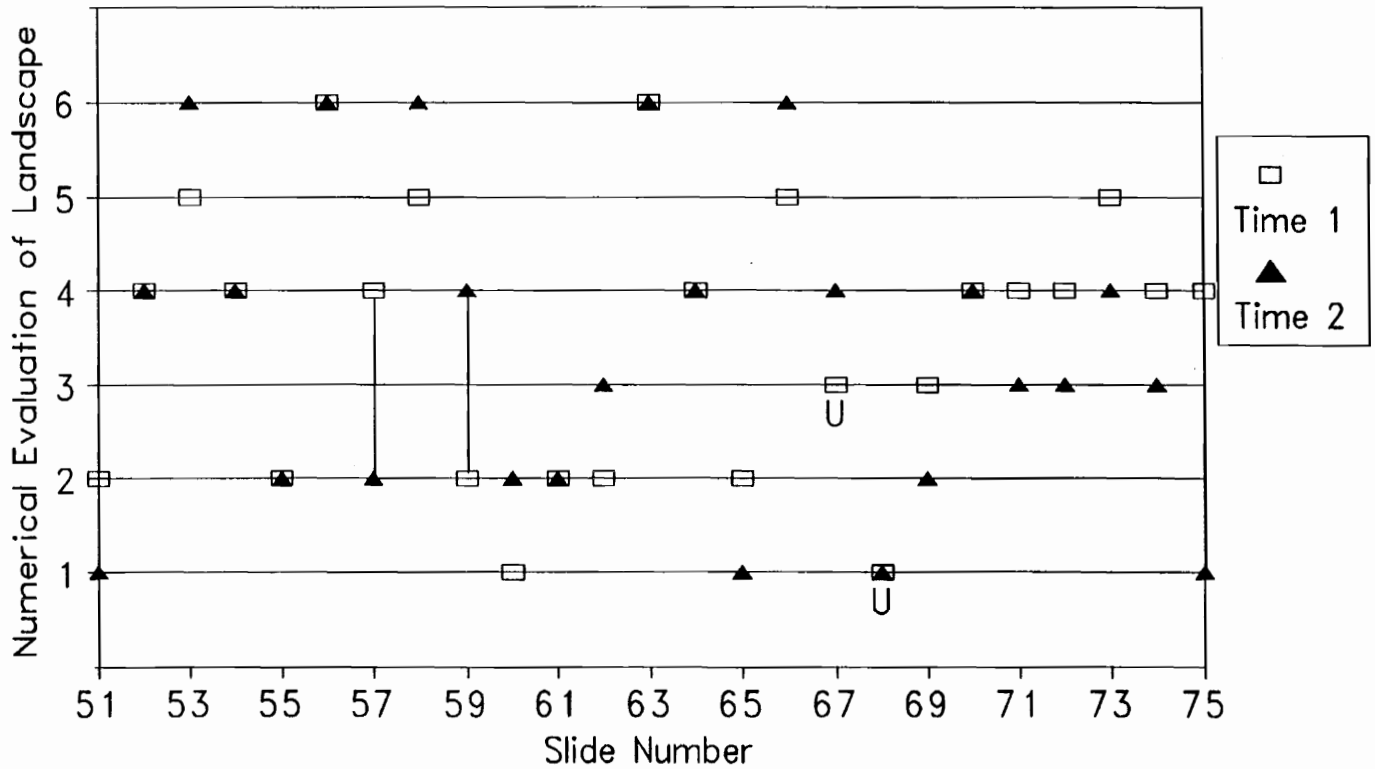

Figure 3C: Agreement:LA1 Time 1, Time 2
Slides # 51-75

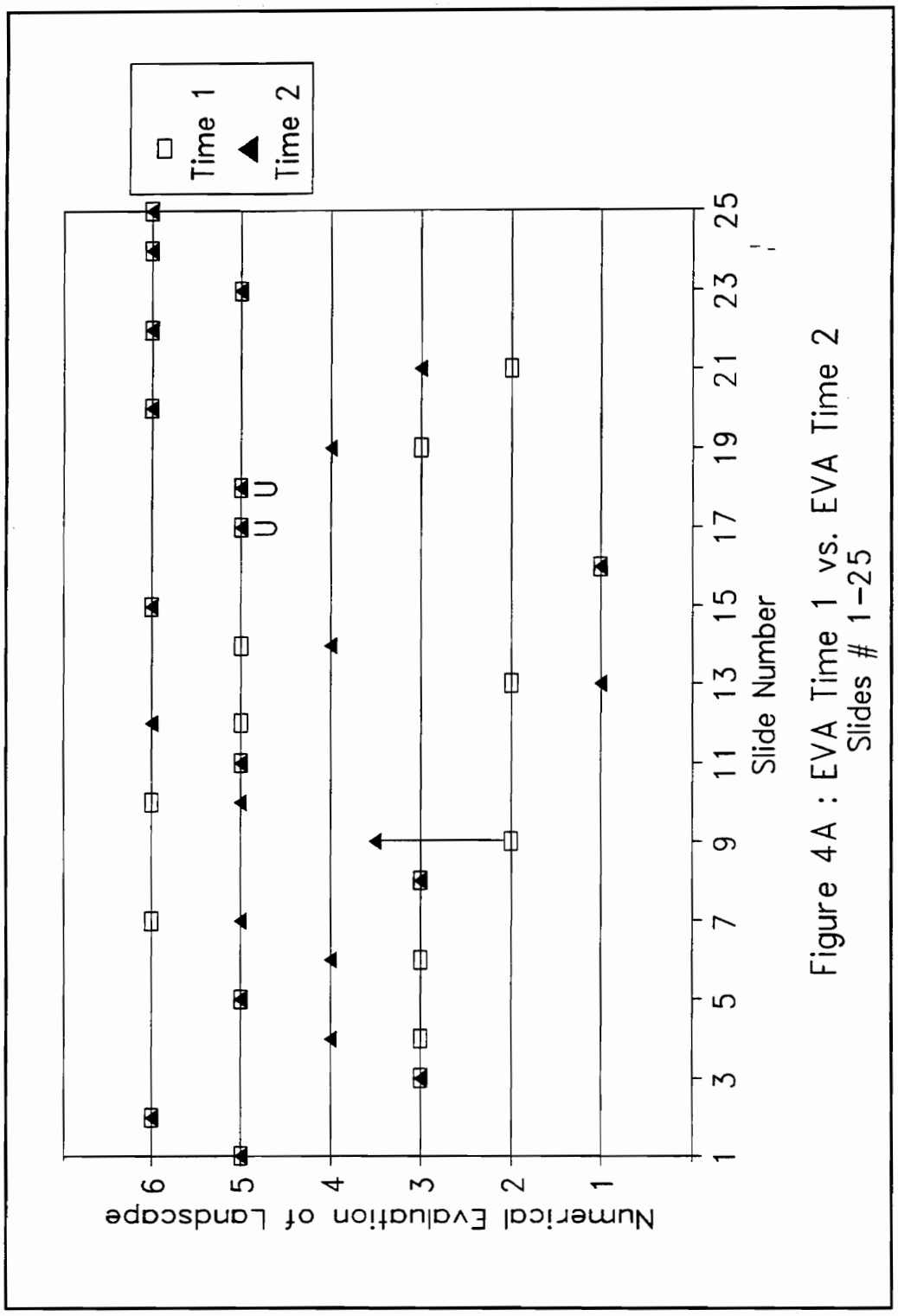

Figure 4A : EVA Time 1 vs. EVA Time 2
Slides # 1–25

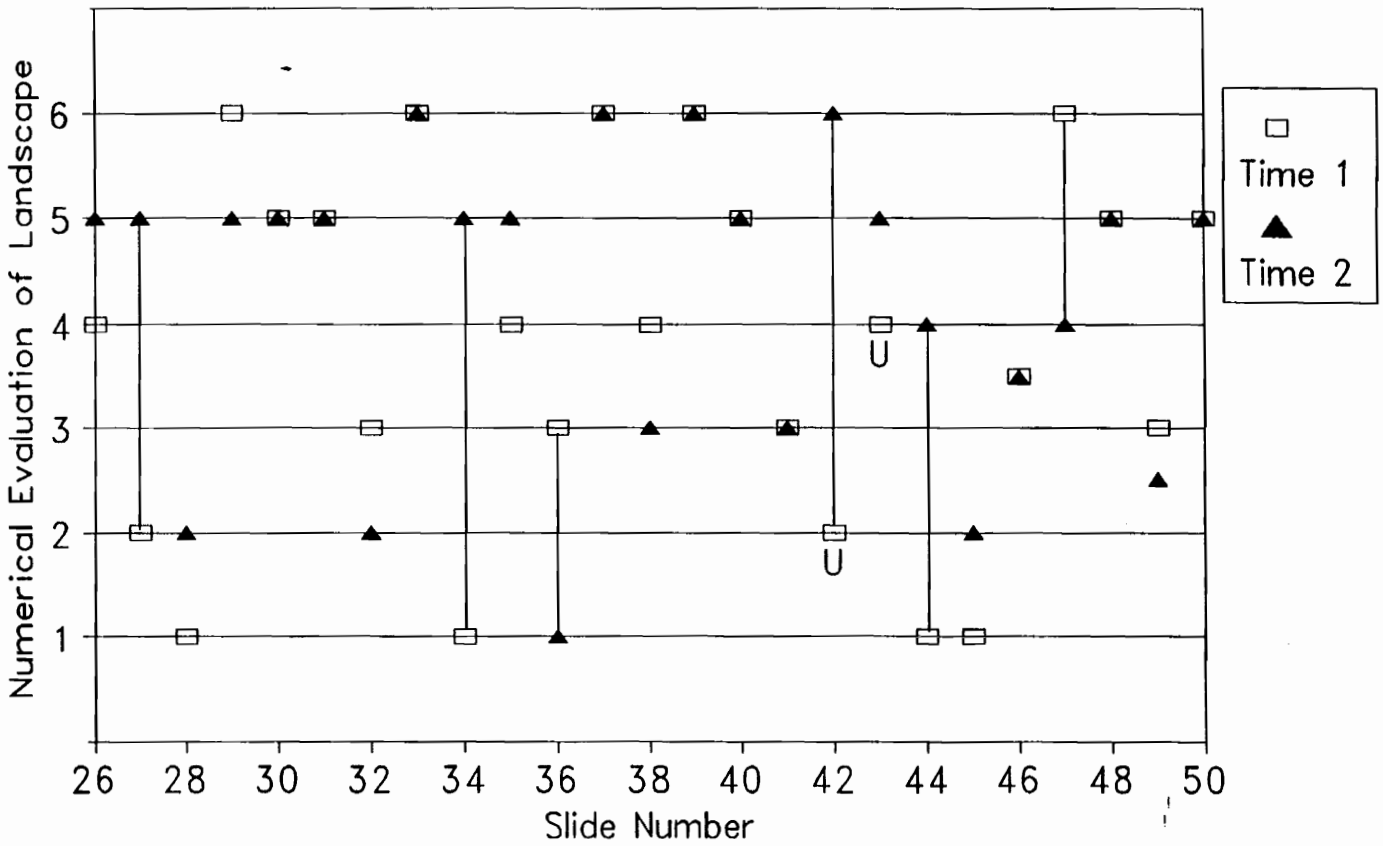

Figure 4B : EVA Time 1 vs. EVA Time 2
Slides # 26-50

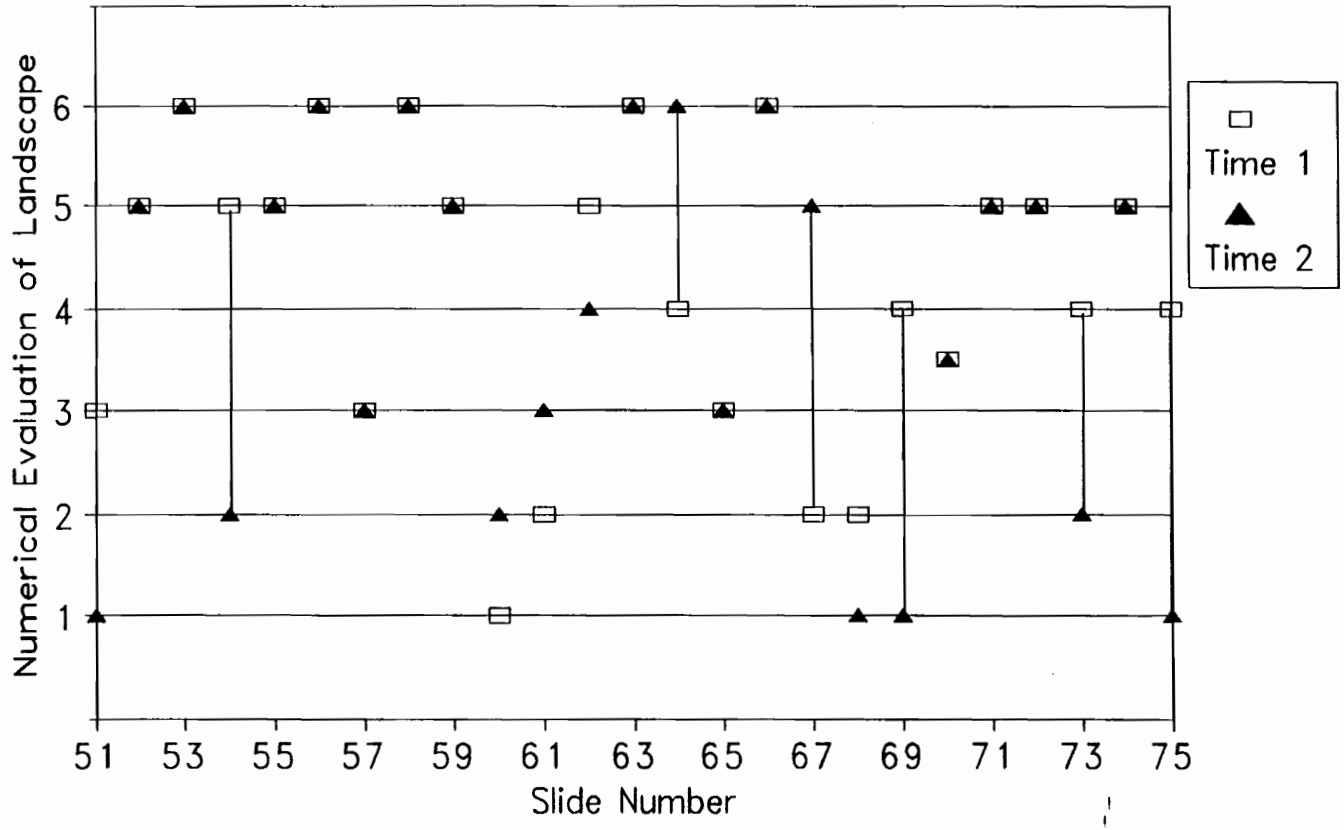

Figure 4C : EVA Time 1 vs. EVA Time 2
Slides # 51-75

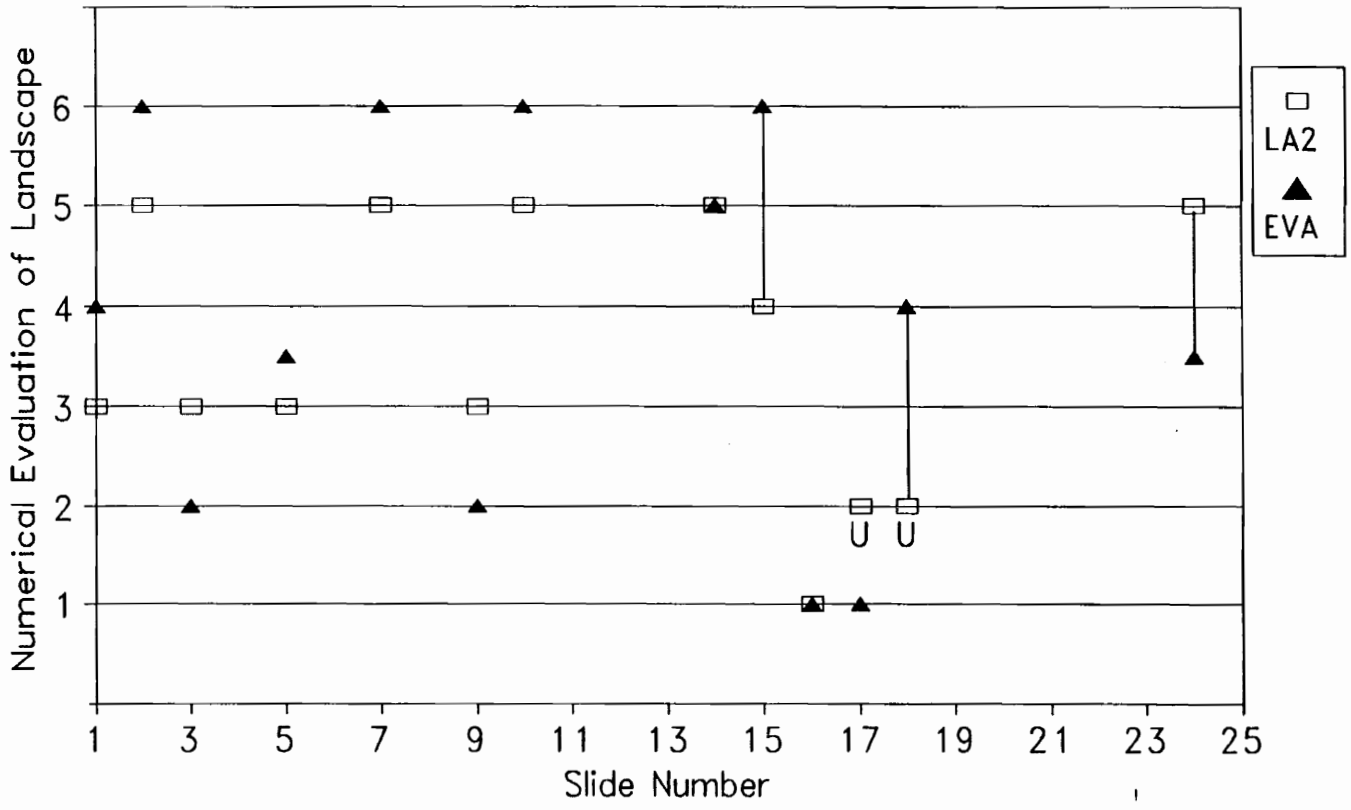

Figure 5A: LA2 vs. EVA
Slides # 1–25

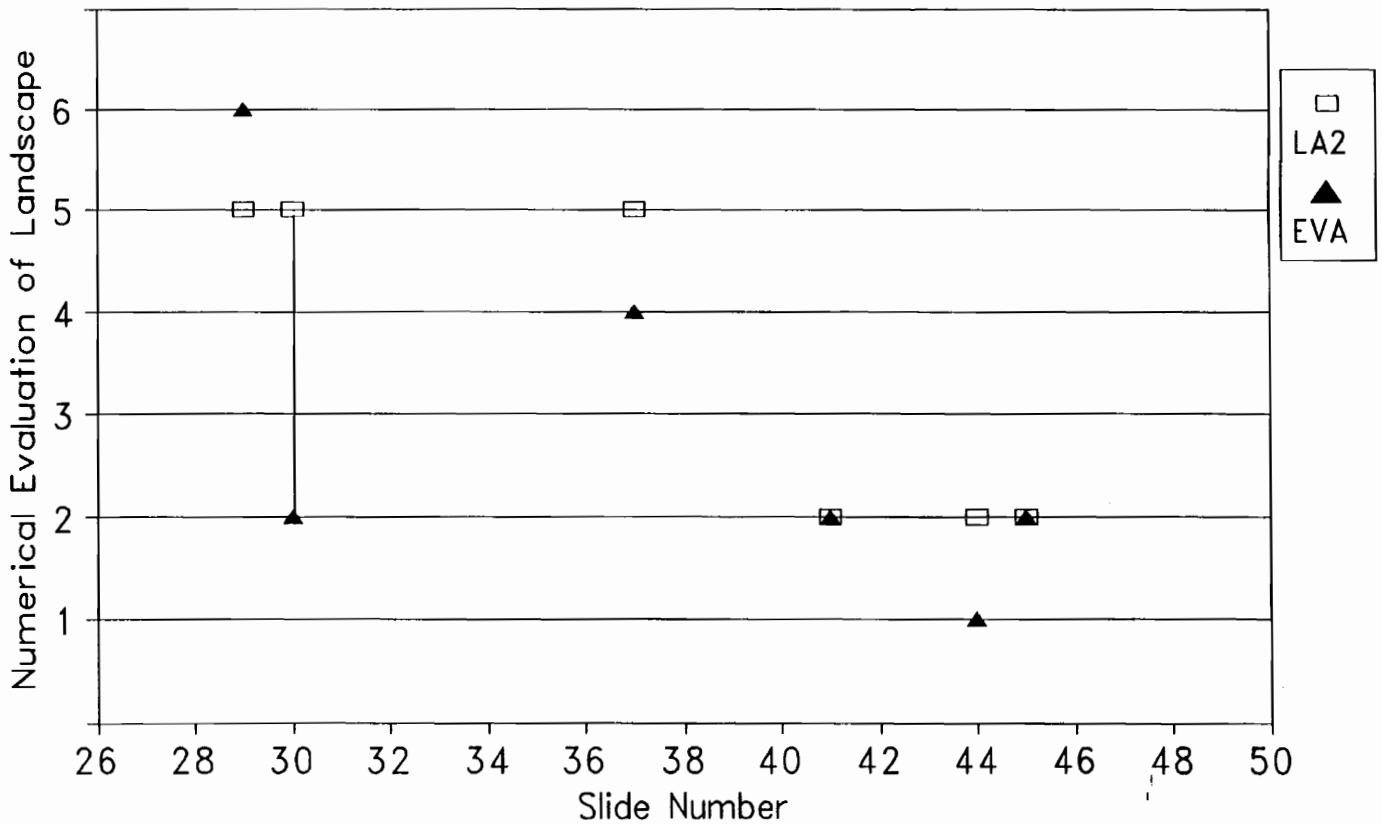

Figure 5B: LA2 vs. EVA
Slides # 26–50

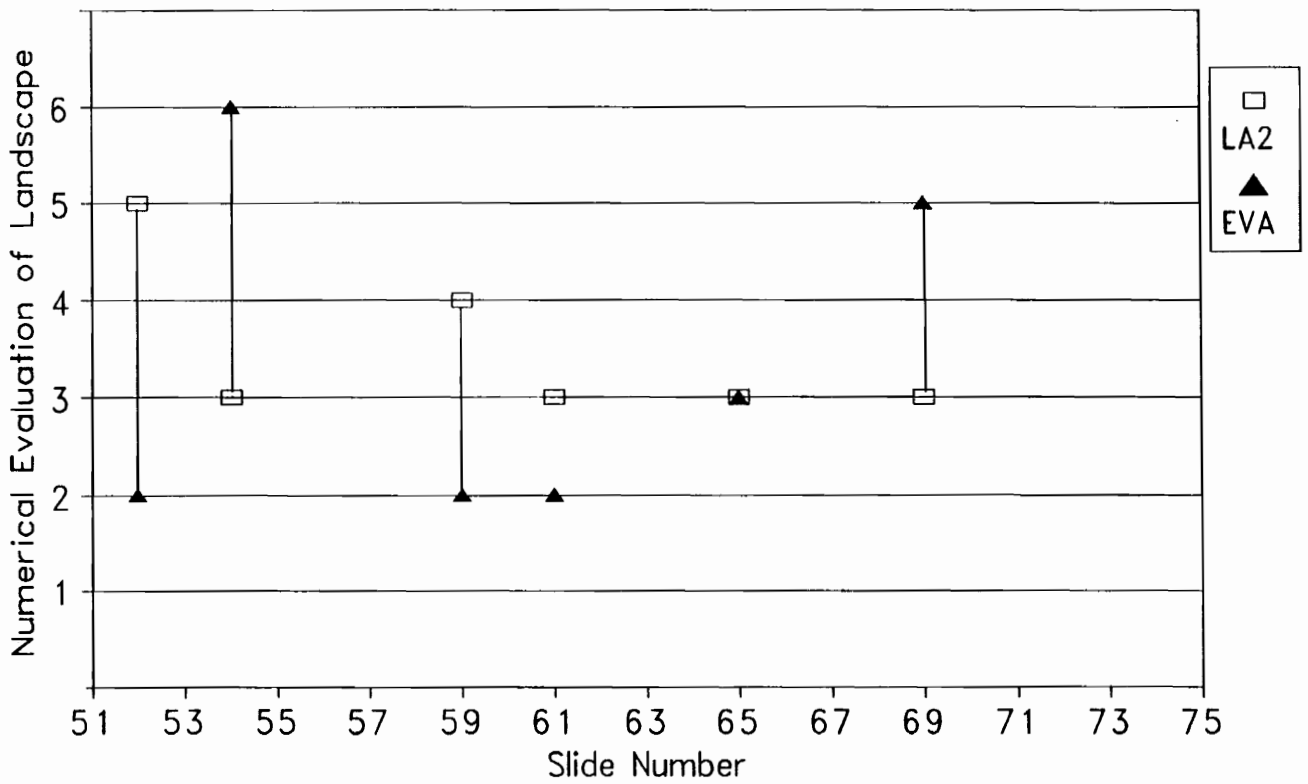

Figure 5C: LA2 vs. EVA
Slides # 51–75

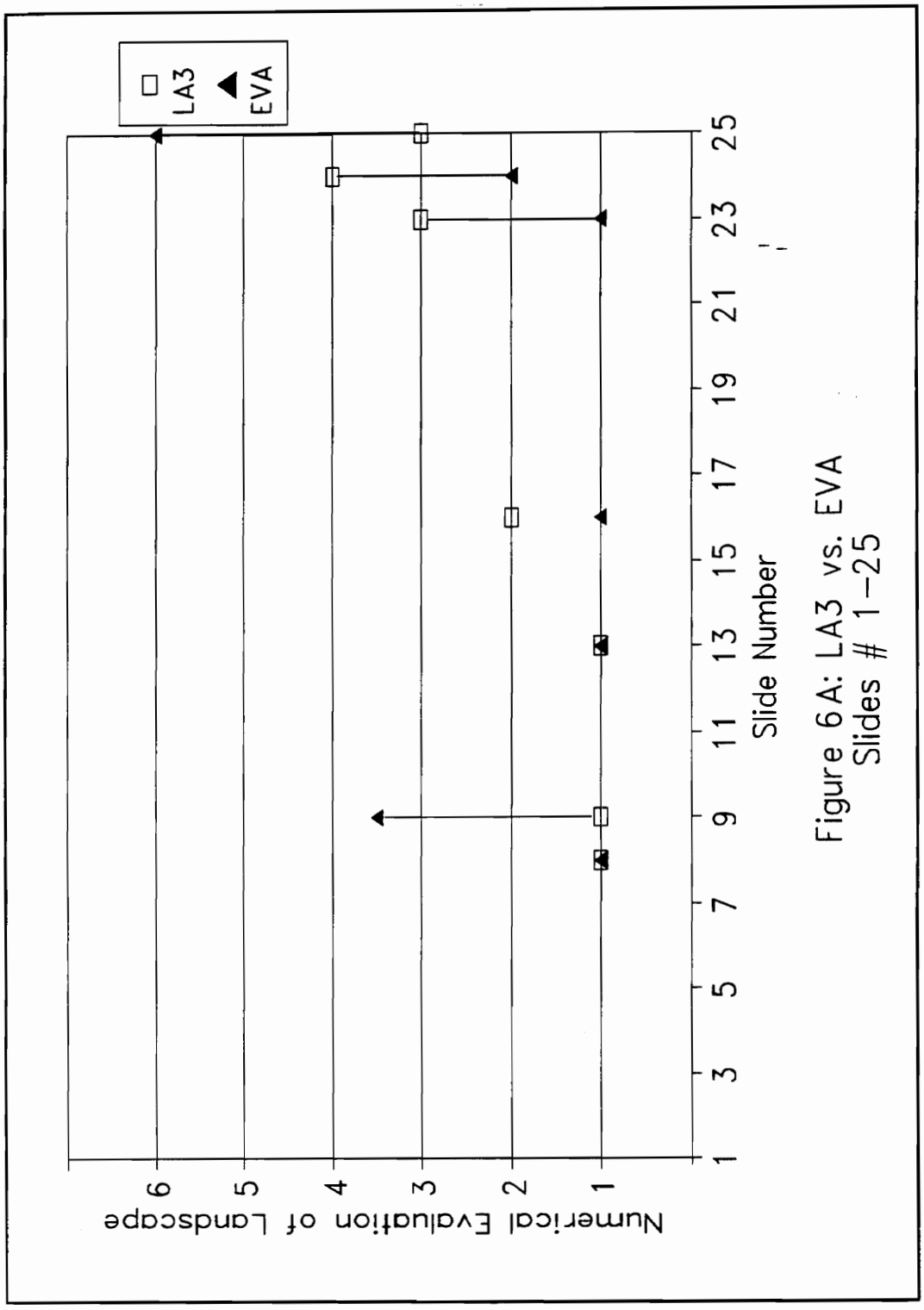

Figure 6A: LA3 vs. EVA
Slides # 1–25

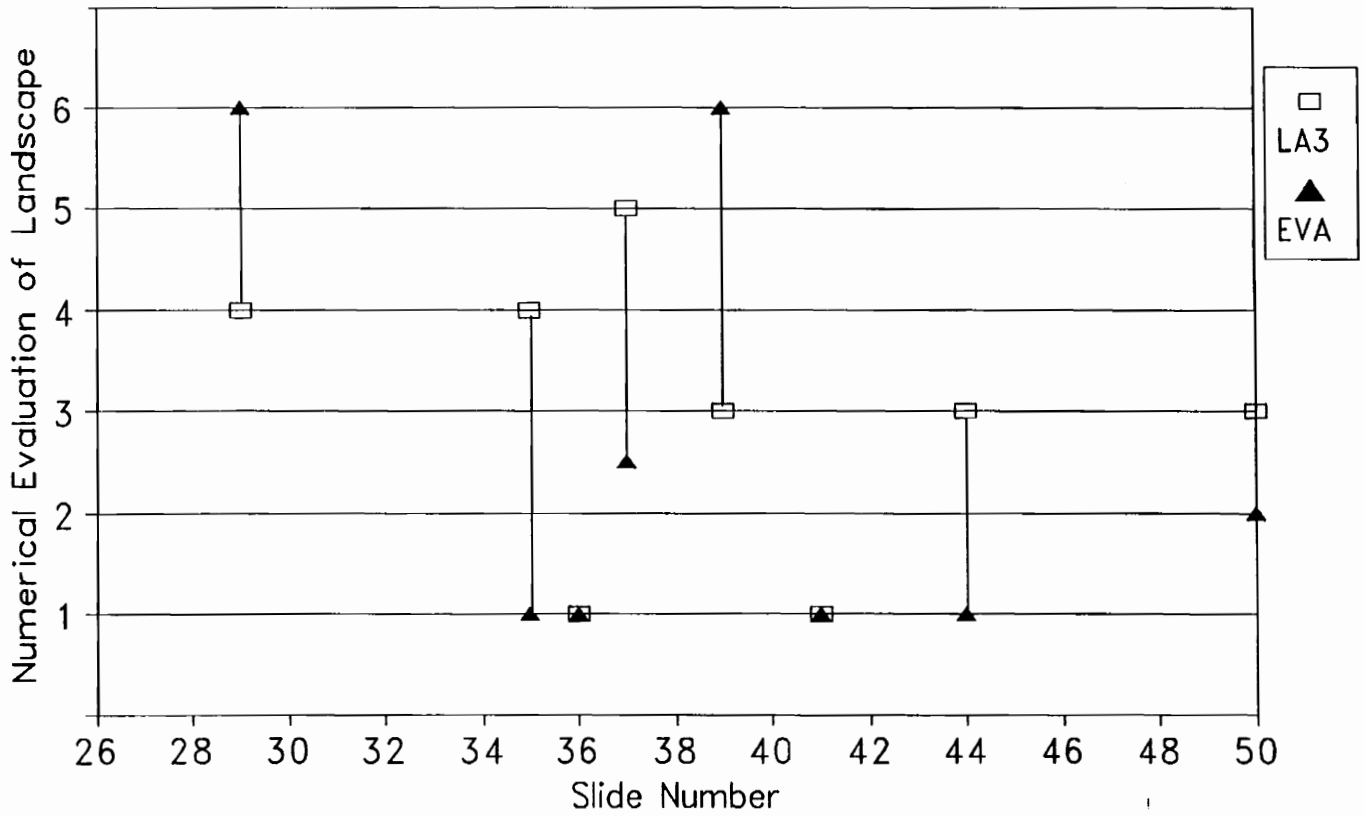

Figure 6B: LA3 vs. EVA
Slides # 26–50

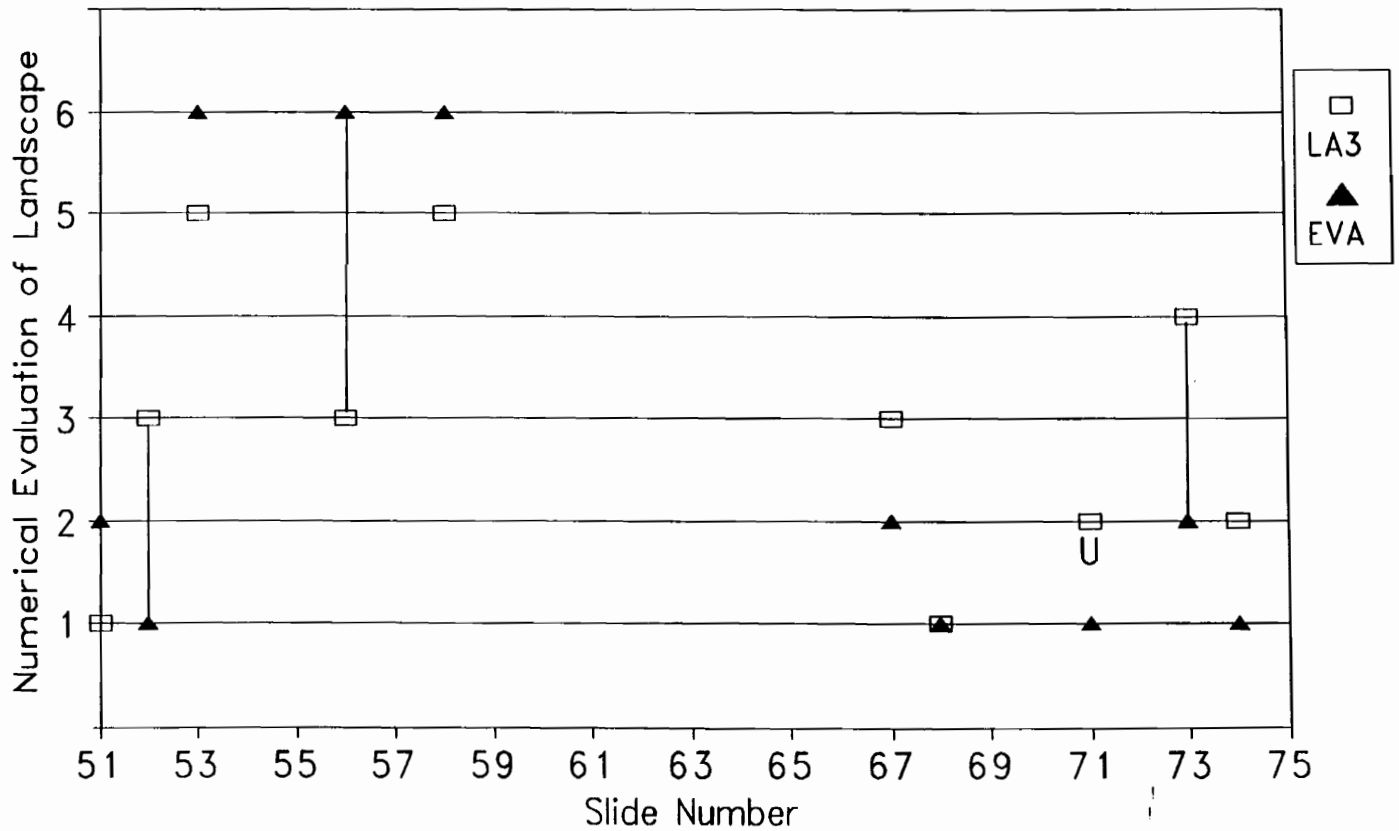

Figure 6C: LA3 vs. EVA
Slides # 51–75

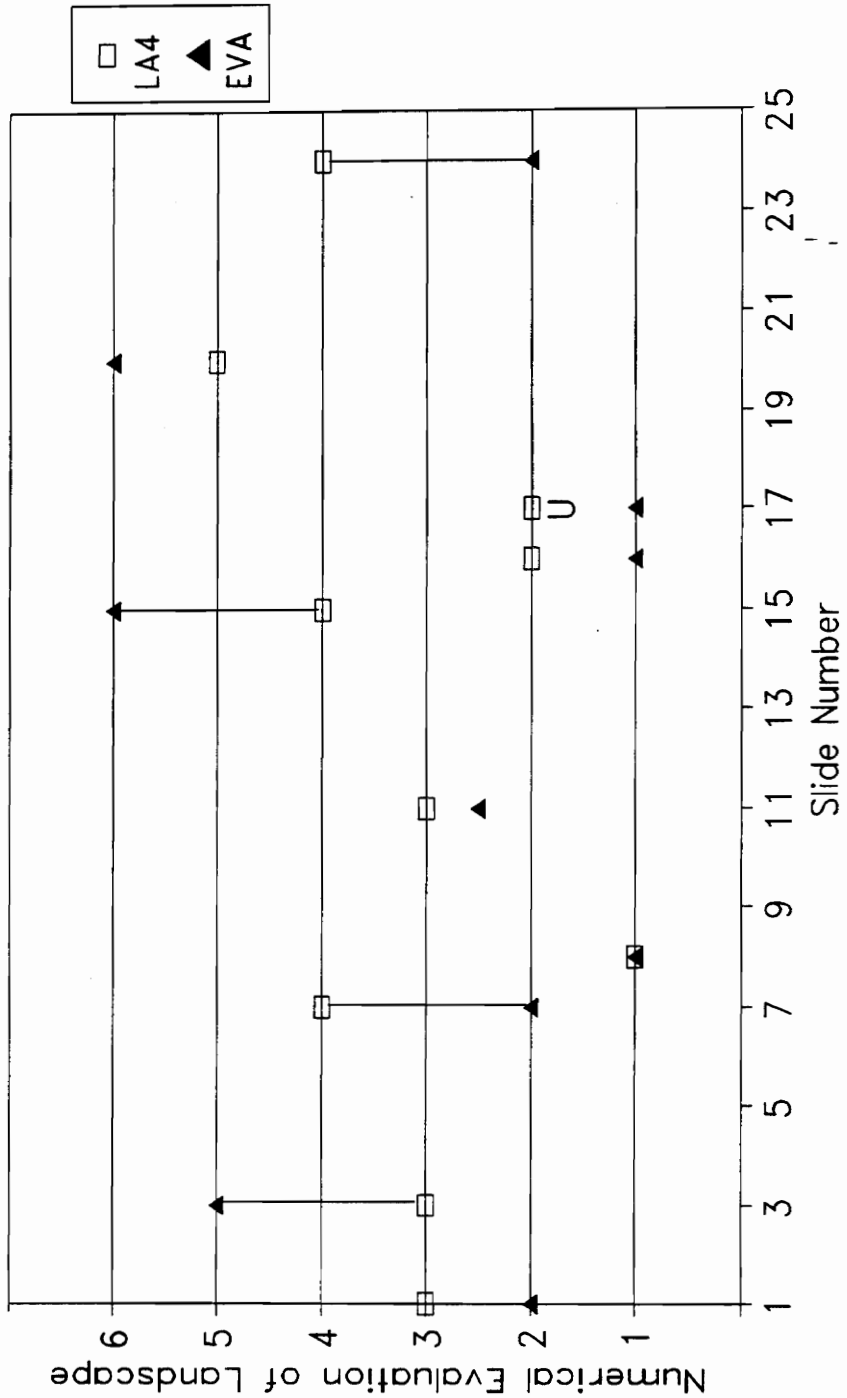

Figure 7A: LA4 vs. EVA
Slides # 1–25

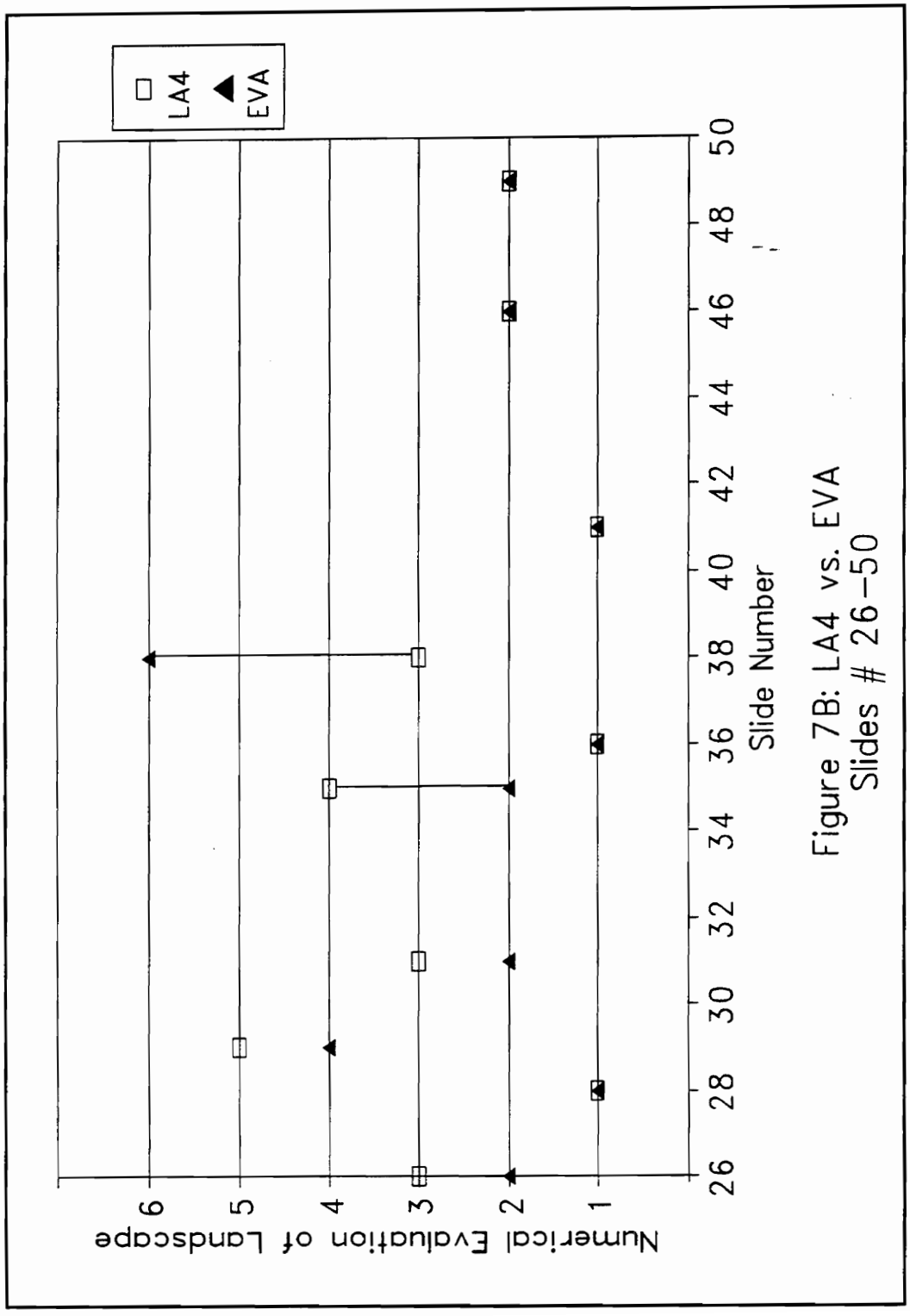

Figure 7B: LA4 vs. EVA
Slides # 26–50

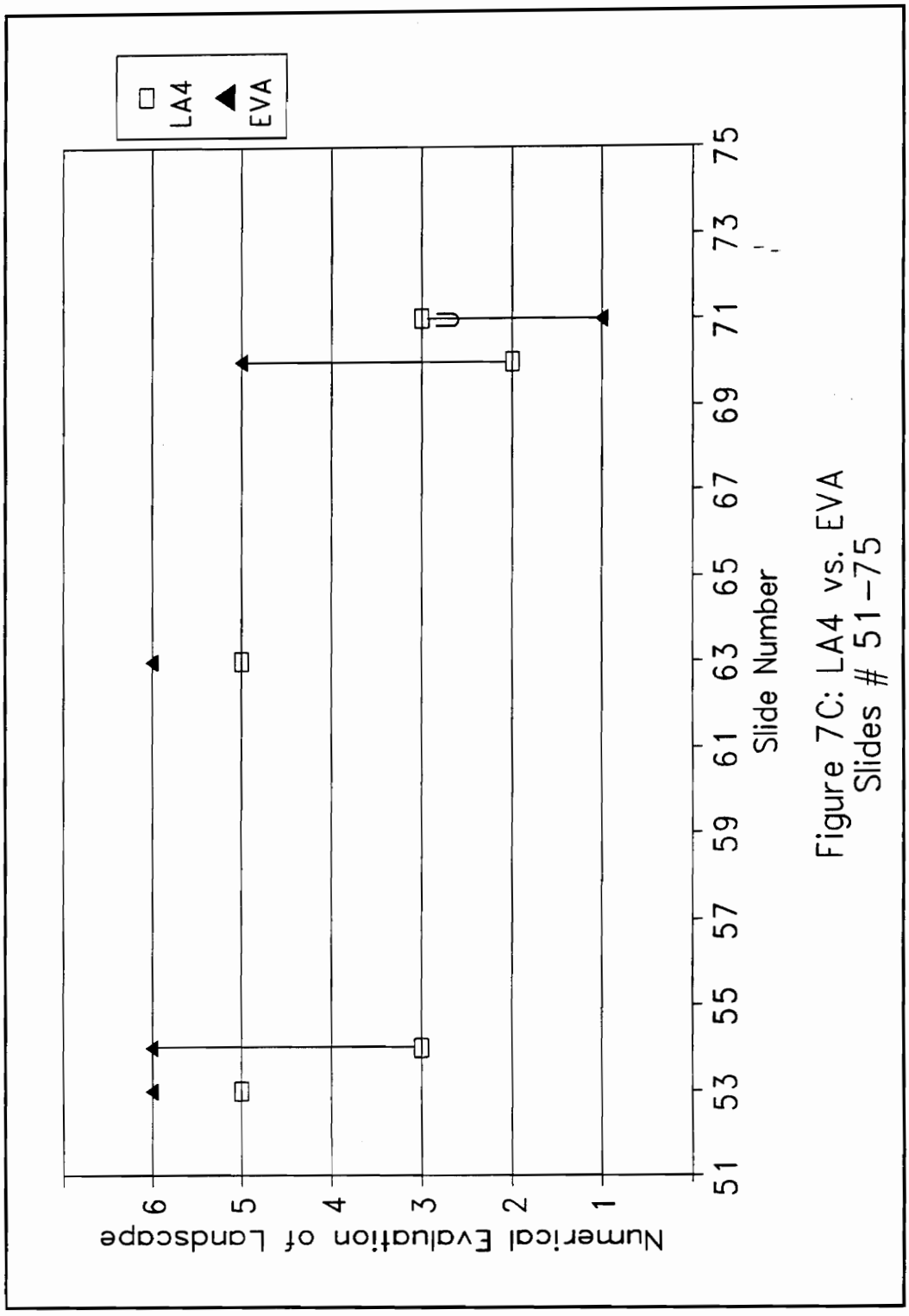

Figure 7C: LA4 vs. EVA
Slides # 51–75

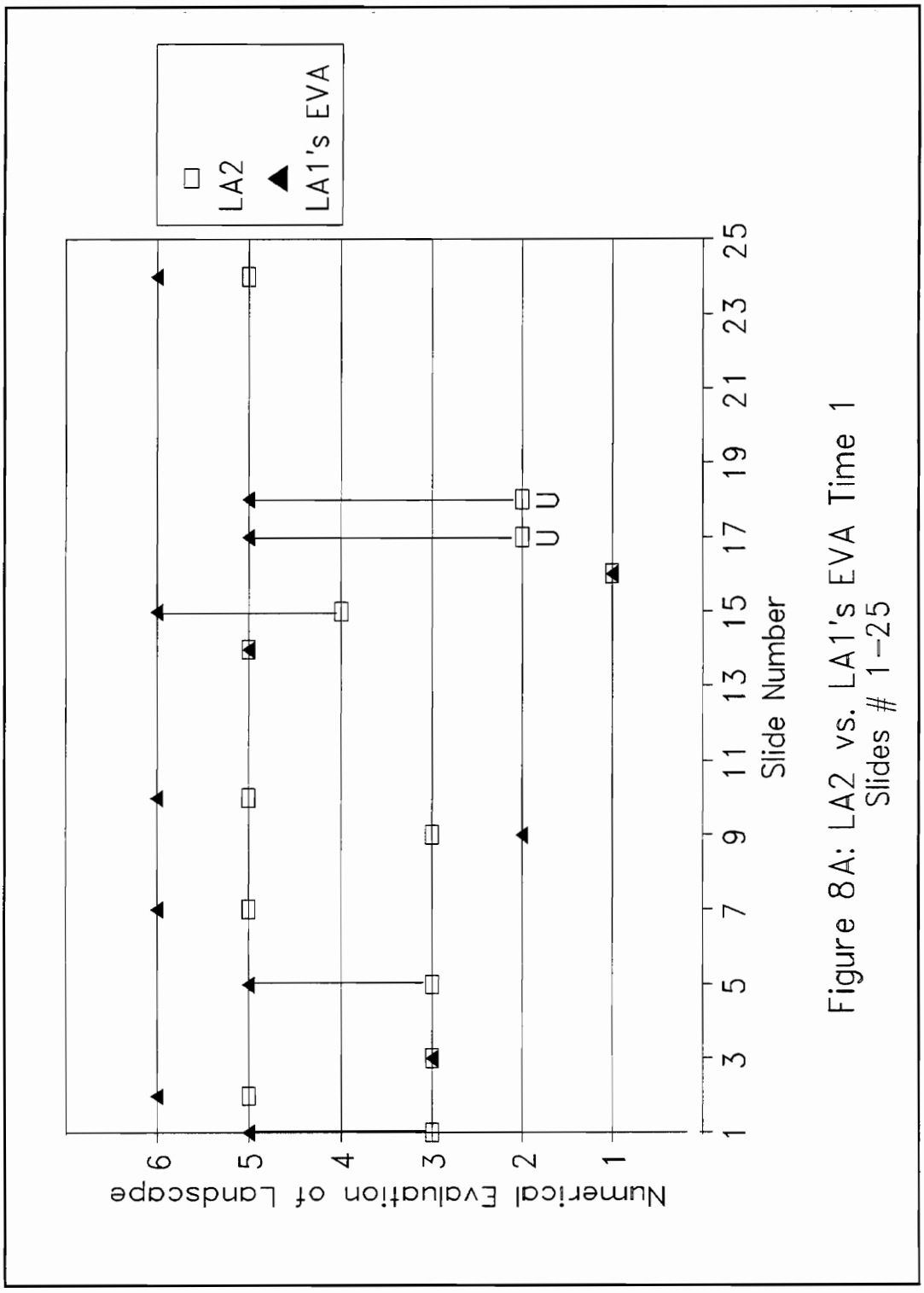

Figure 8A: LA2 vs. LA1's EVA Time 1
Slides # 1–25

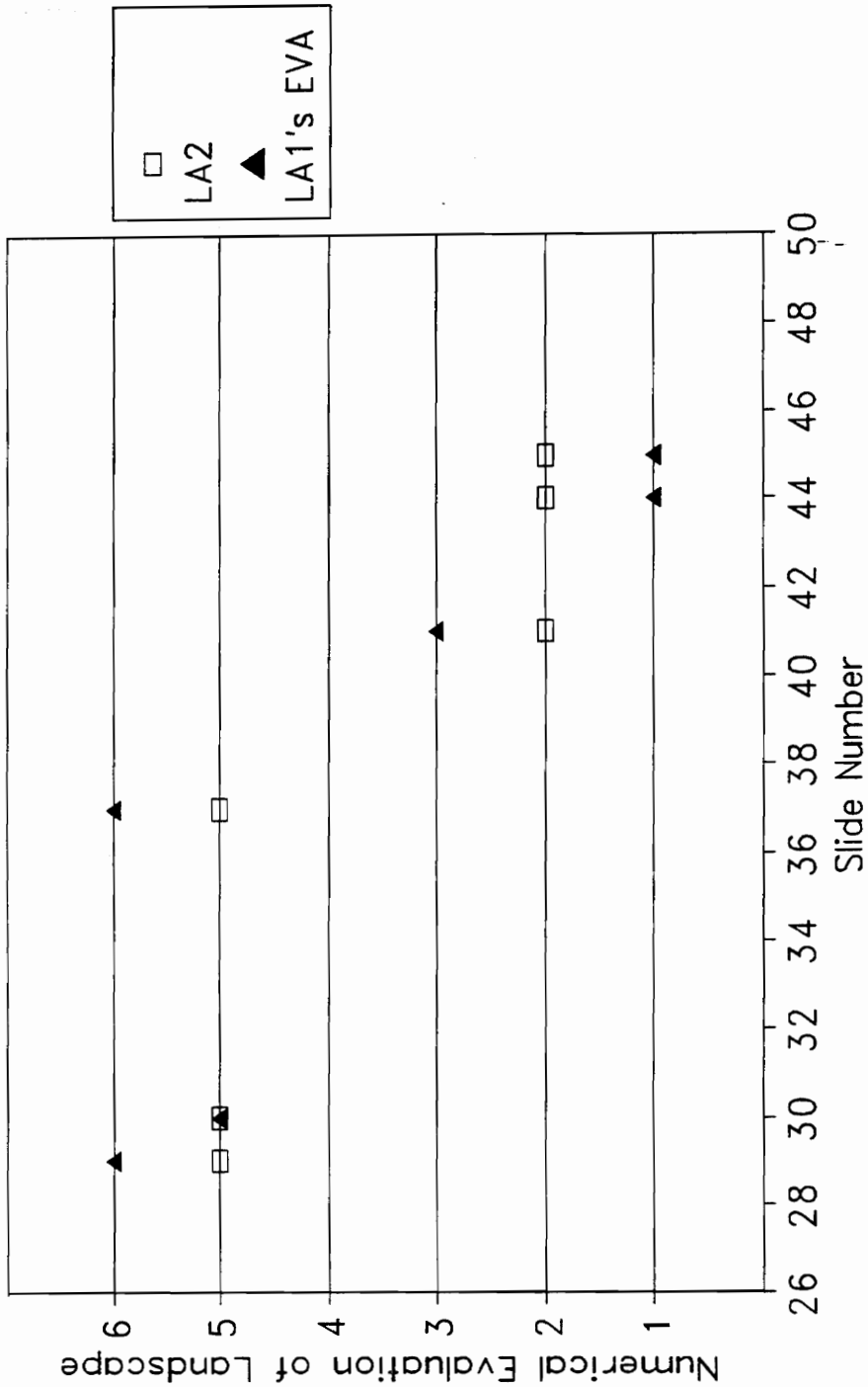

Figure 8B: LA2 vs. LA1's EVA Time 1
Slides # 26–50

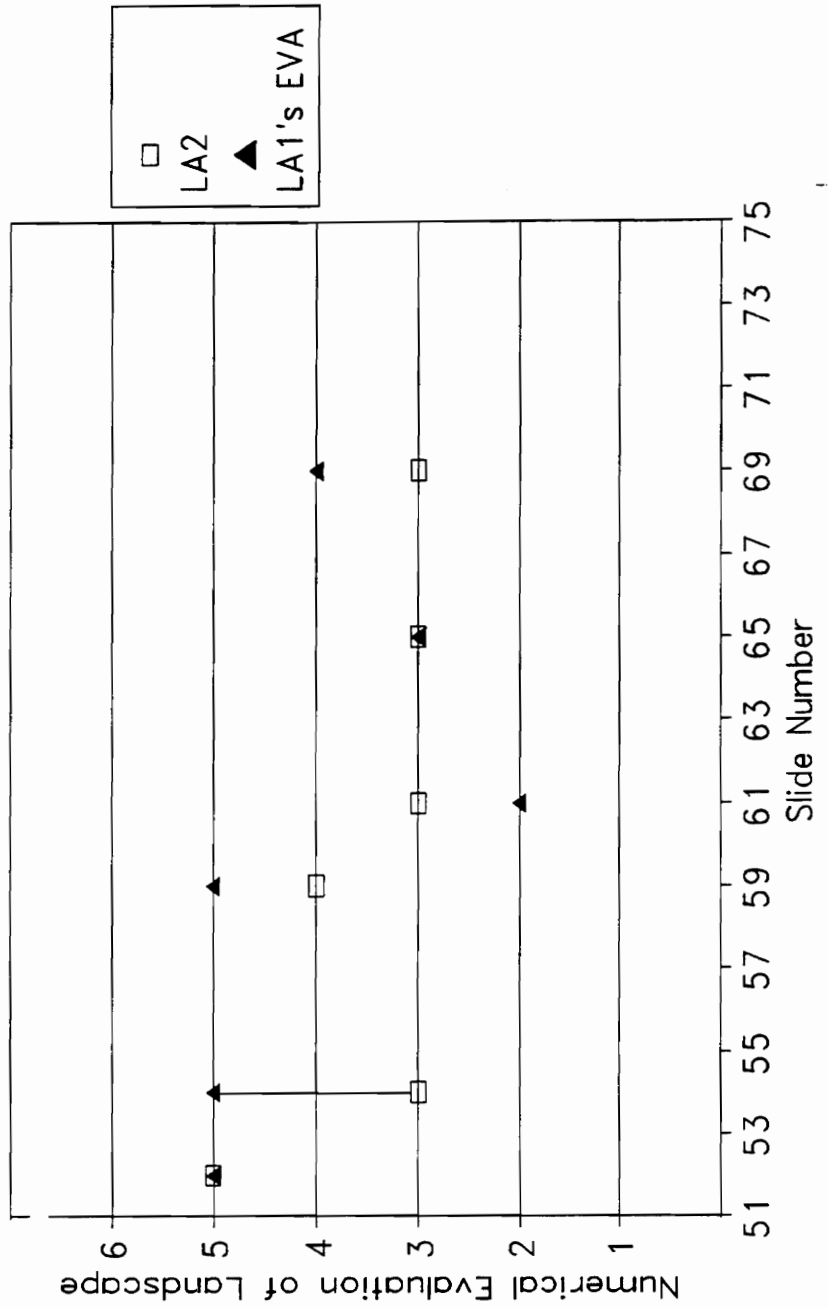

Figure 8C: LA2 vs. LA1's EVA Time 1
Slides # 51–75

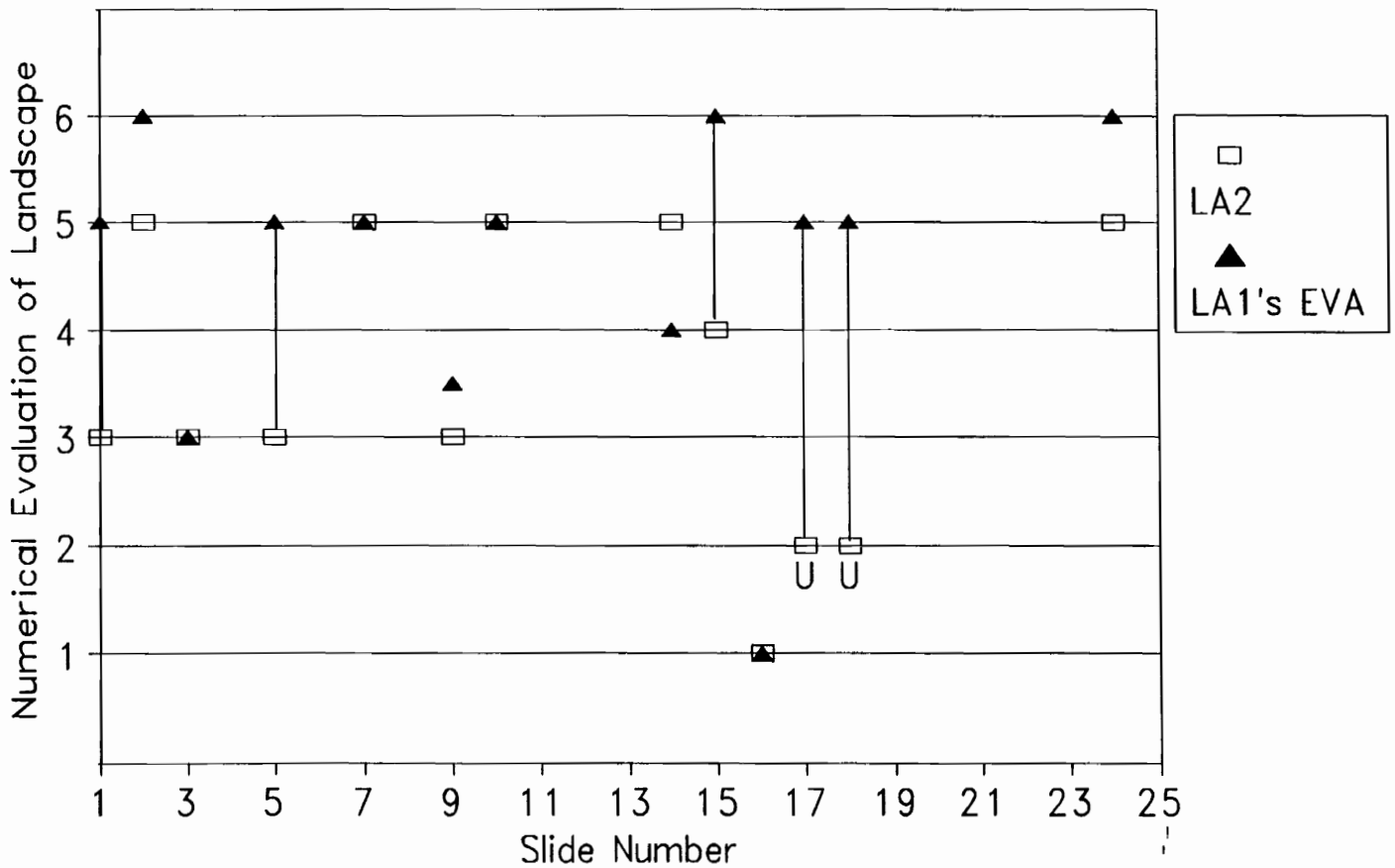

Figure 8D: LA2 vs. LA1's EVA Time 2
Slides # 1-25

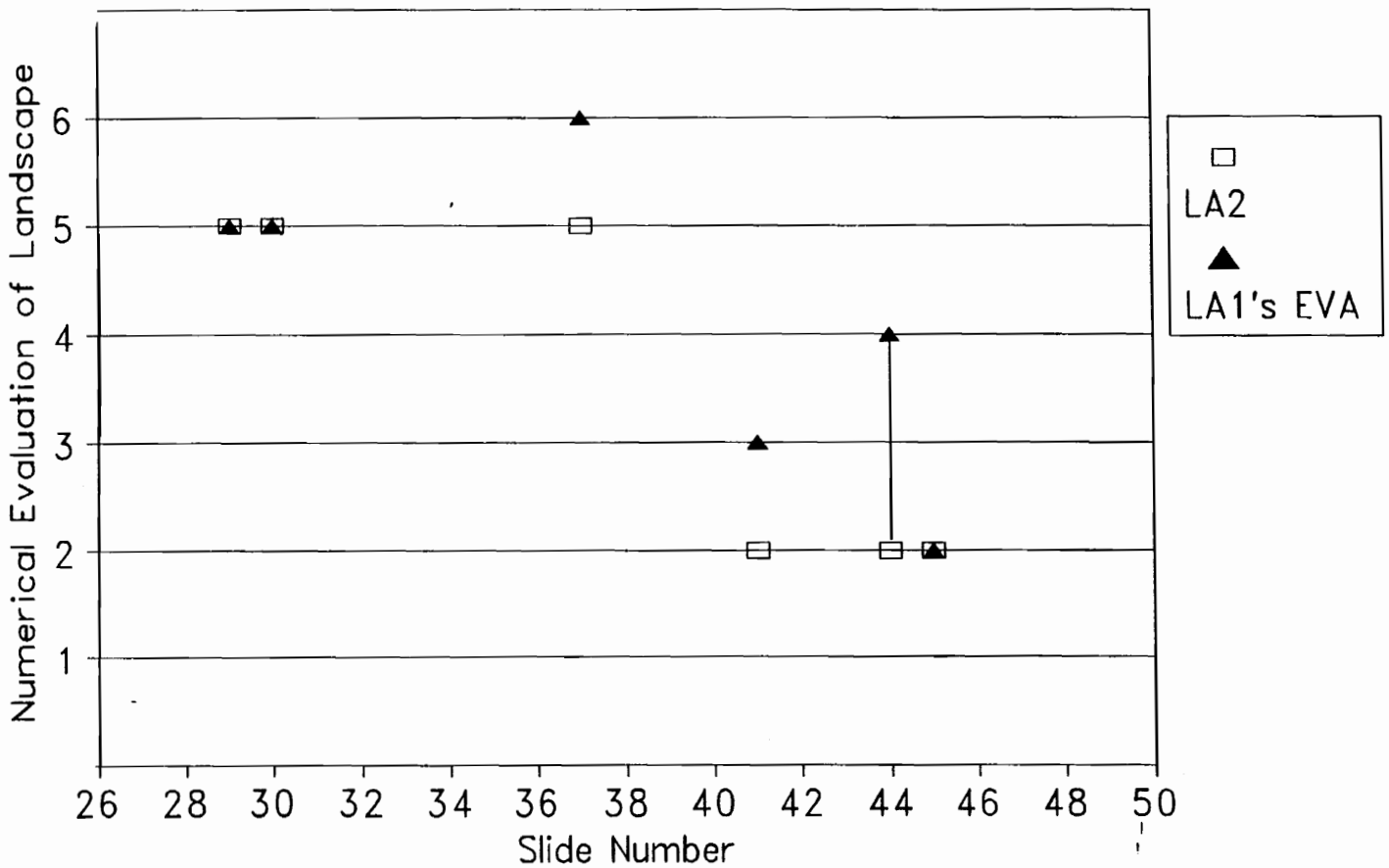

Figure 8E: LA2 vs. LA1's EVA Time 2
Slides # 26–50

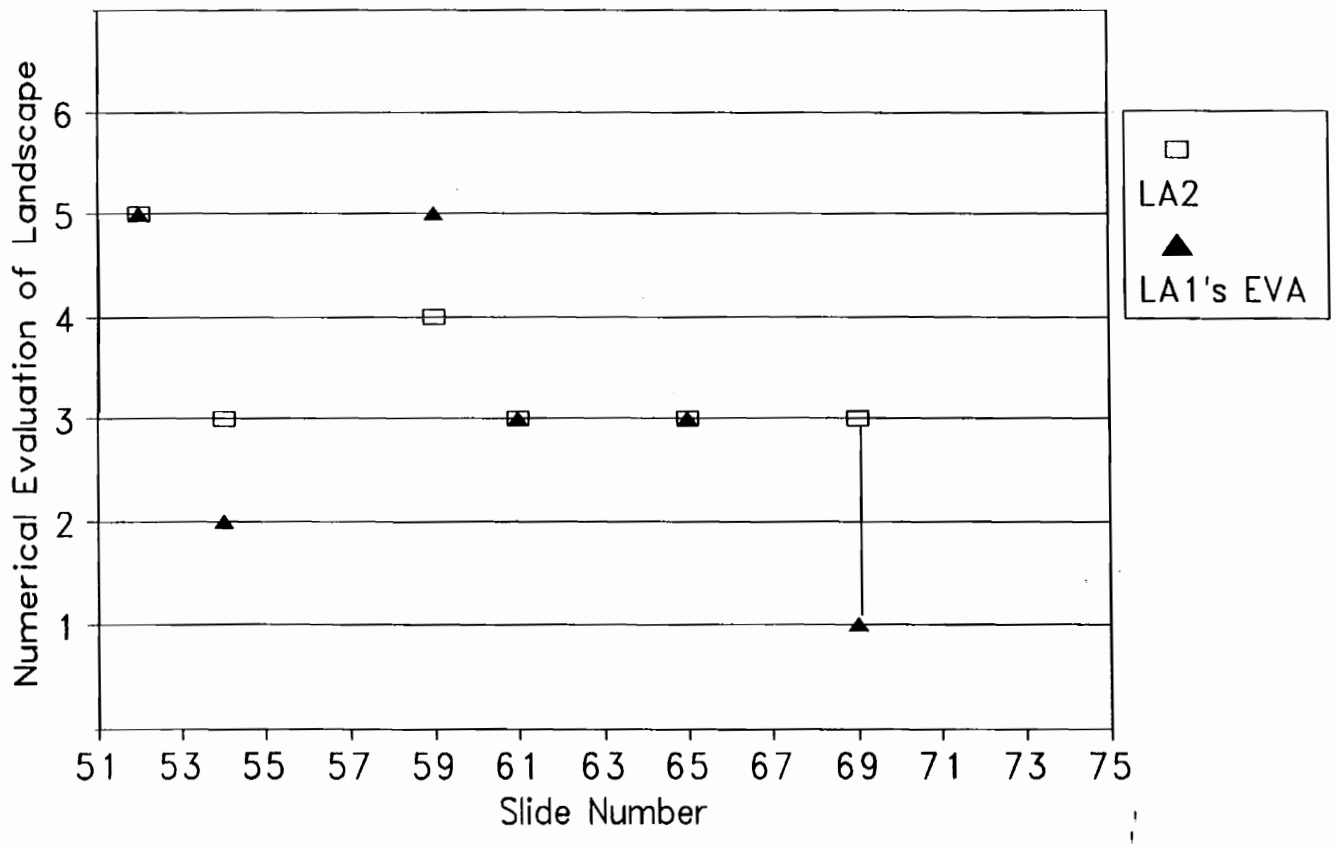

Figure 8F: LA2 vs. LA1's EVA Time 2
Slides # 51-75

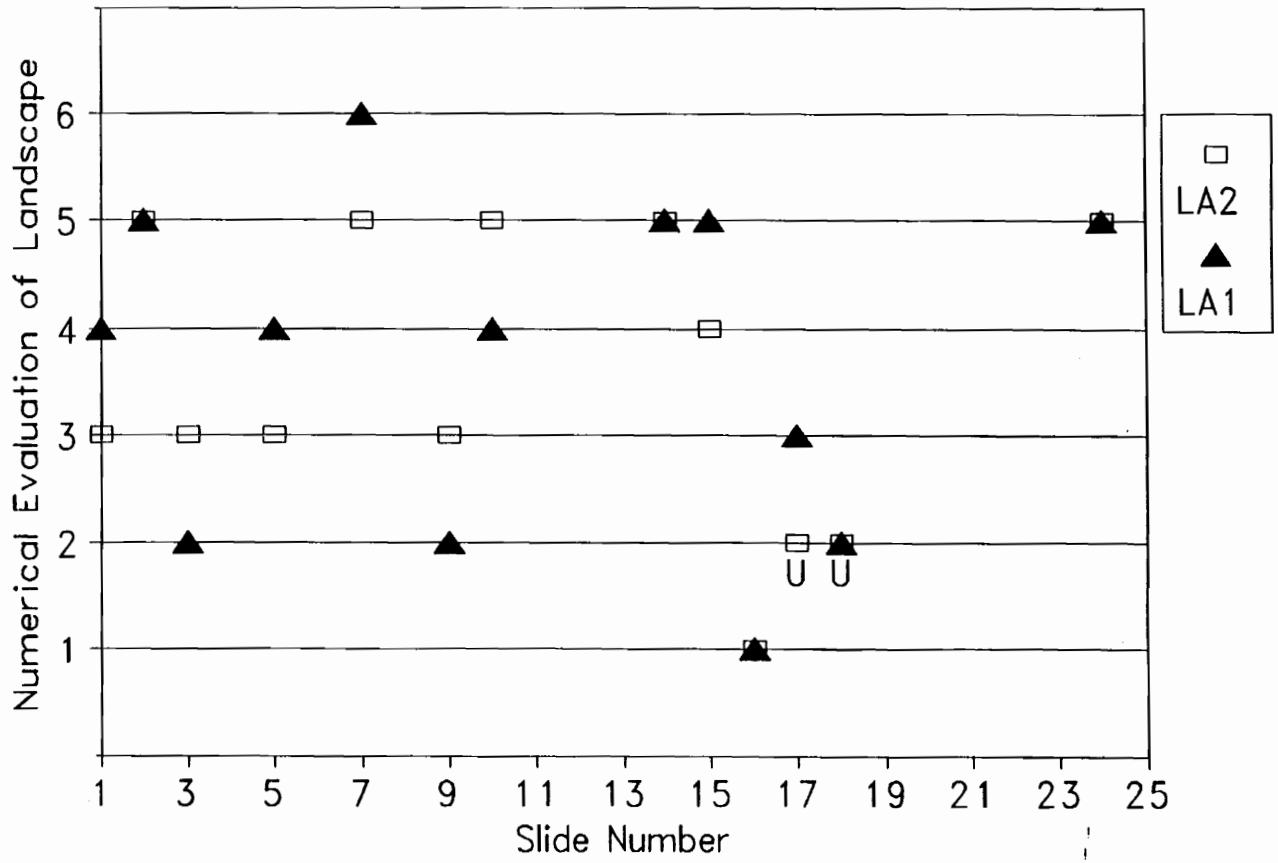

Figure 9A: LA2 vs. LA1 Time 1
Slides # 1-25

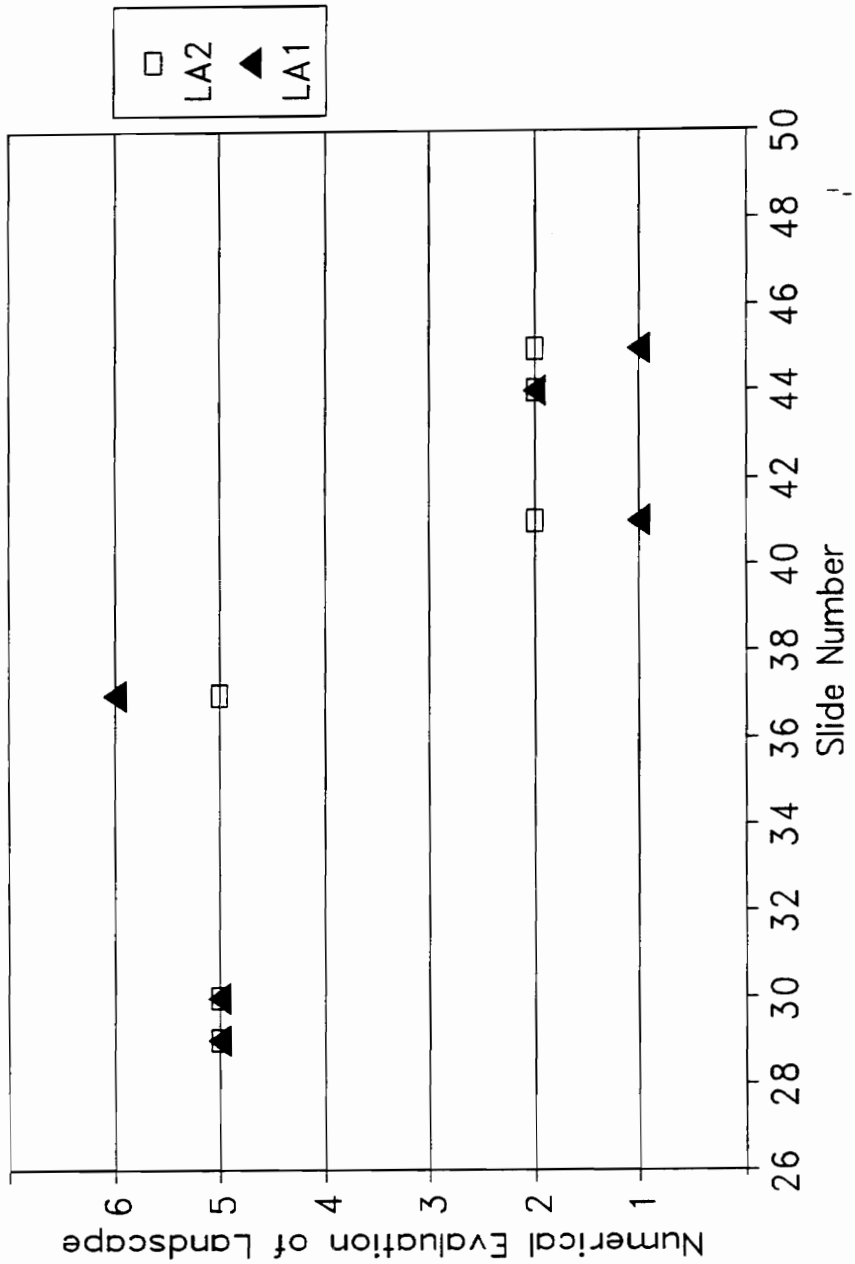

Figure 9B: LA2 vs. LA1 Time 1
Slides # 26–50

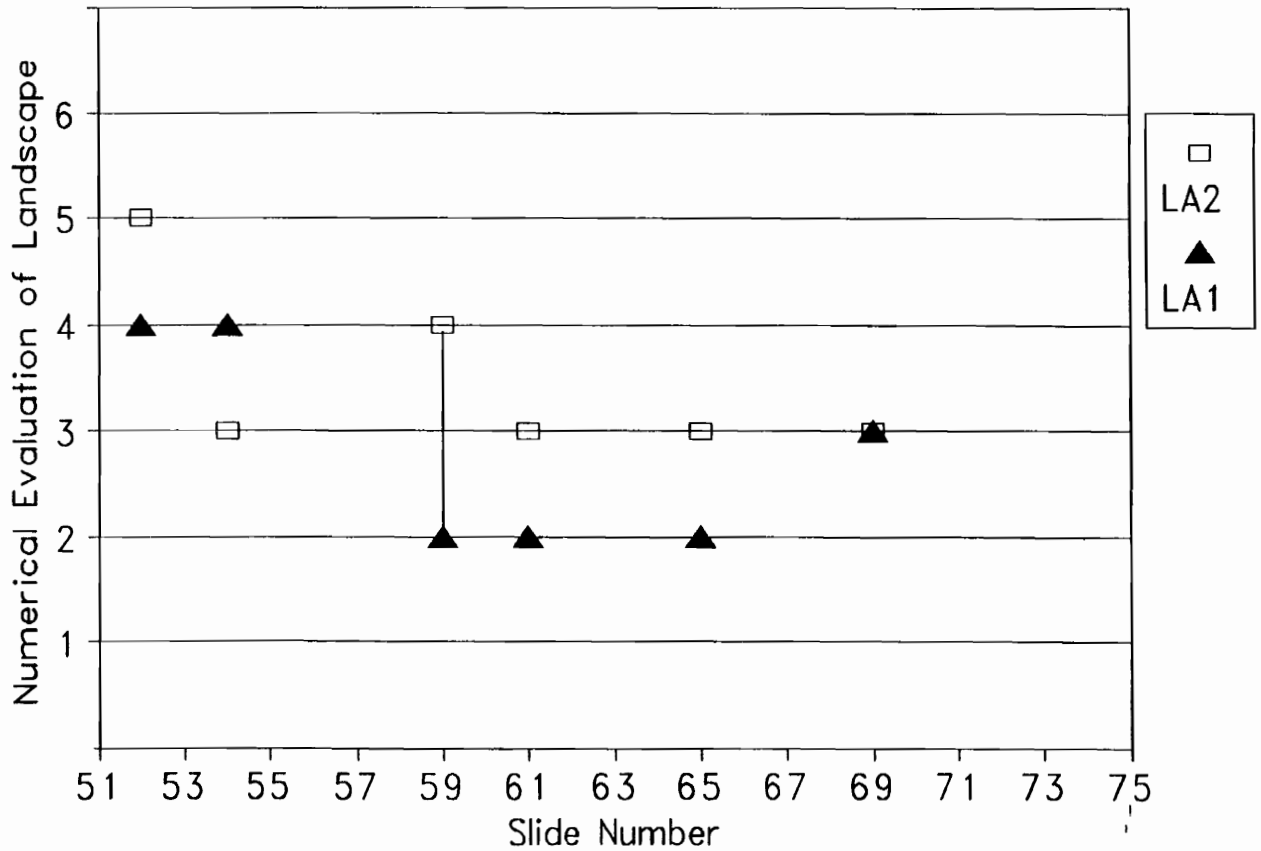

Figure 9C: LA2 vs. LA1 Time 1
Slides # 51-75

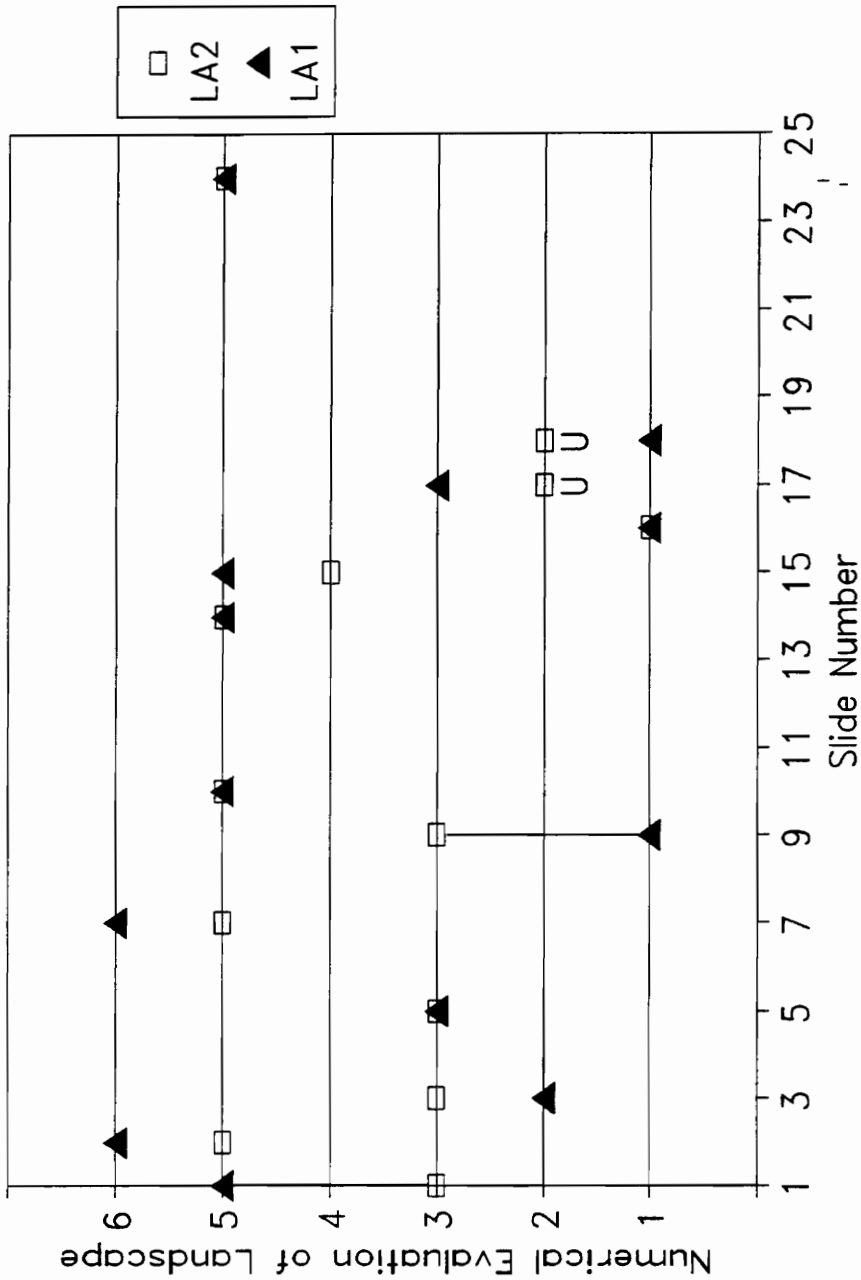

Figure 9D: LA2 vs. LA1 Time 2
Slides # 1–25

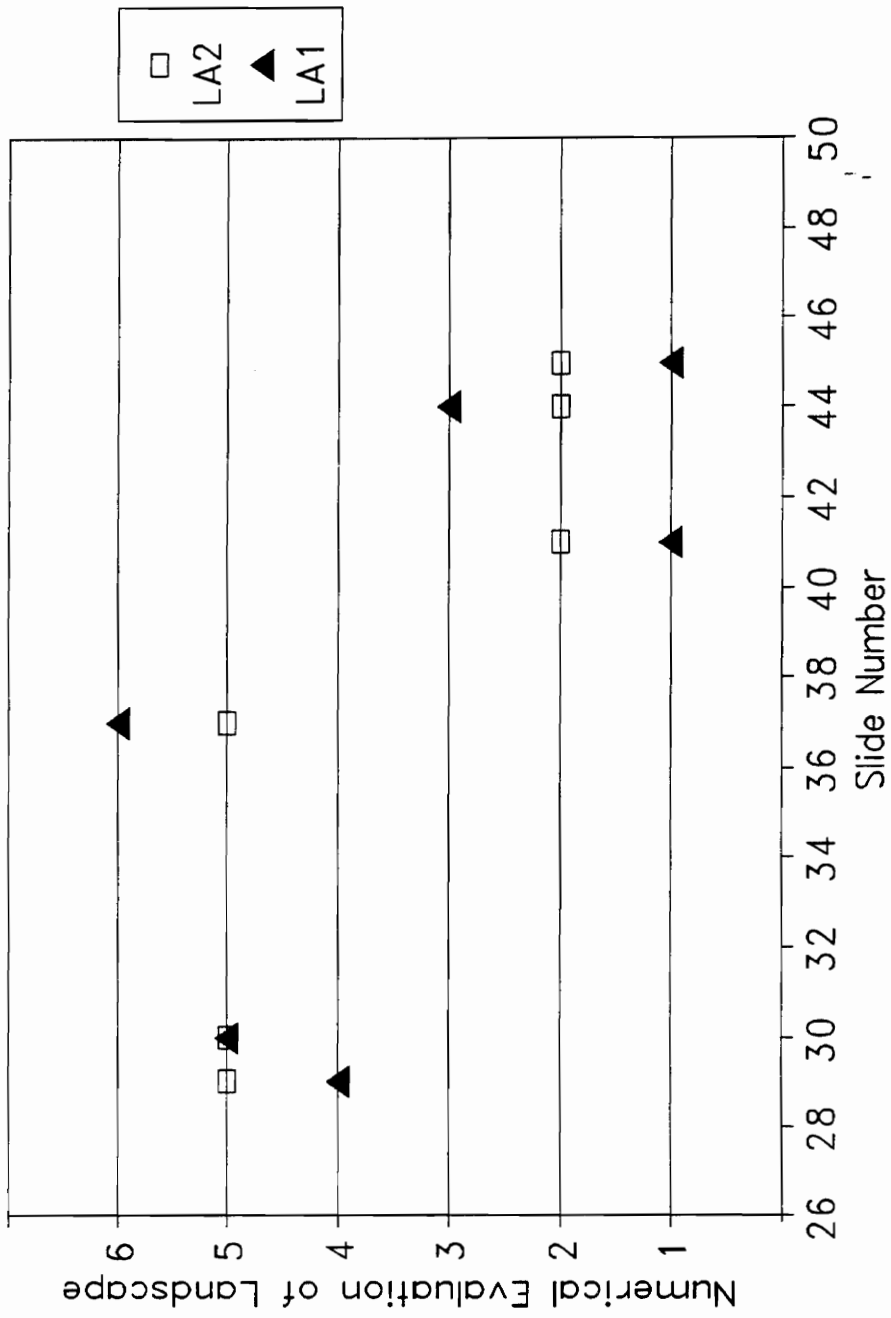

Figure 9E: LA2 vs. LA1 Time 2
Slides # 26–50

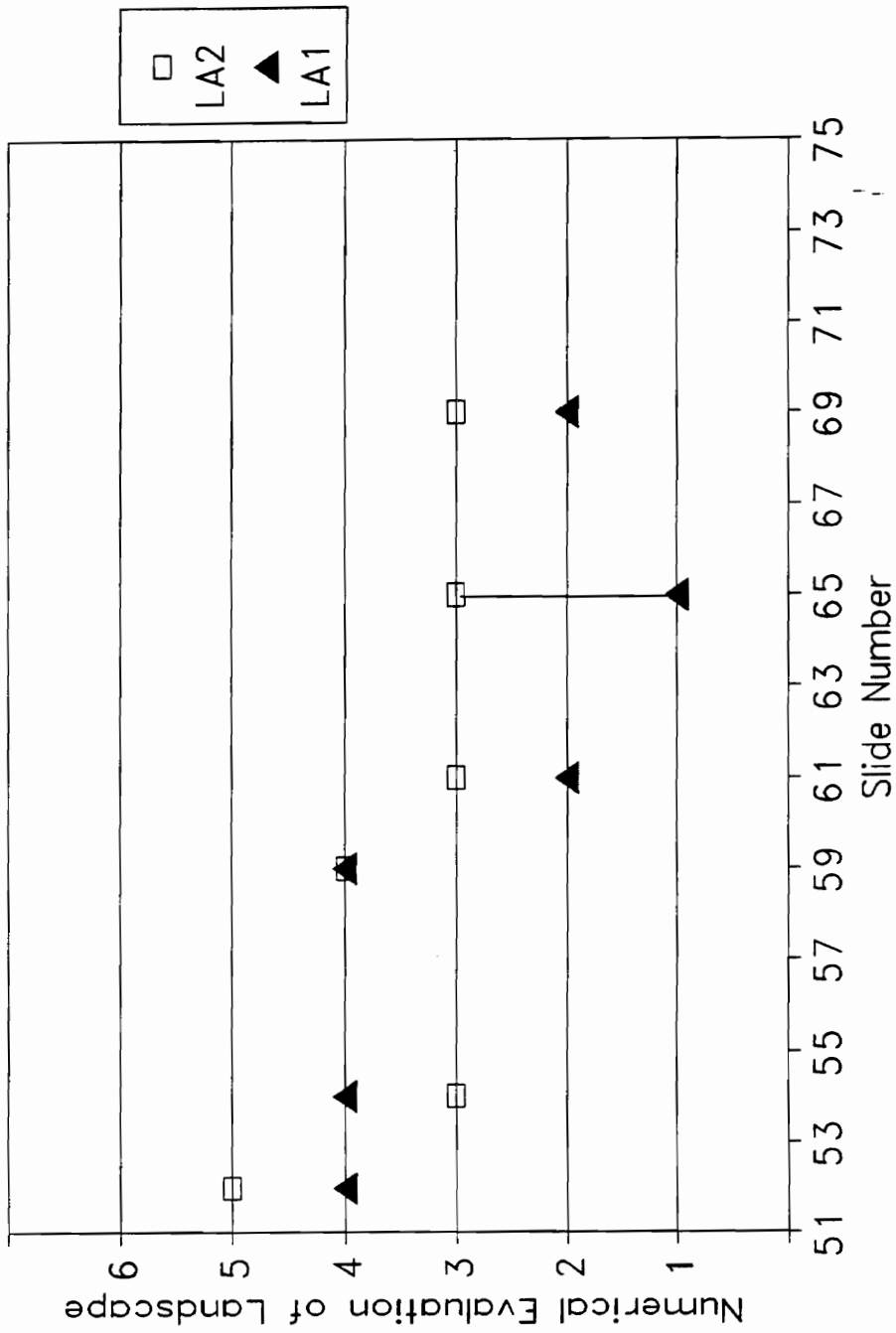

Figure 9F: LA2 vs. LA1 Time 2
Slides # 51–75

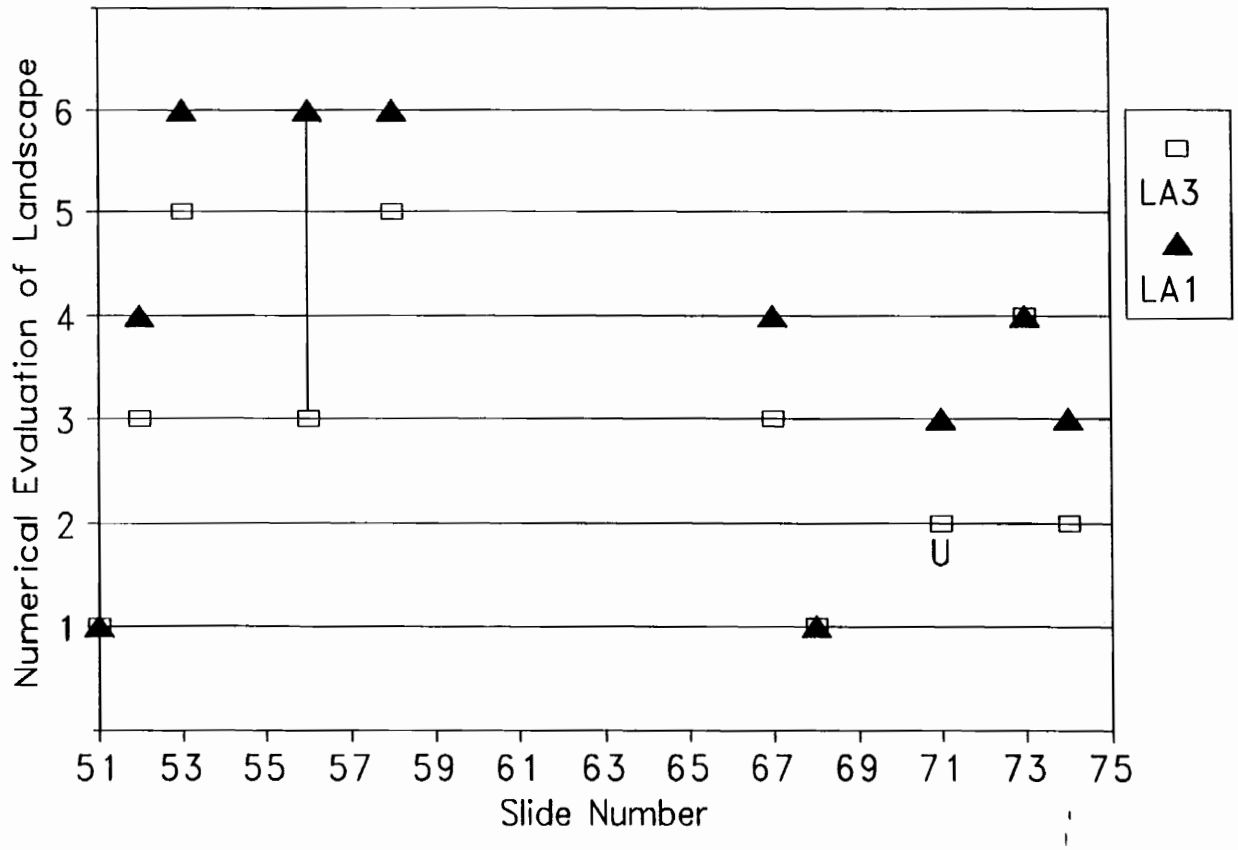

Figure 10A: LA3 vs. LA1 Time 2
Slides # 1-25

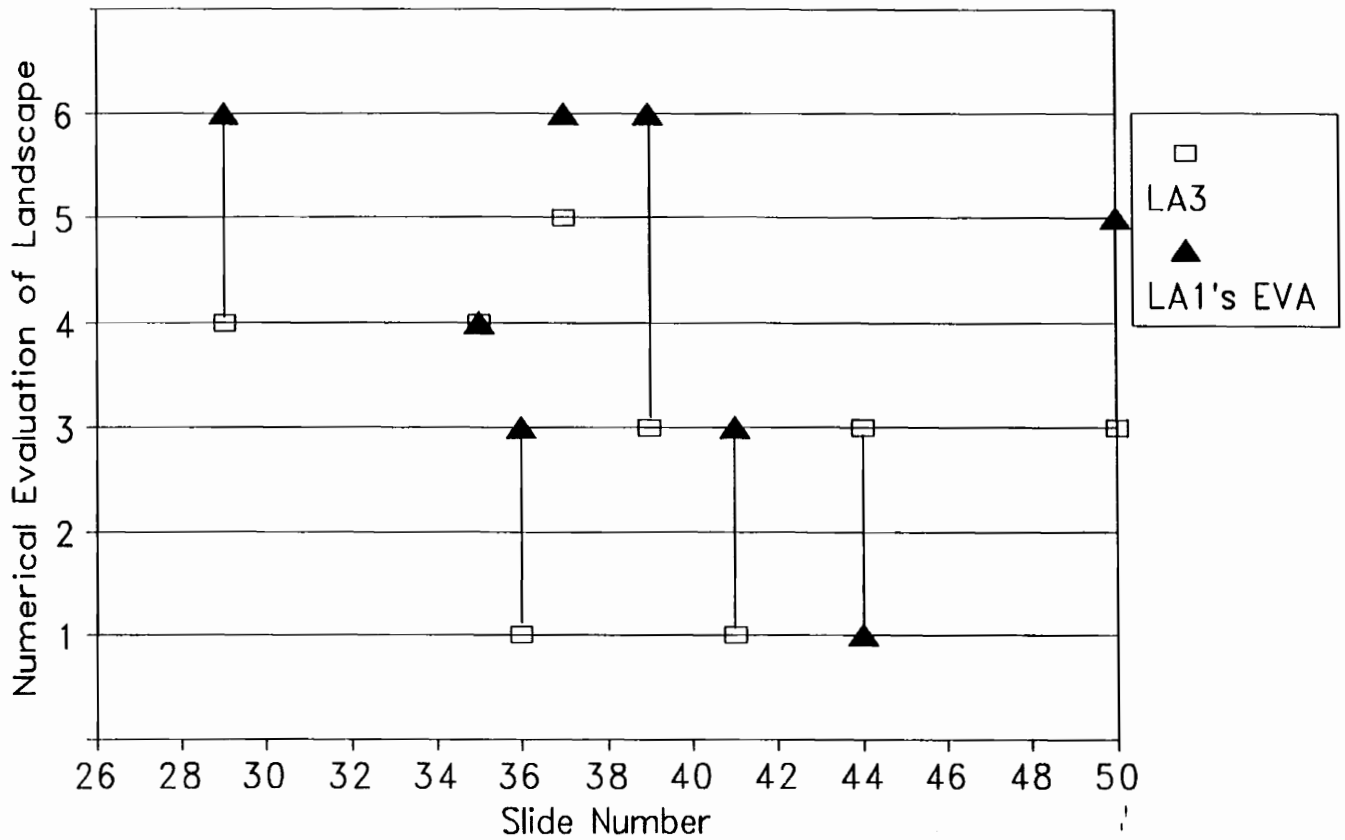

Figure 10B: LA3 vs. LA1's EVA Time 1
Slides # 26–50

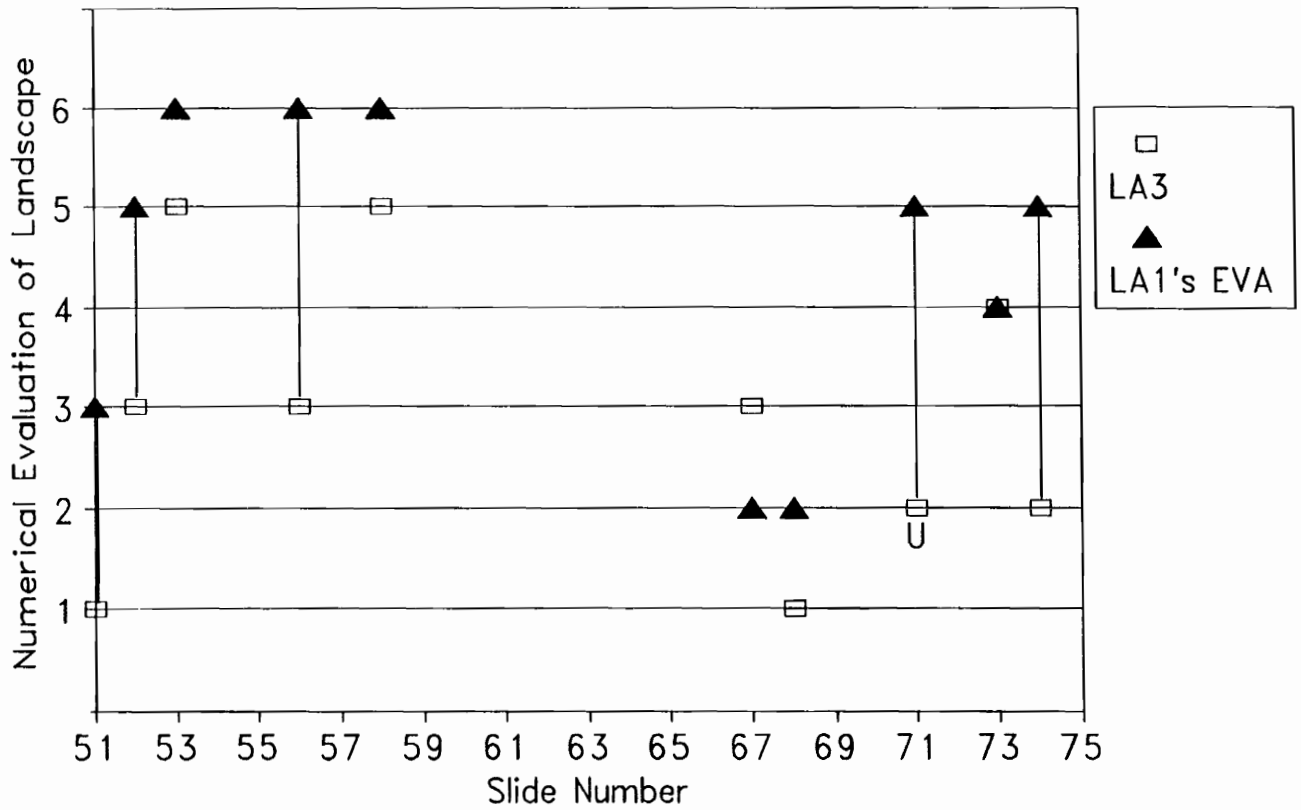

Figure 10C: LA3 vs. LA1's EVA Time 1
Slides # 51-75

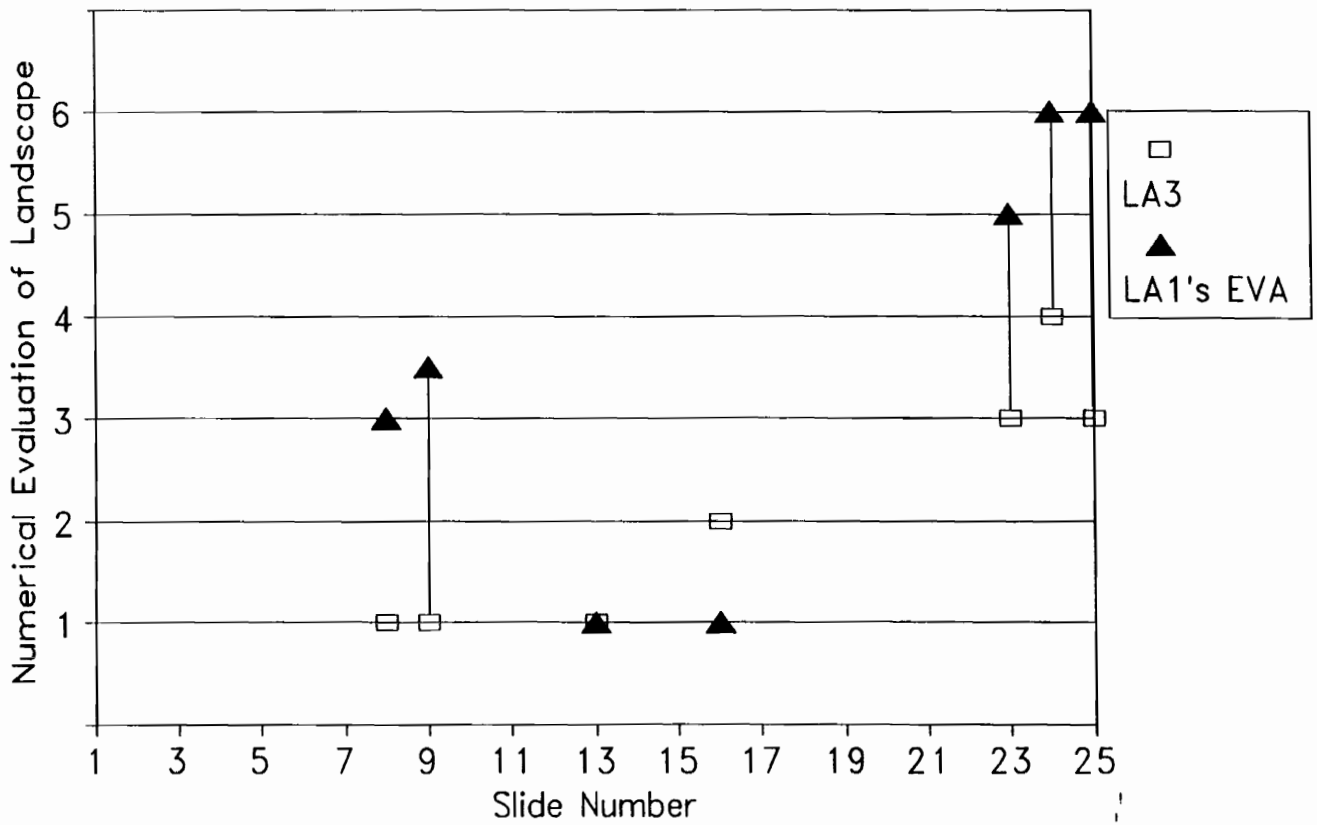

Figure 10D: LA3 vs. LA1's EVA Time 2
Slides # 1–25

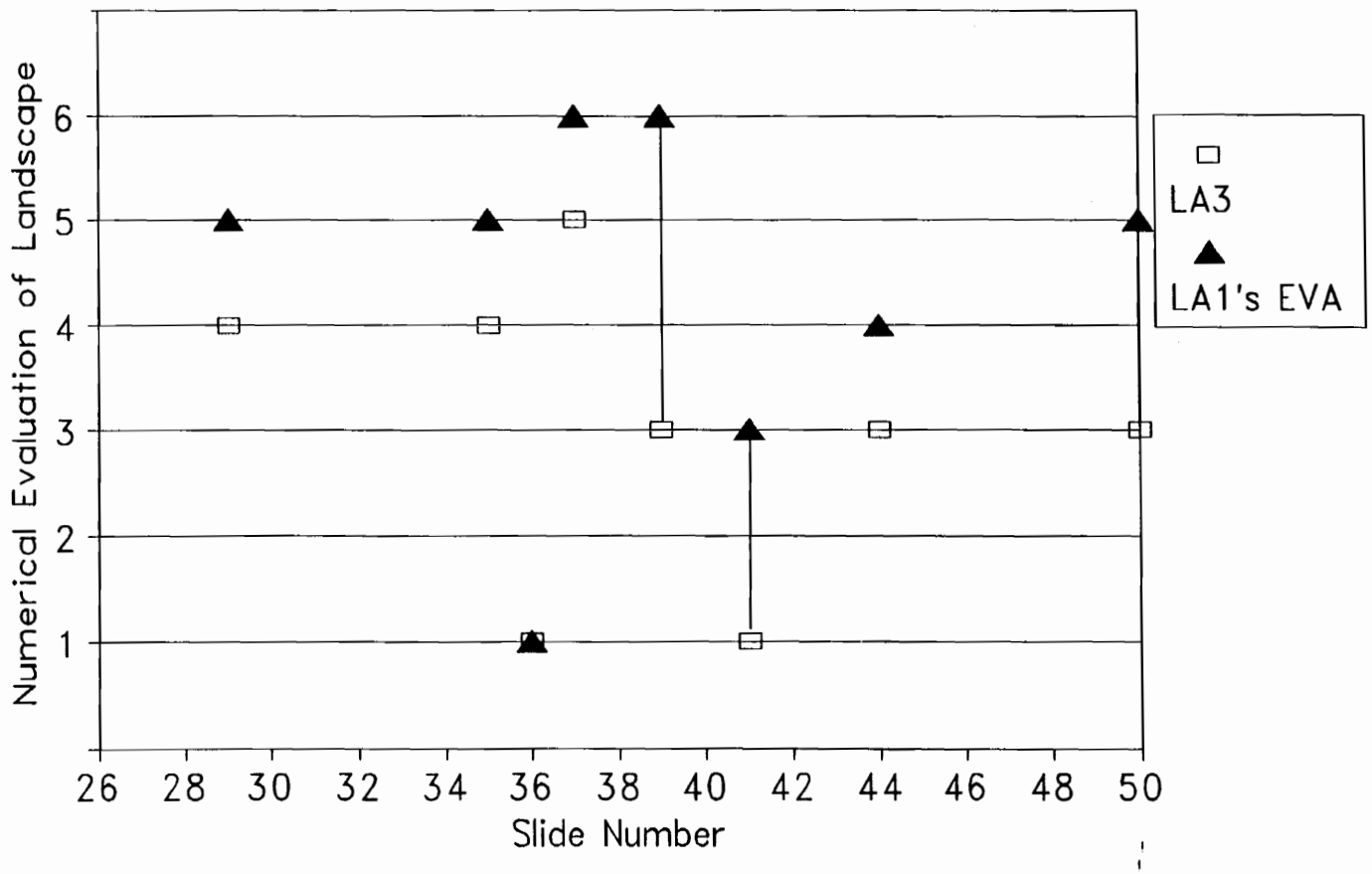

Figure 10E: LA3 vs. LA1's EVA Time 2
Slides # 26–50

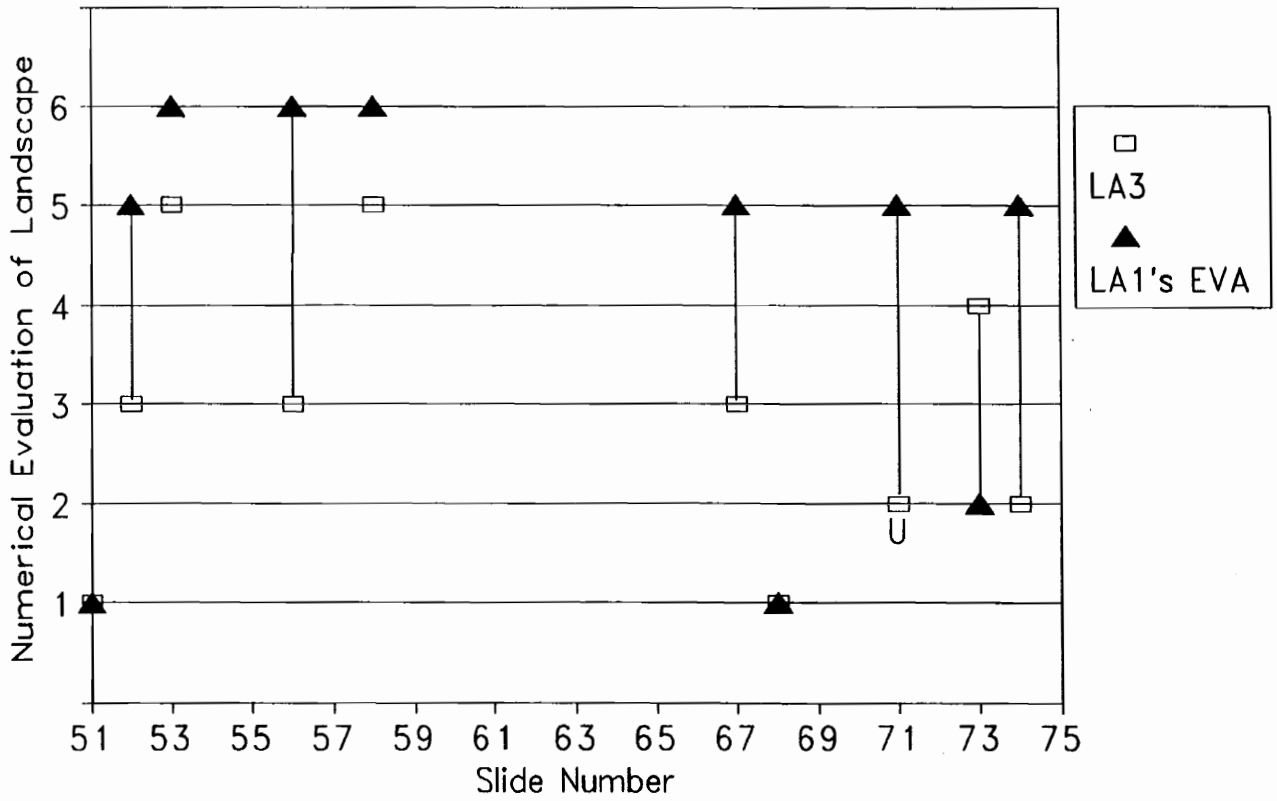

Figure 10F: LA3 vs. LA1's EVA Time 2
Slides # 51–75

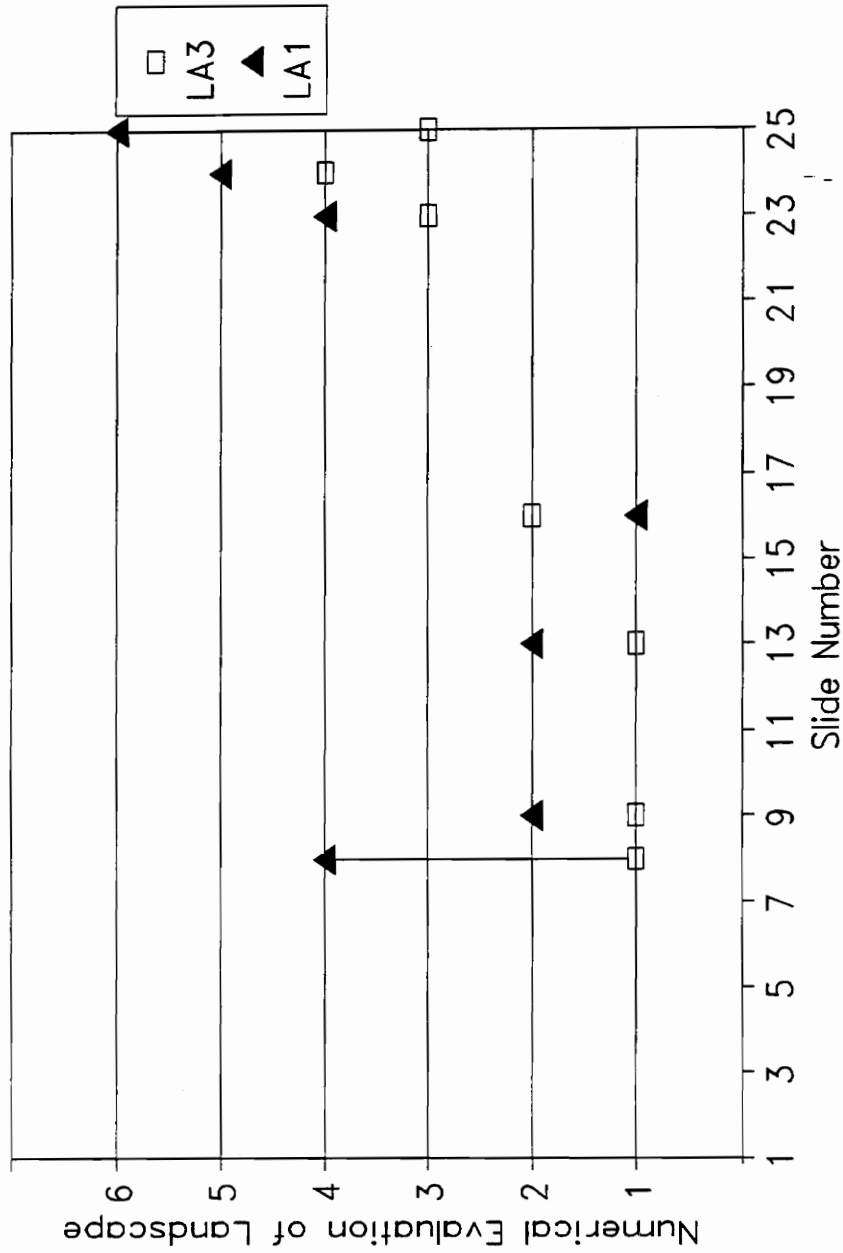

Figure 11A: LA3 vs. LA1 Time 1
Slides # 1–25

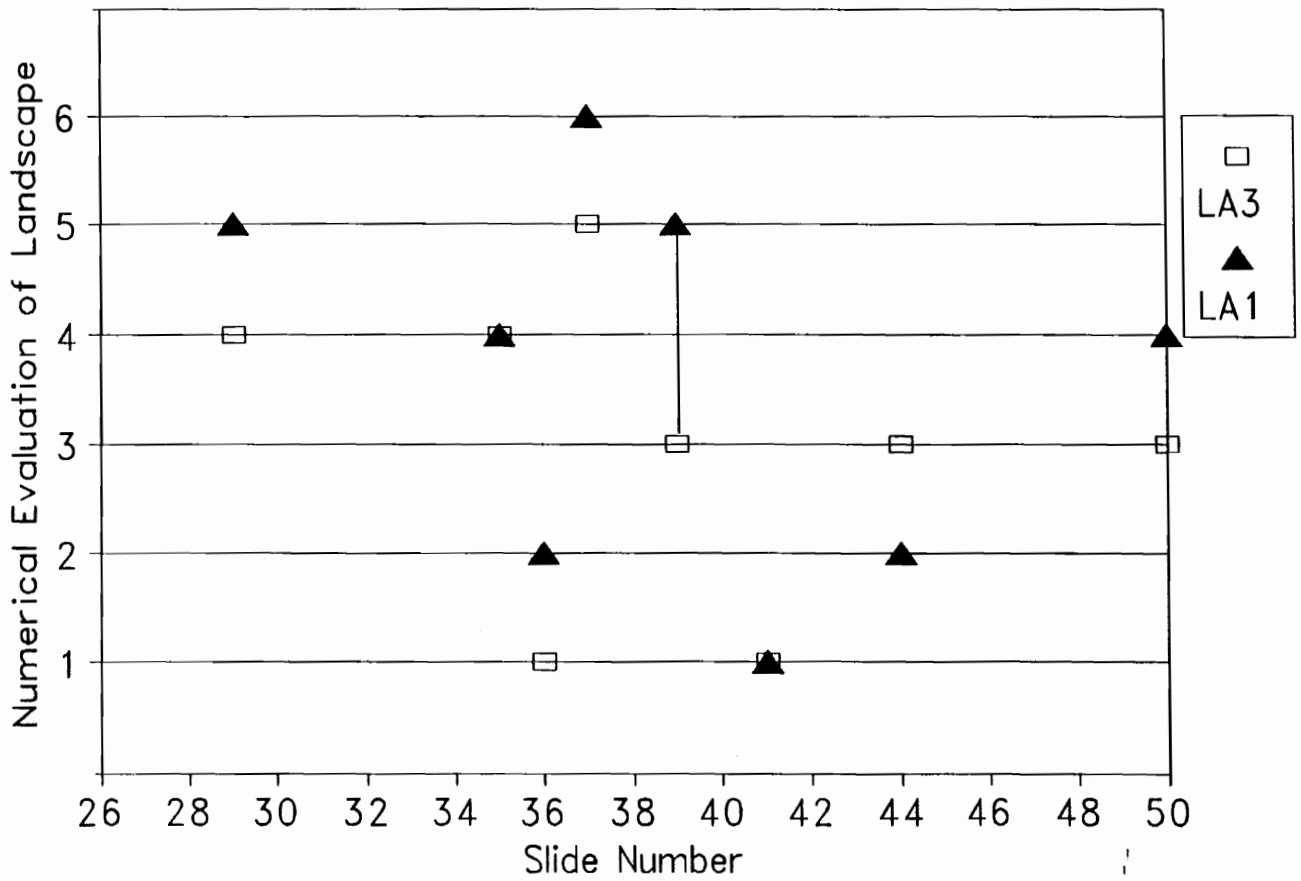

Figure 11B: LA3 vs. LA1 Time 1
Slides # 26–50

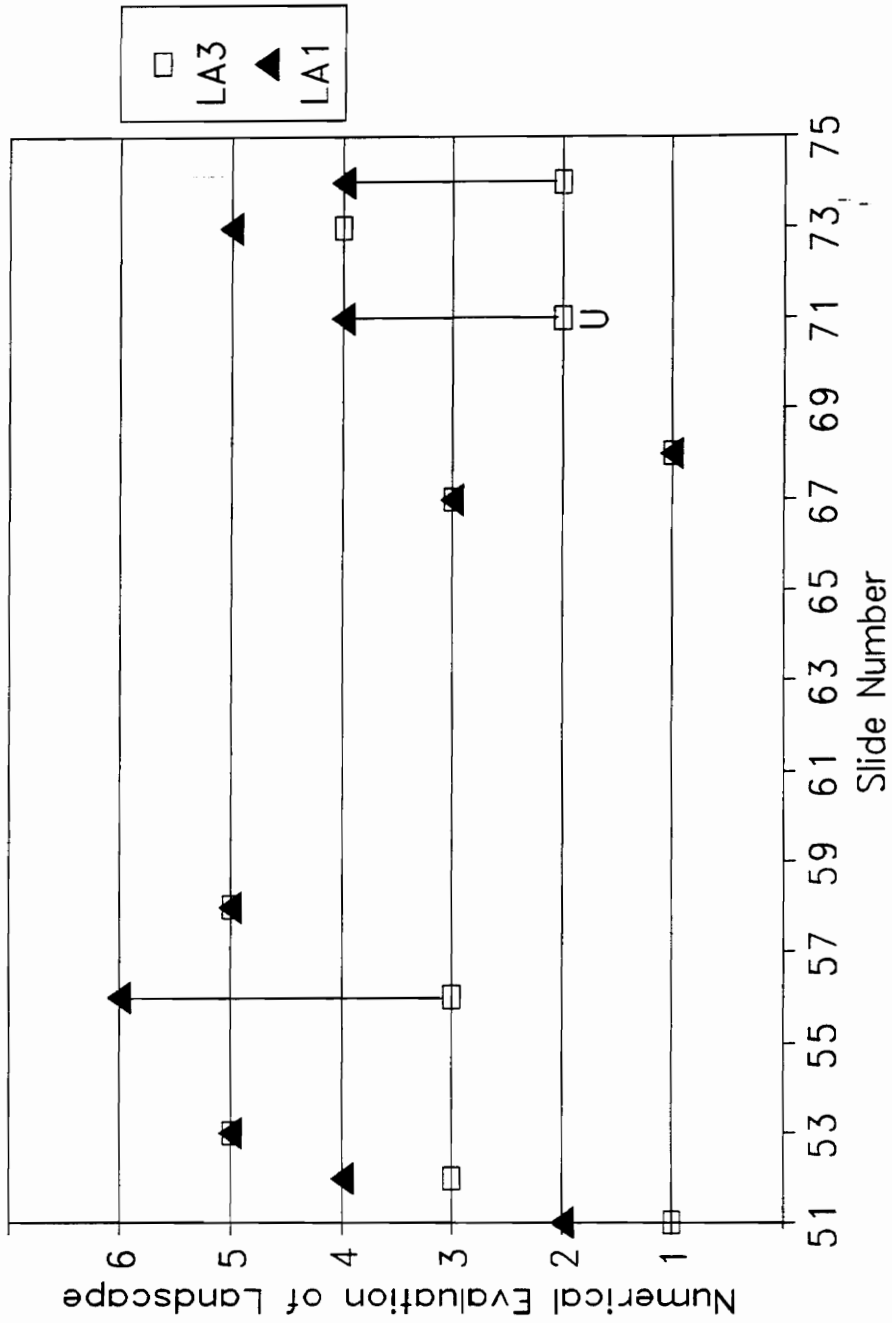

Figure 11C: LA3 vs. LA1 Time 1
Slides # 51–75

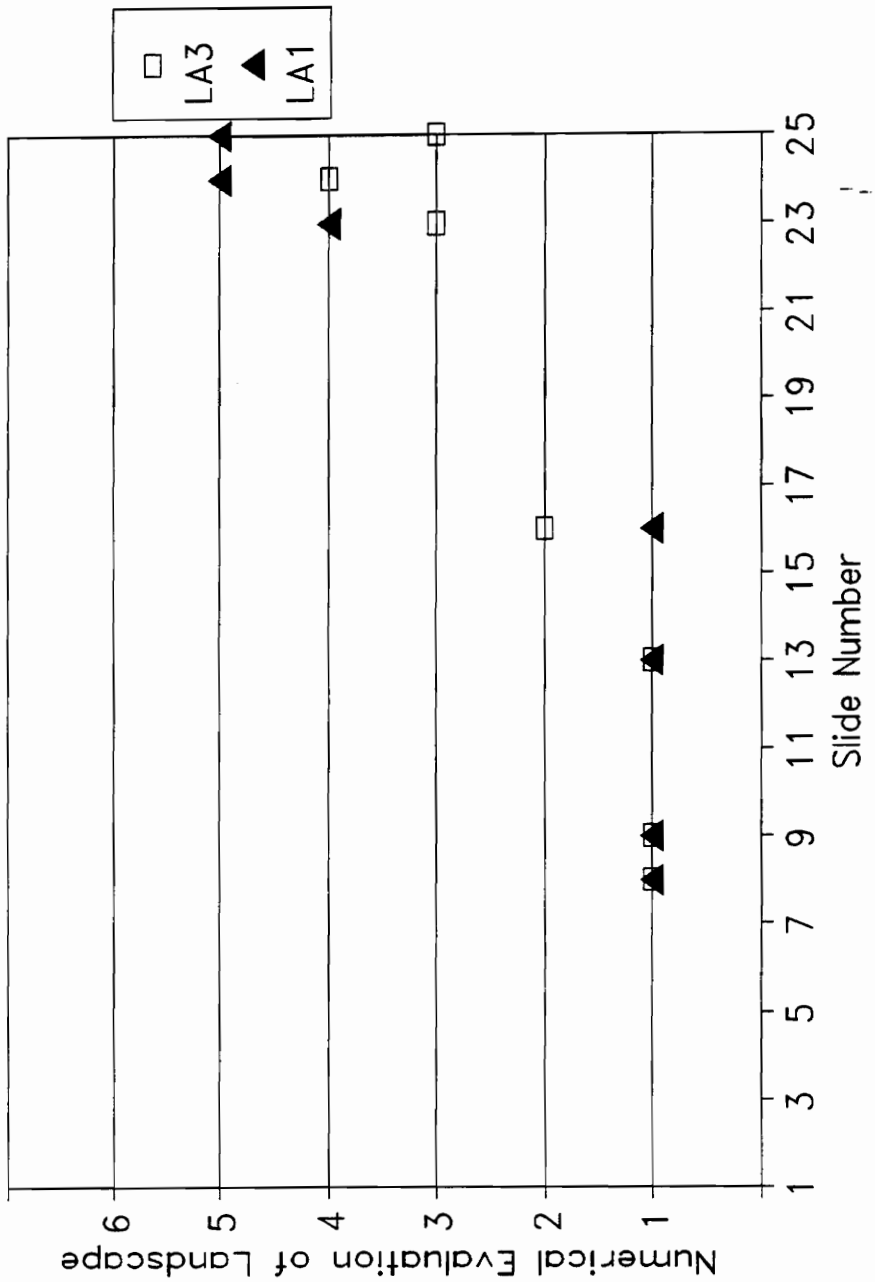

Figure 11D: LA3 vs. LA1 Time 2
Slides # 1–25

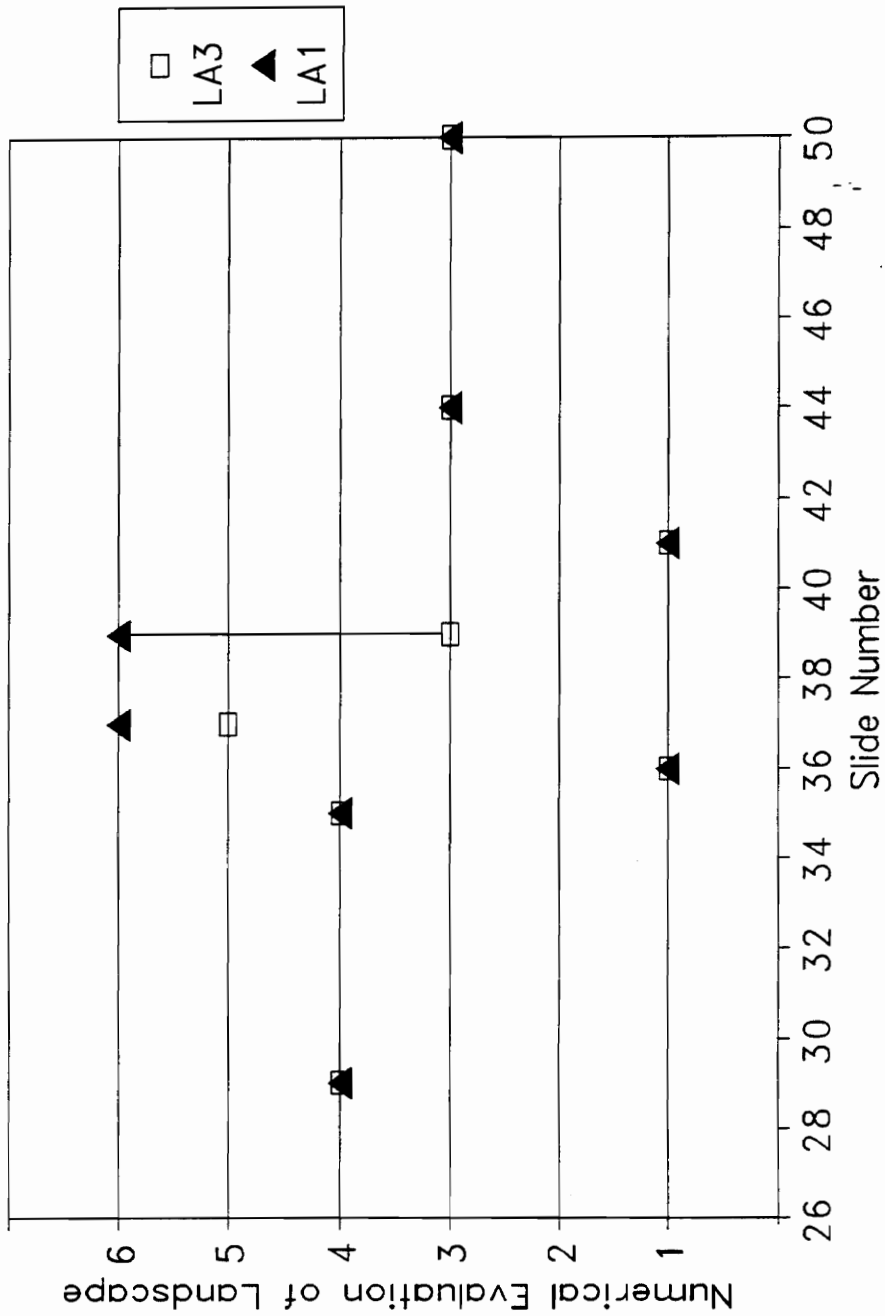

Figure 11E: LA3 vs. LA1 Time 2
Slides # 26–50

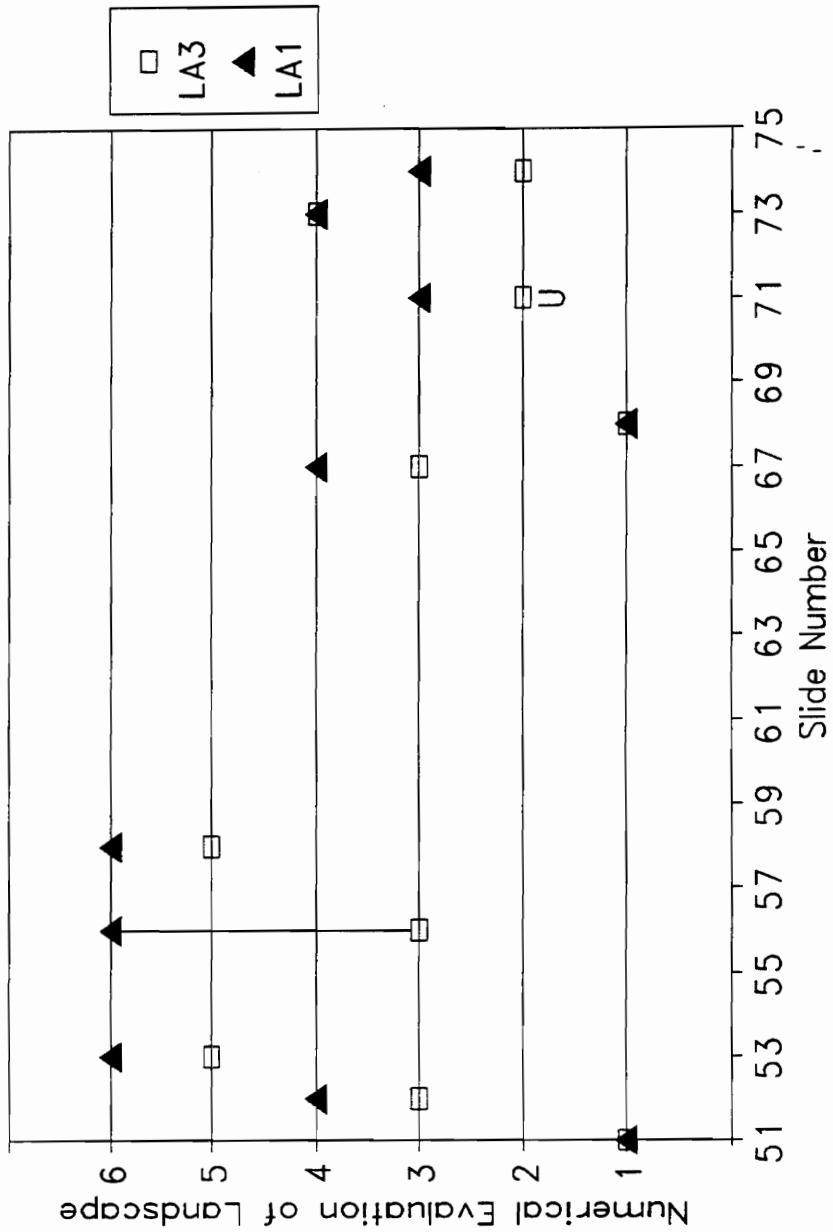

Figure 11F: LA3 vs. LA1 Time 2
Slides # 51–75

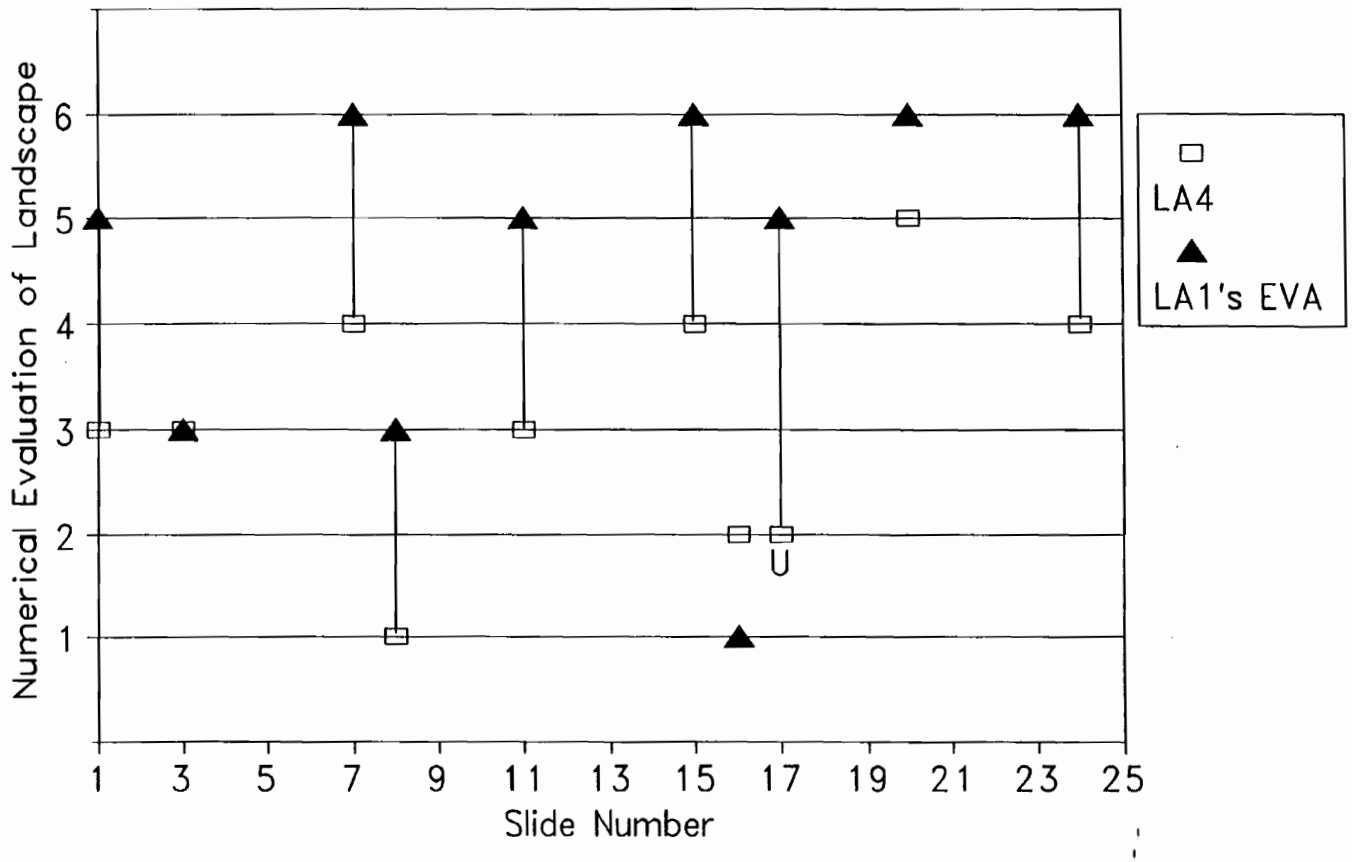

Figure 12A: LA4 vs. LA1's EVA Time 1
Slides # 1-25

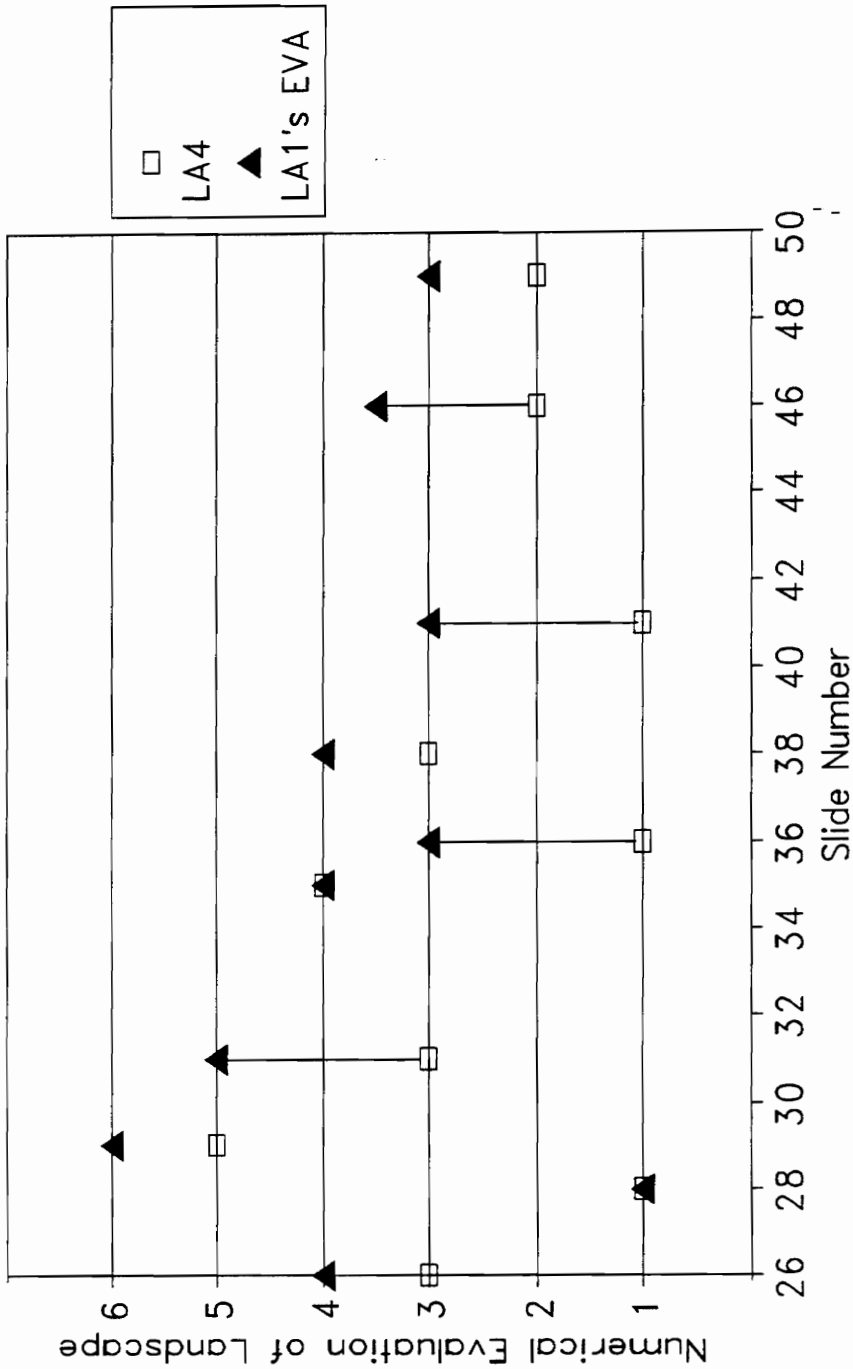

Figure 12B: LA4 vs. LA1's EVA Time 1
Slides # 26–50

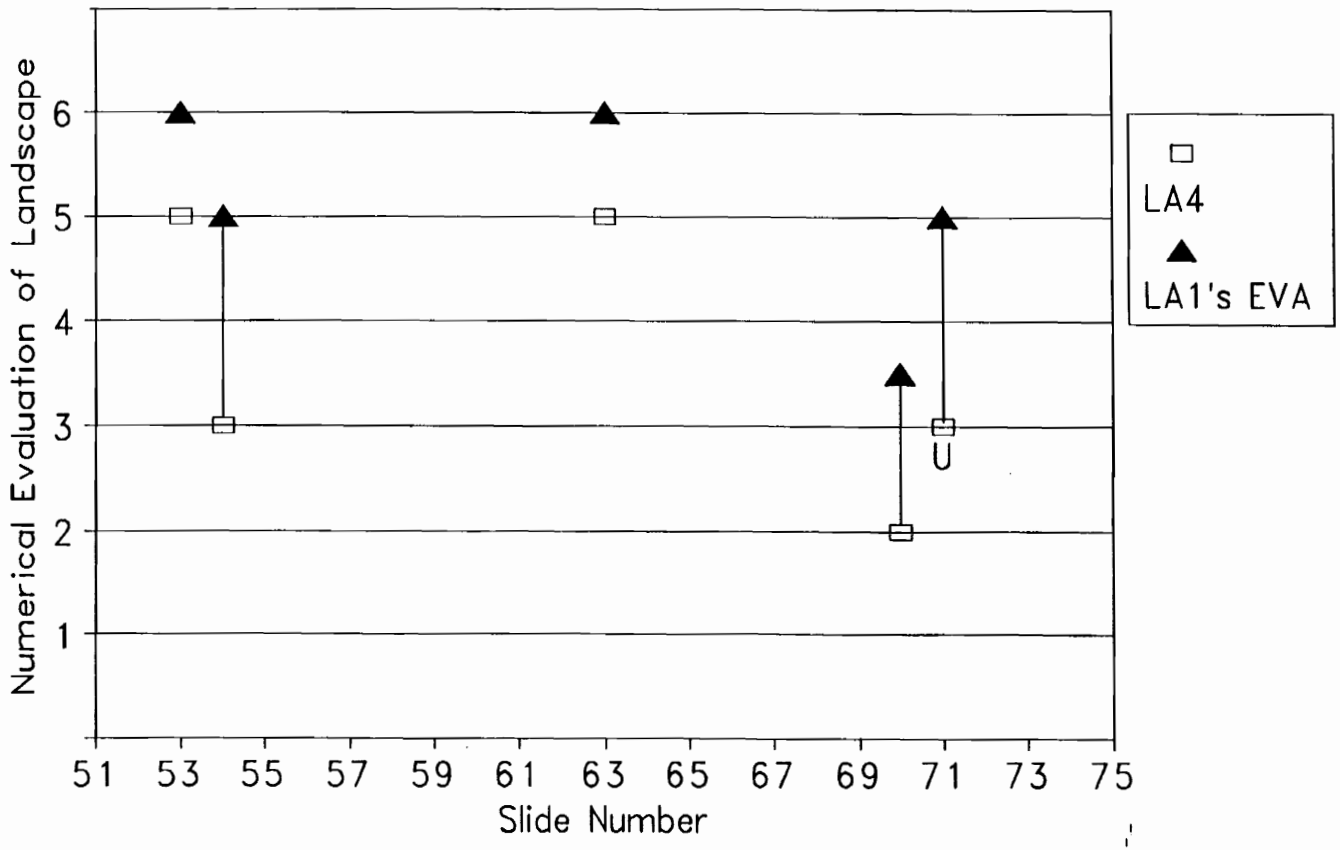

Figure 12C: LA4 vs. LA1's EVA Time 1
Slides # 51-75

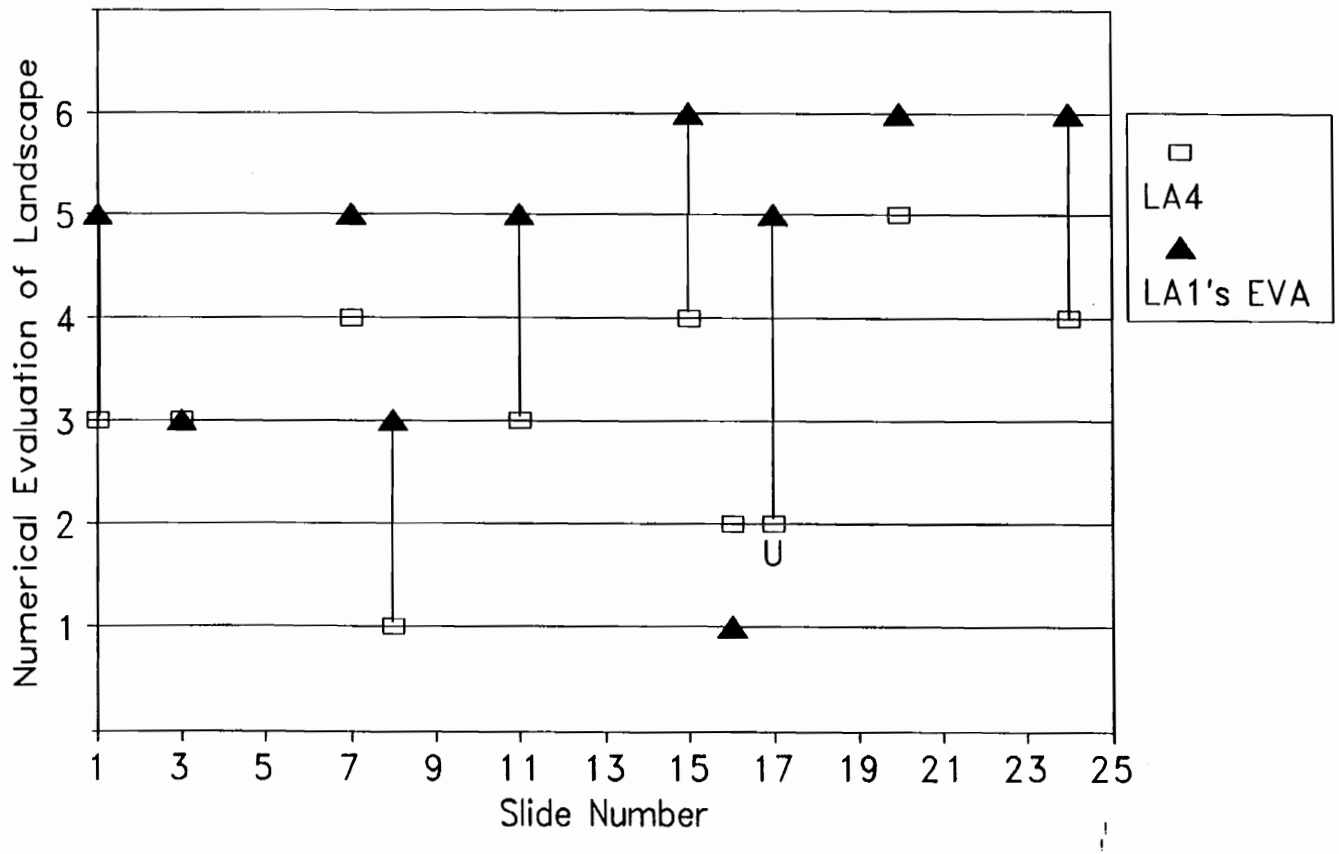

Figure 12D: LA4 vs. LA1's EVA Time 2
Slides # 1–25

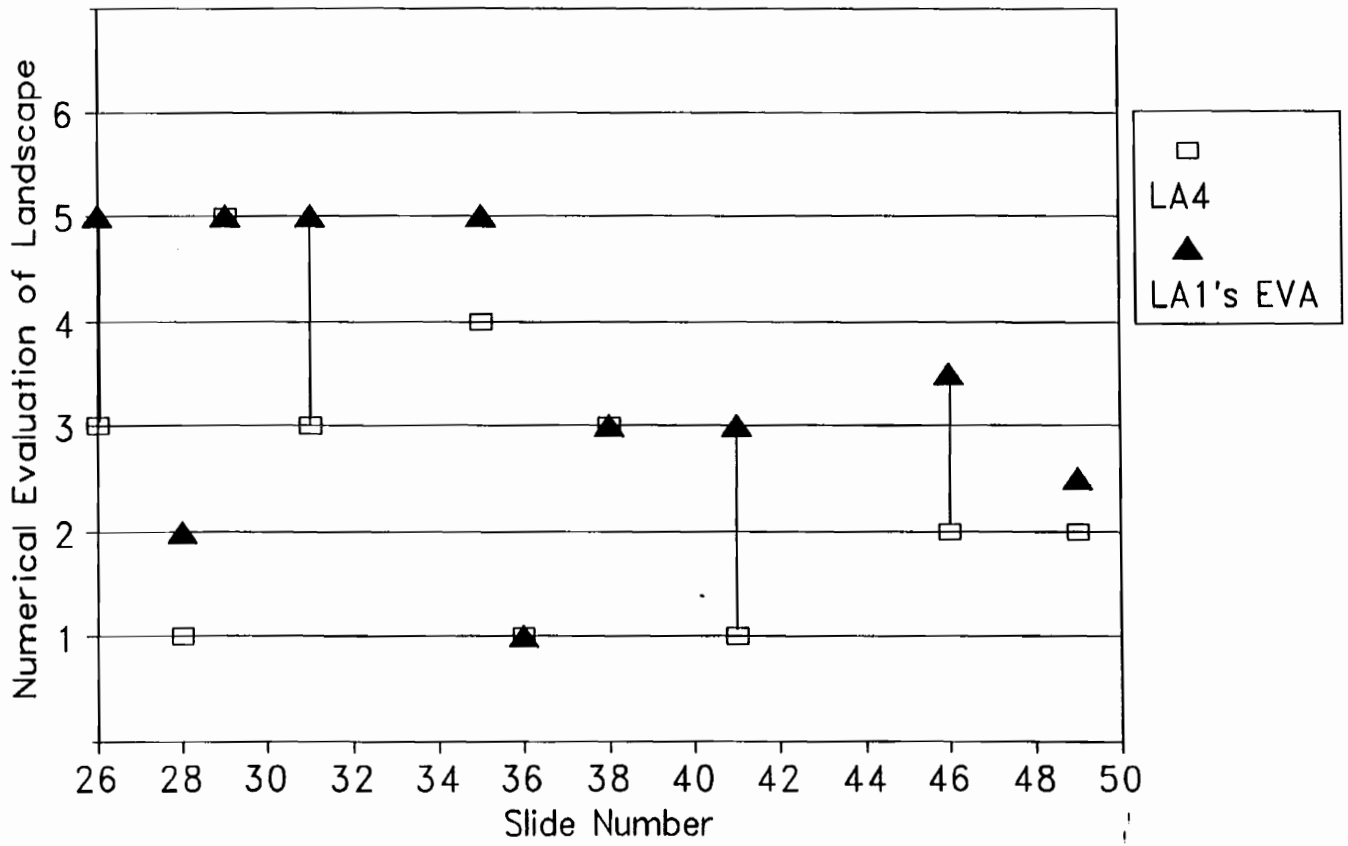

Figure 12E: LA4 vs. LA1's EVA Time 2
Slides # 26–50

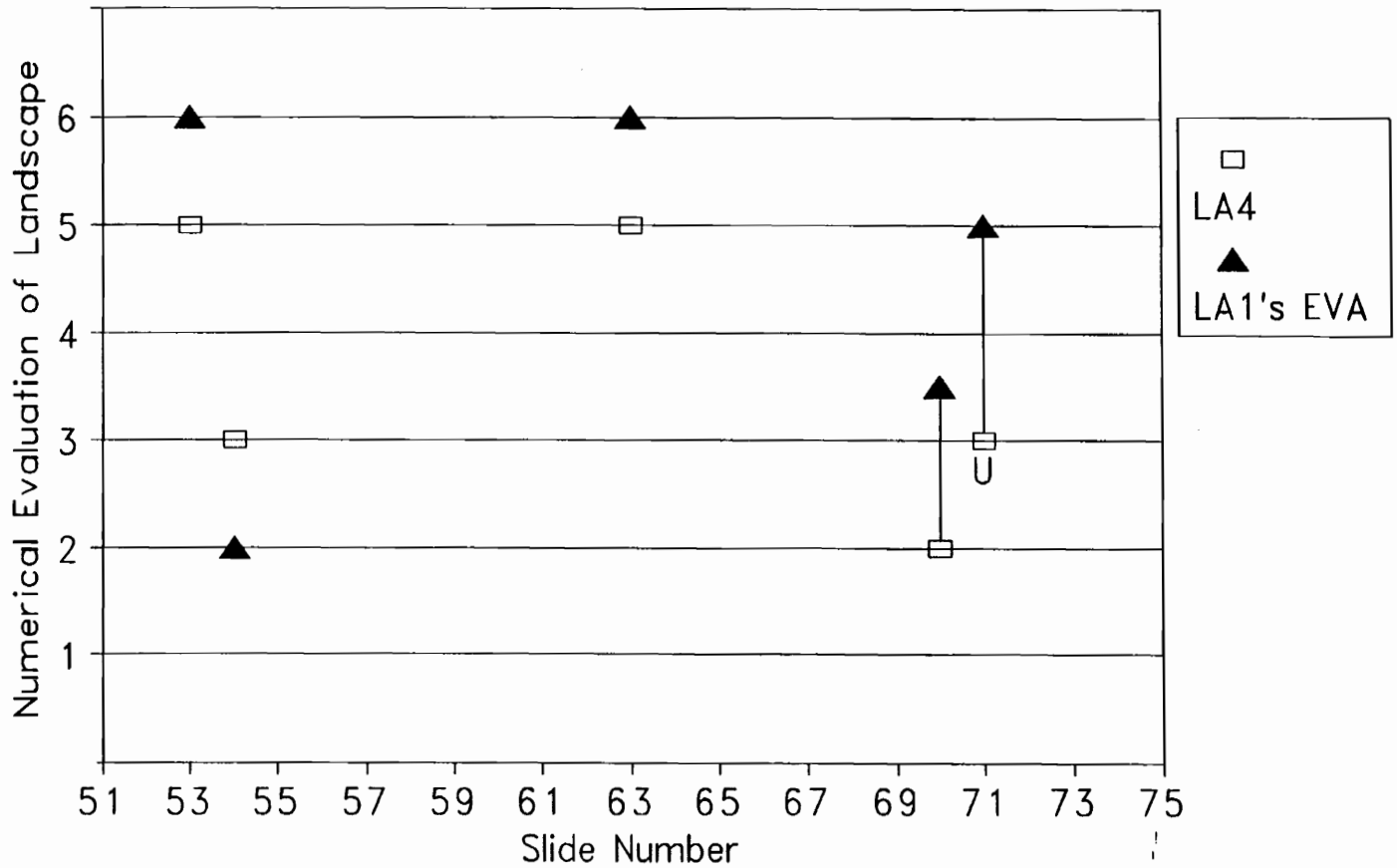

Figure 12F: LA4 vs. LA1's EVA Time 2
Slides # 51–75

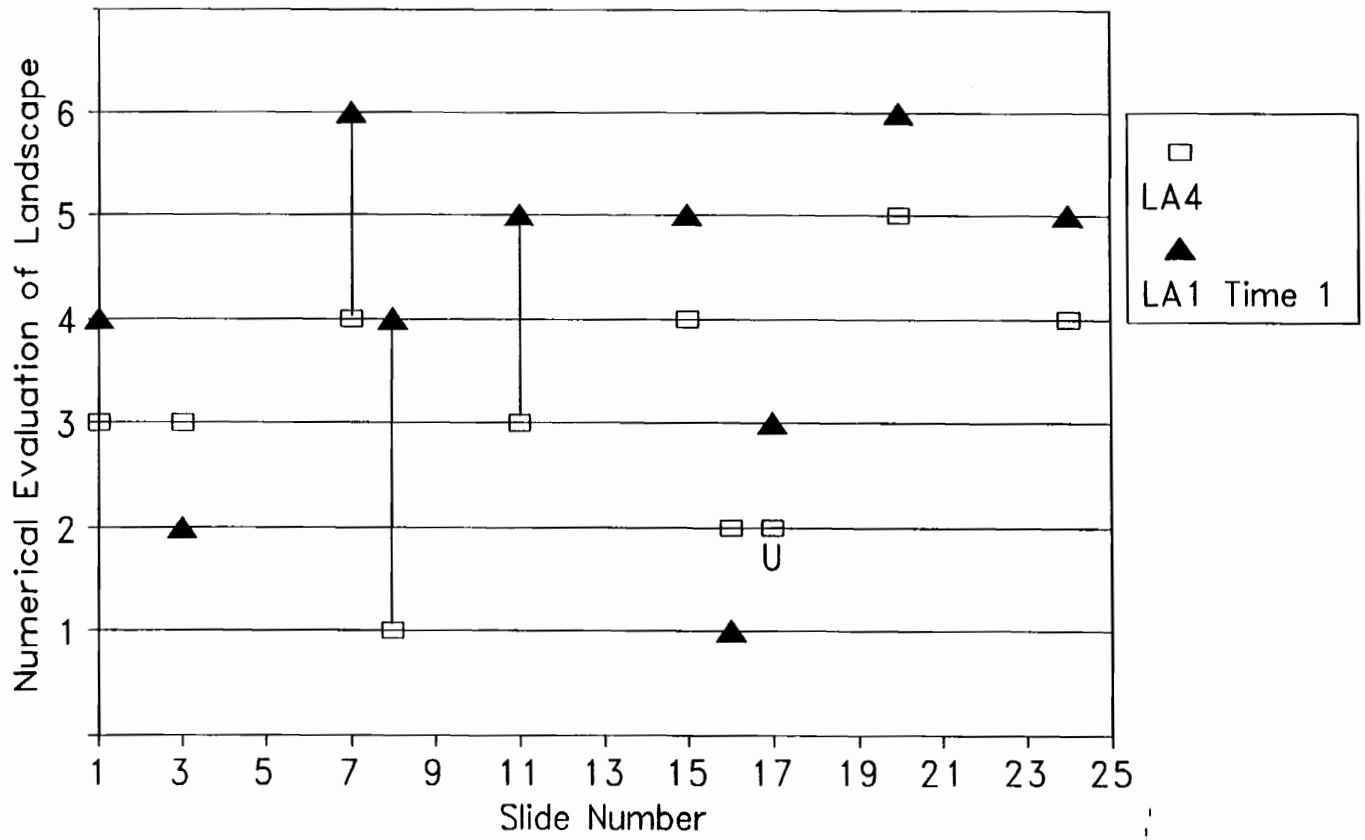

Figure 13A: LA4 vs. LA1 Time 1
Slides # 1–25

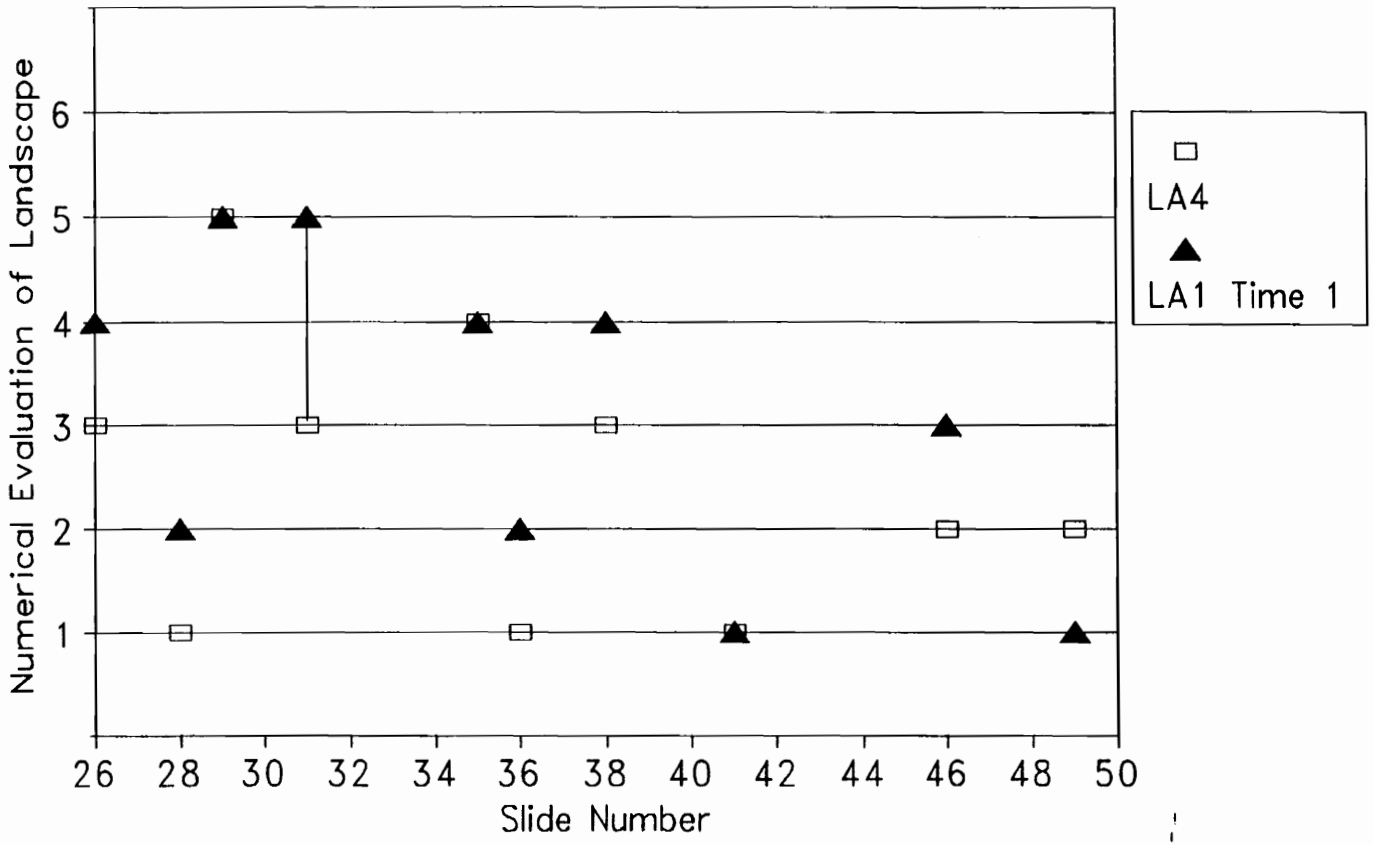

Figure 13B: LA4 vs. LA1 Time 1
Slides # 26-50

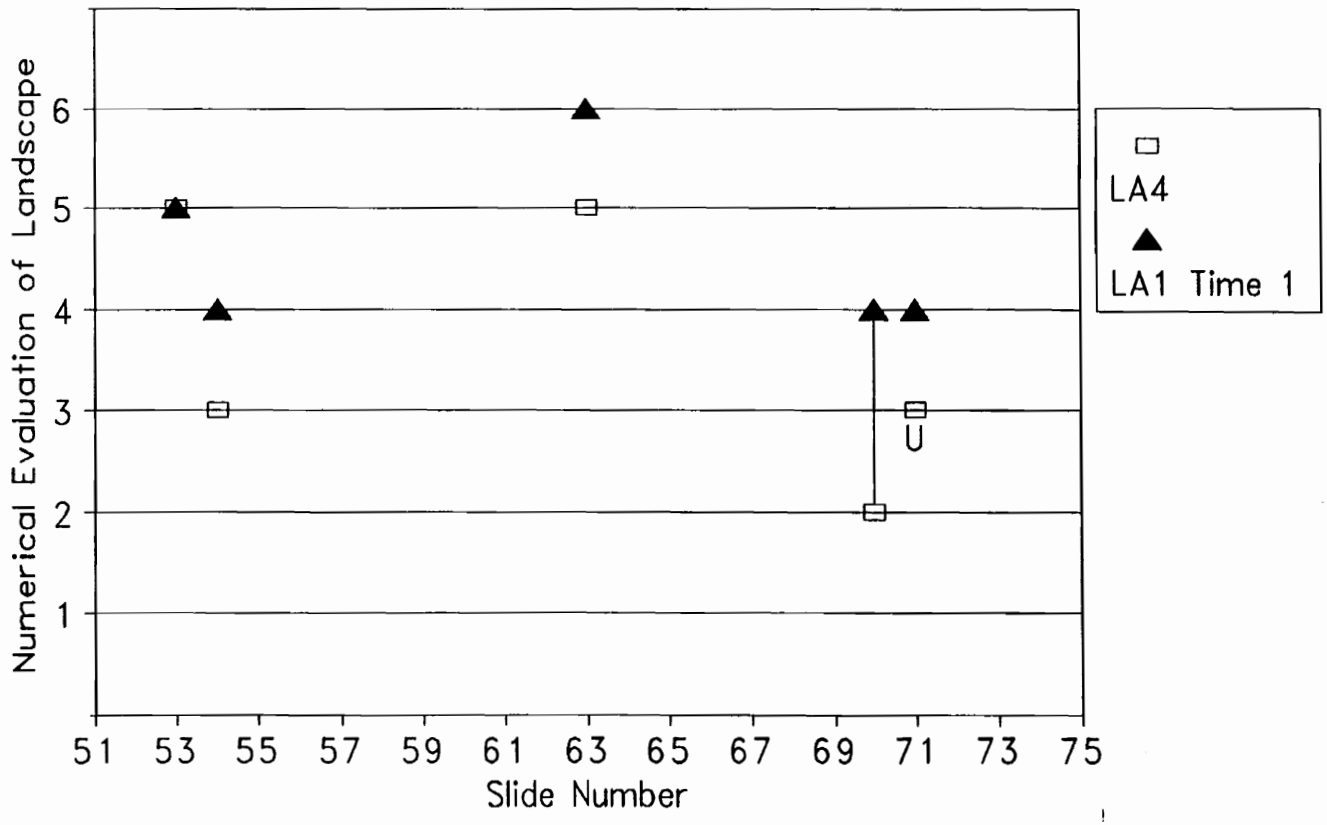

Figure 13C: LA4 vs. LA1 Time 1
Slides # 51-75

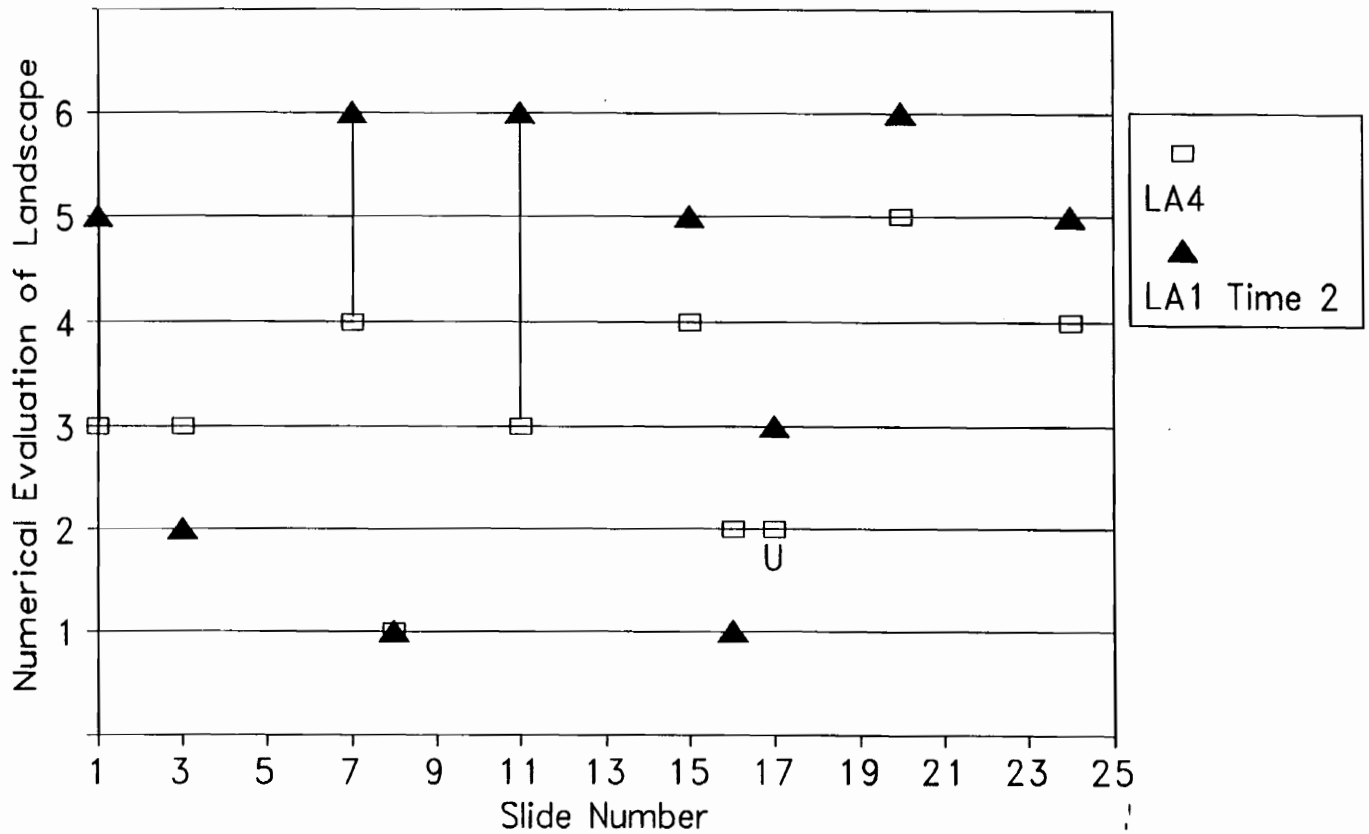

Figure 13D: LA4 vs. LA1 Time 2
Slides # 1-25

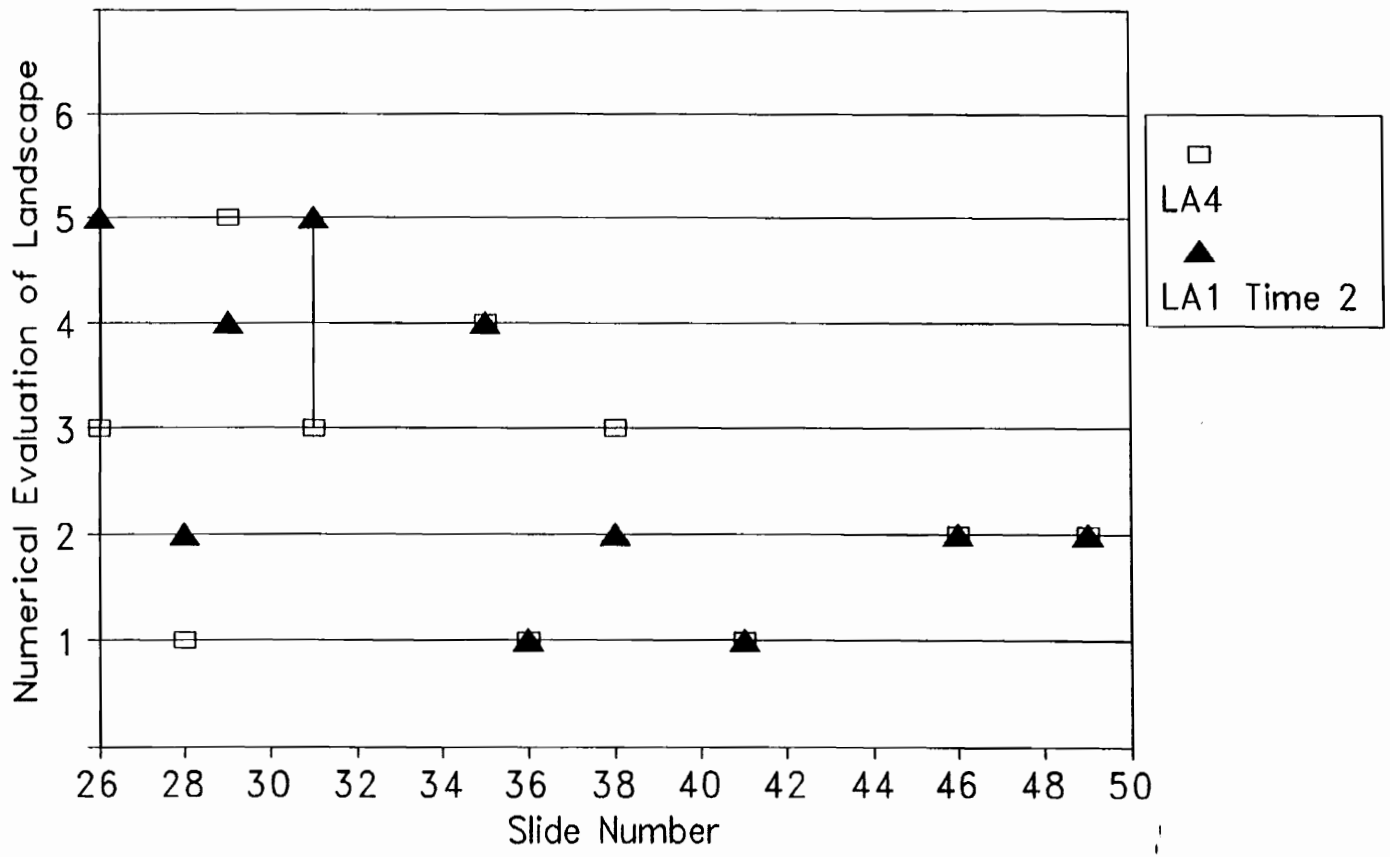

Figure 13E: LA4 vs. LA1 Time 2
Slides # 26-50

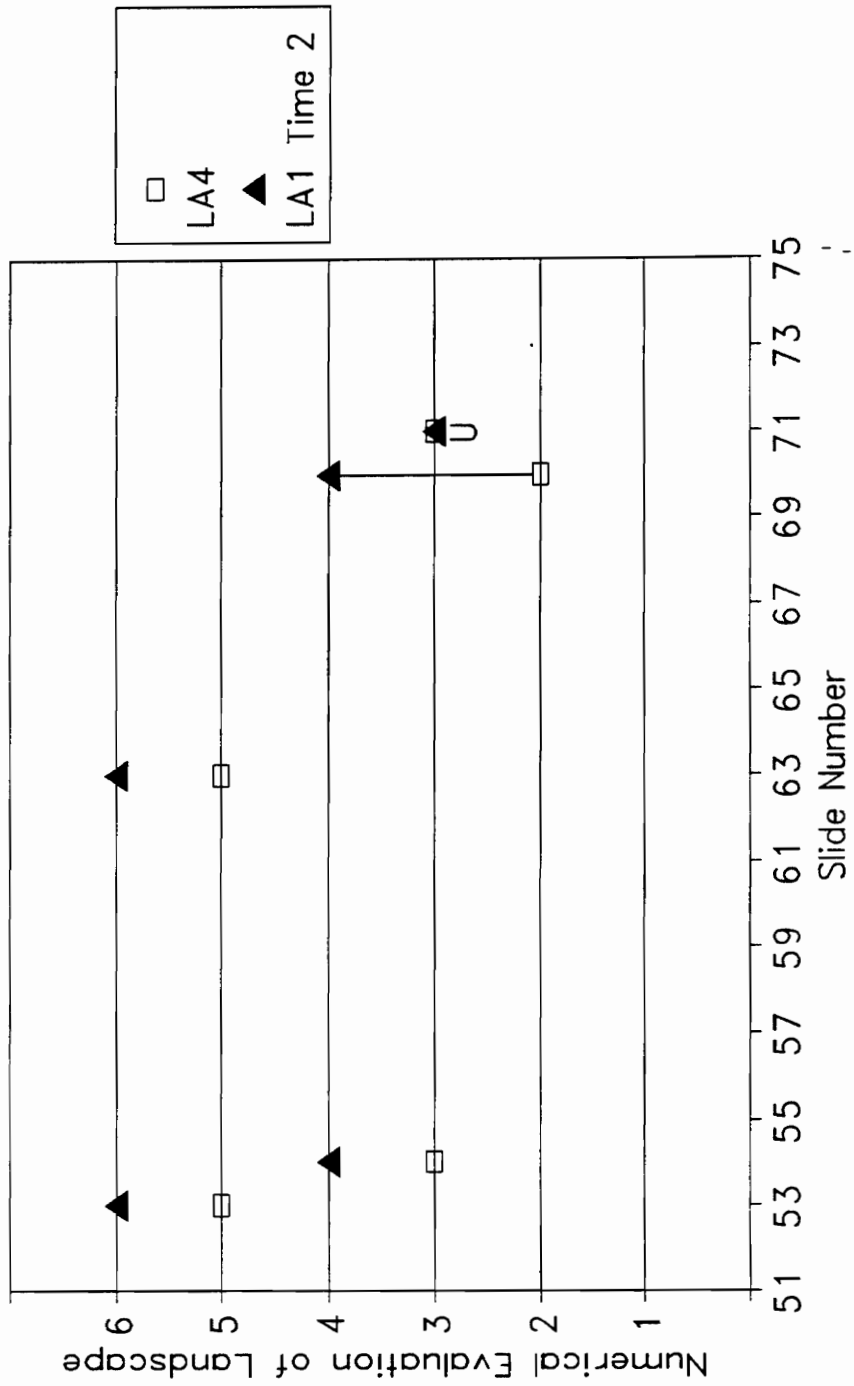

Figure 13F: LA4 vs. LA1 Time 2
Slides # 51–75

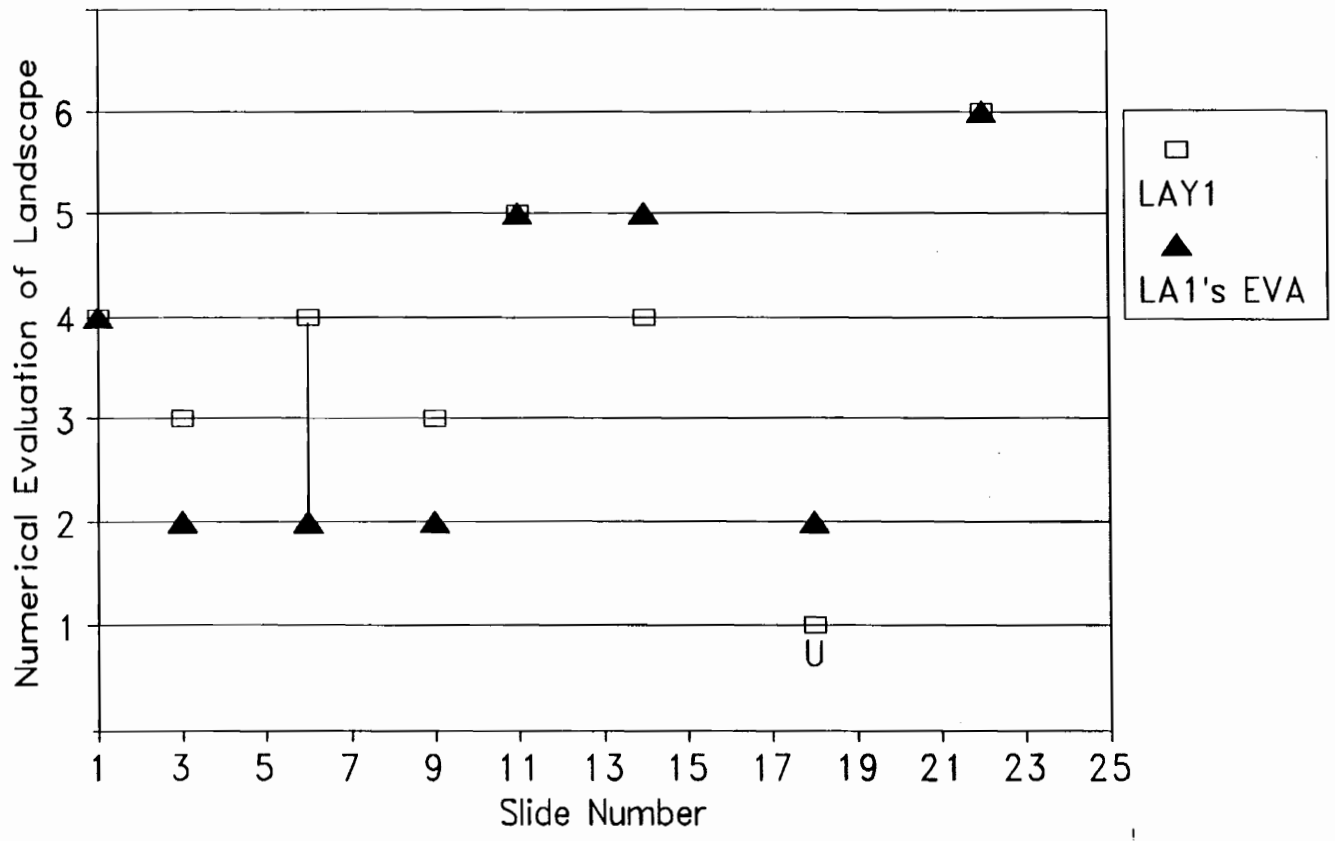

Figure 14A: Lay1 vs. LA1's EVA Time 1
Slides # 1-25

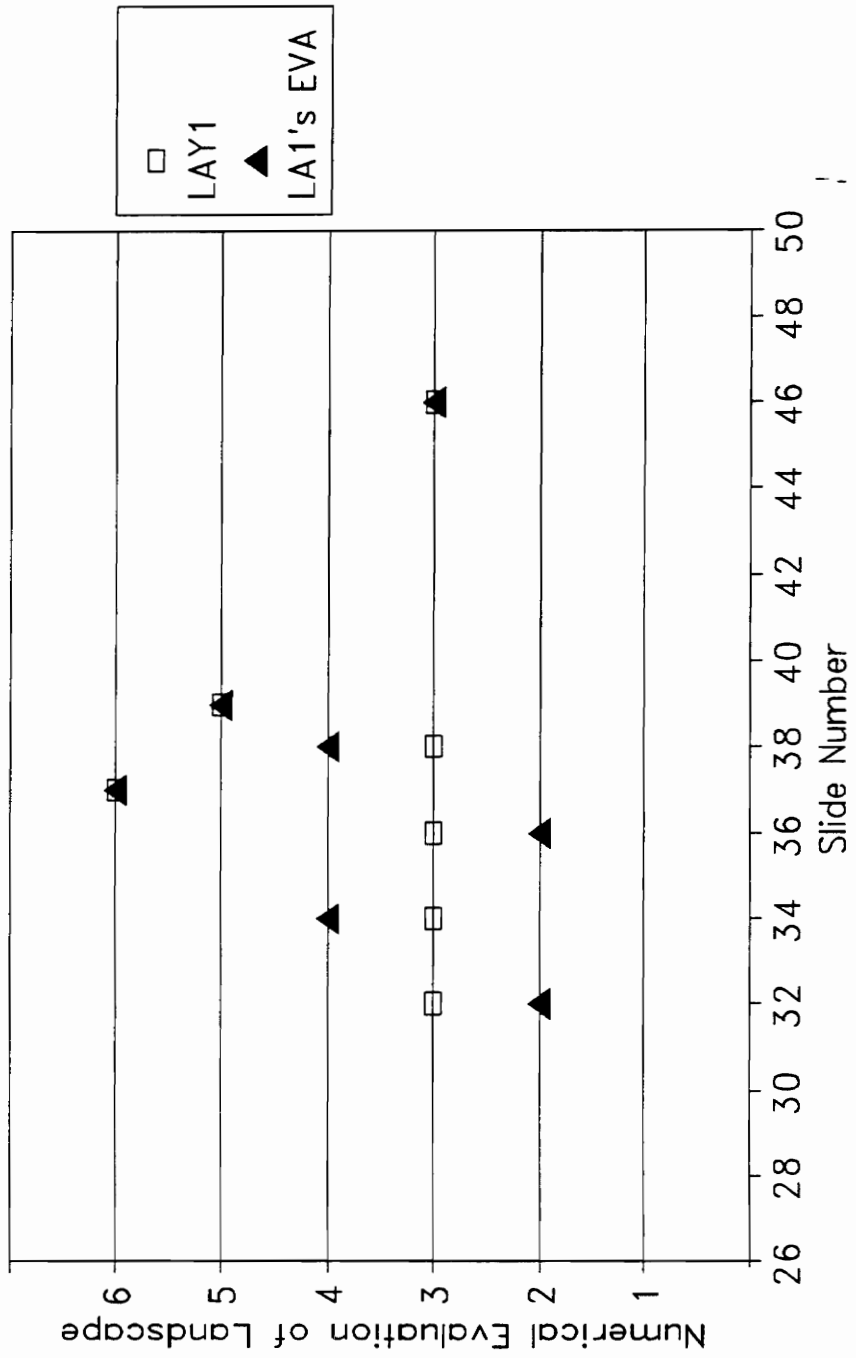

Figure 14B: Lay1 vs. LA1's EVA Time 1
Slides # 26–50

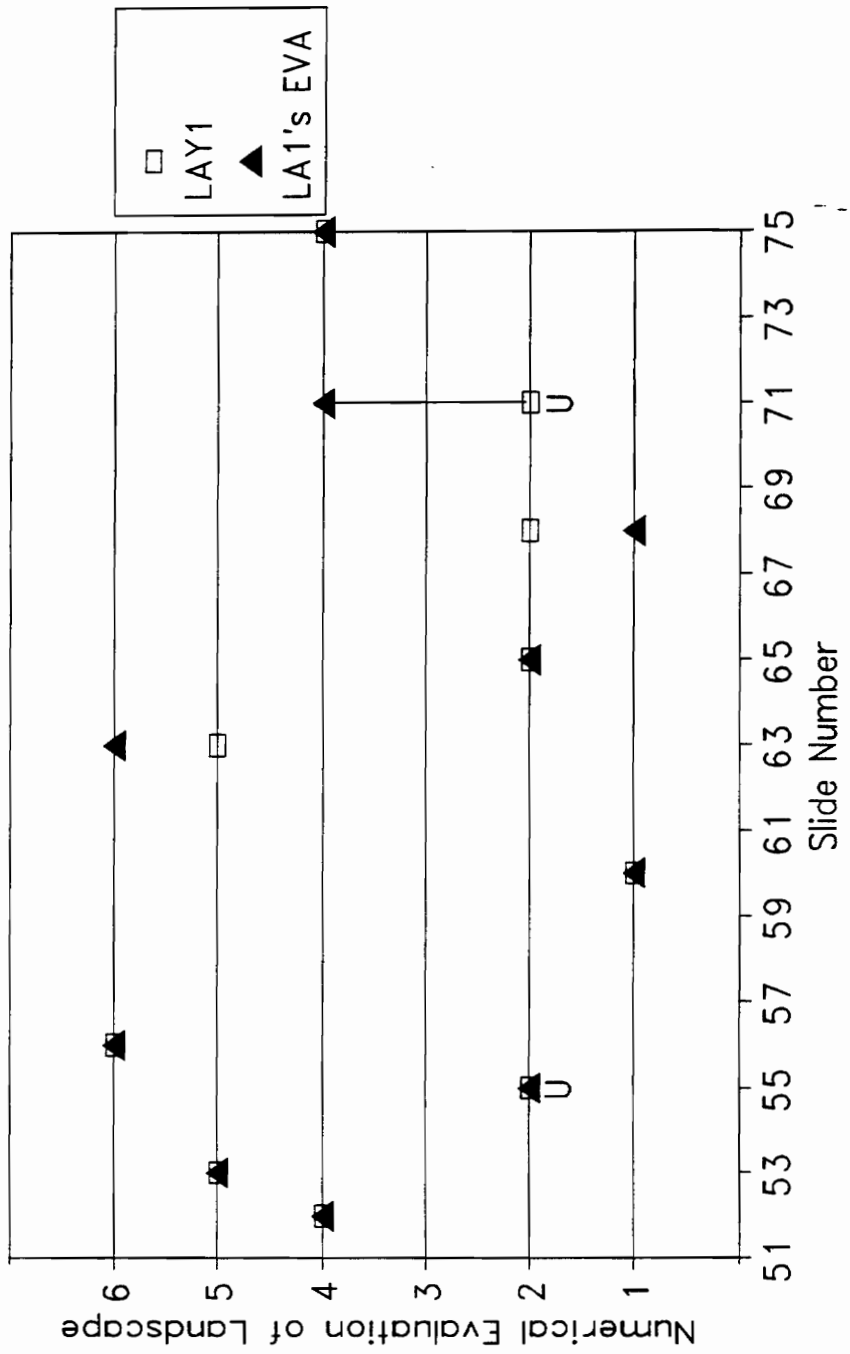

Figure 14C: Lay1 vs. LA1's EVA Time 1
Slides # 51–75

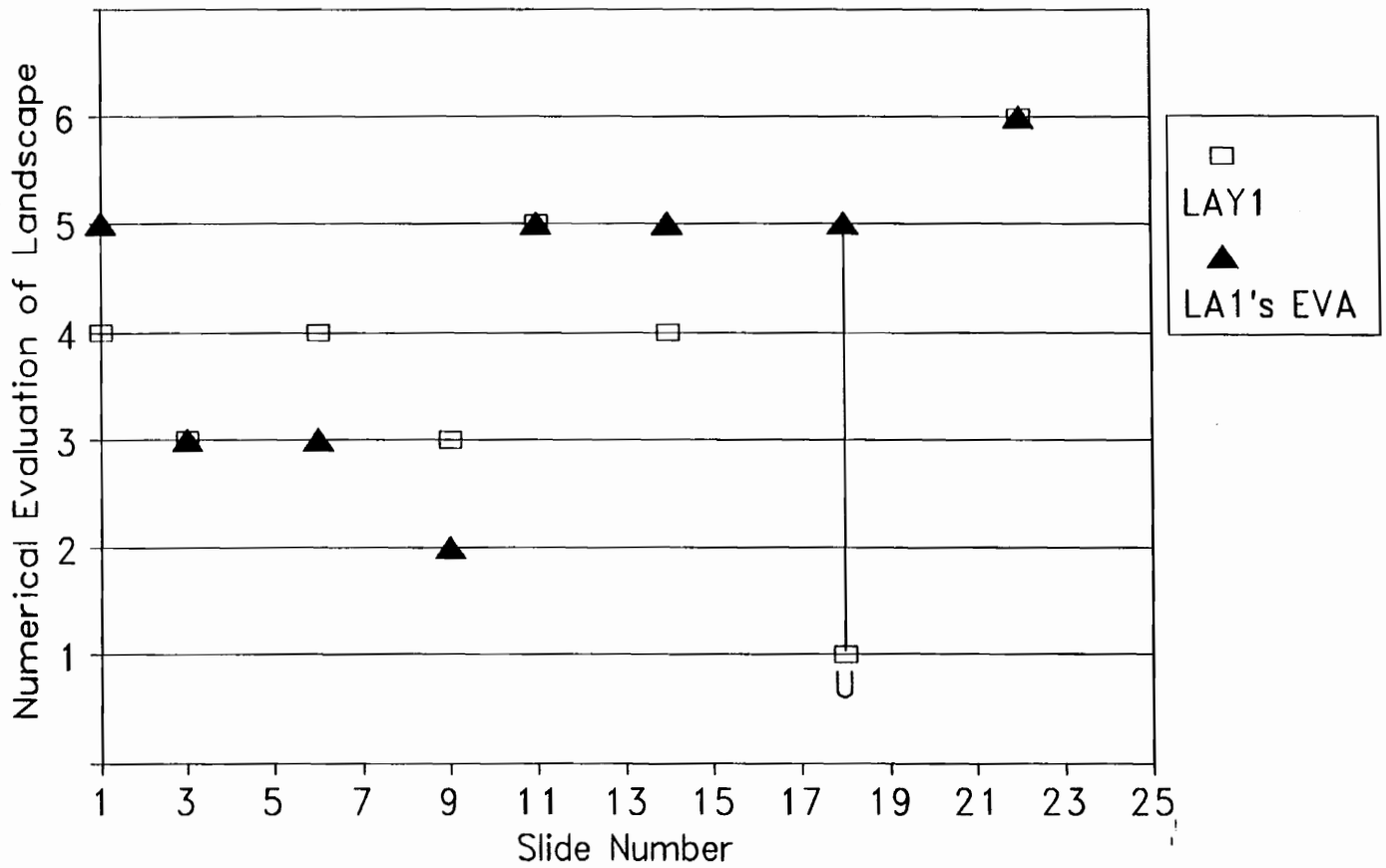

Figure 14D: Lay1 vs. LA1's EVA Time 2
Slides # 1–25

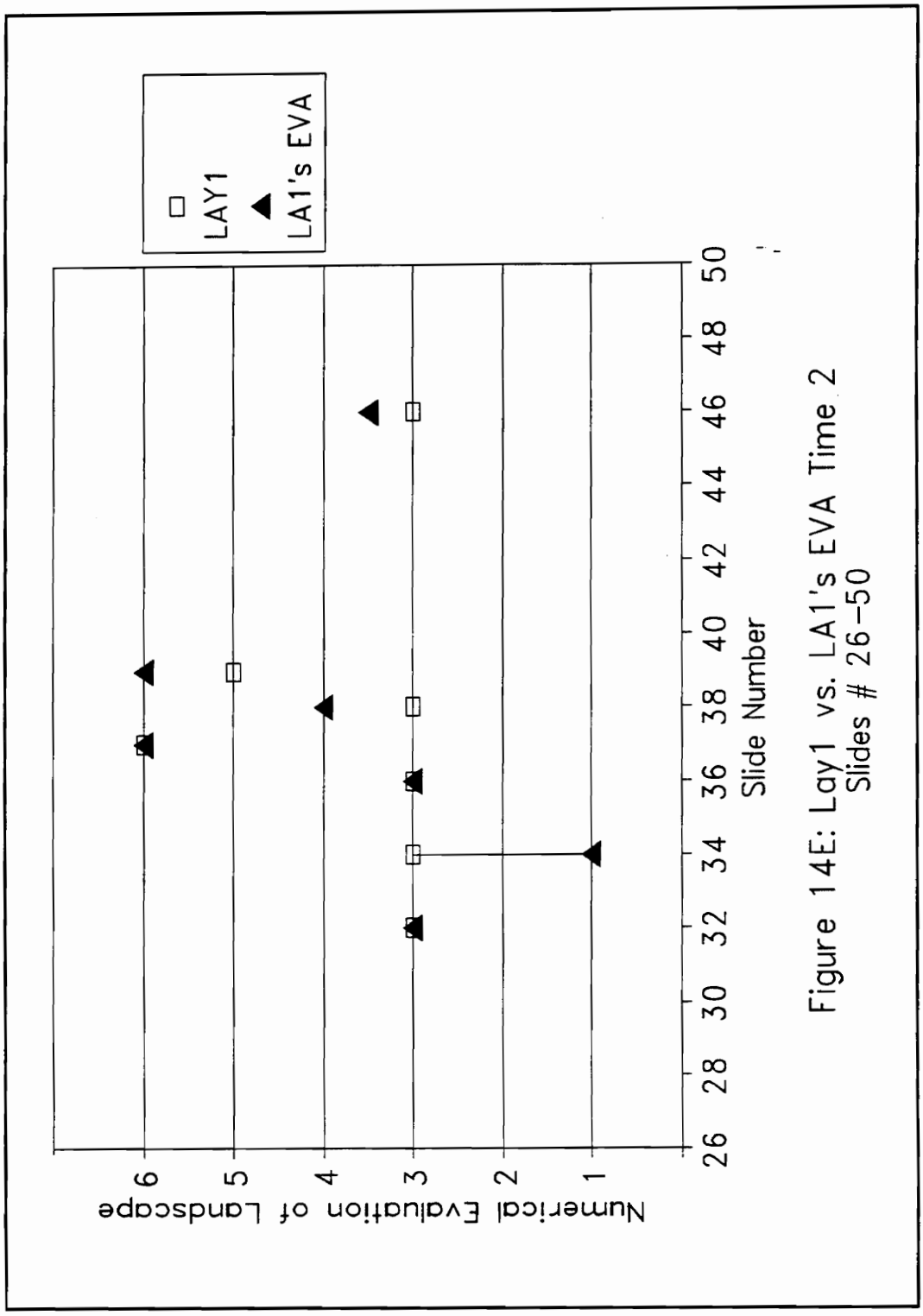

Figure 14E: Lay1 vs. LA1's EVA Time 2
Slides # 26–50

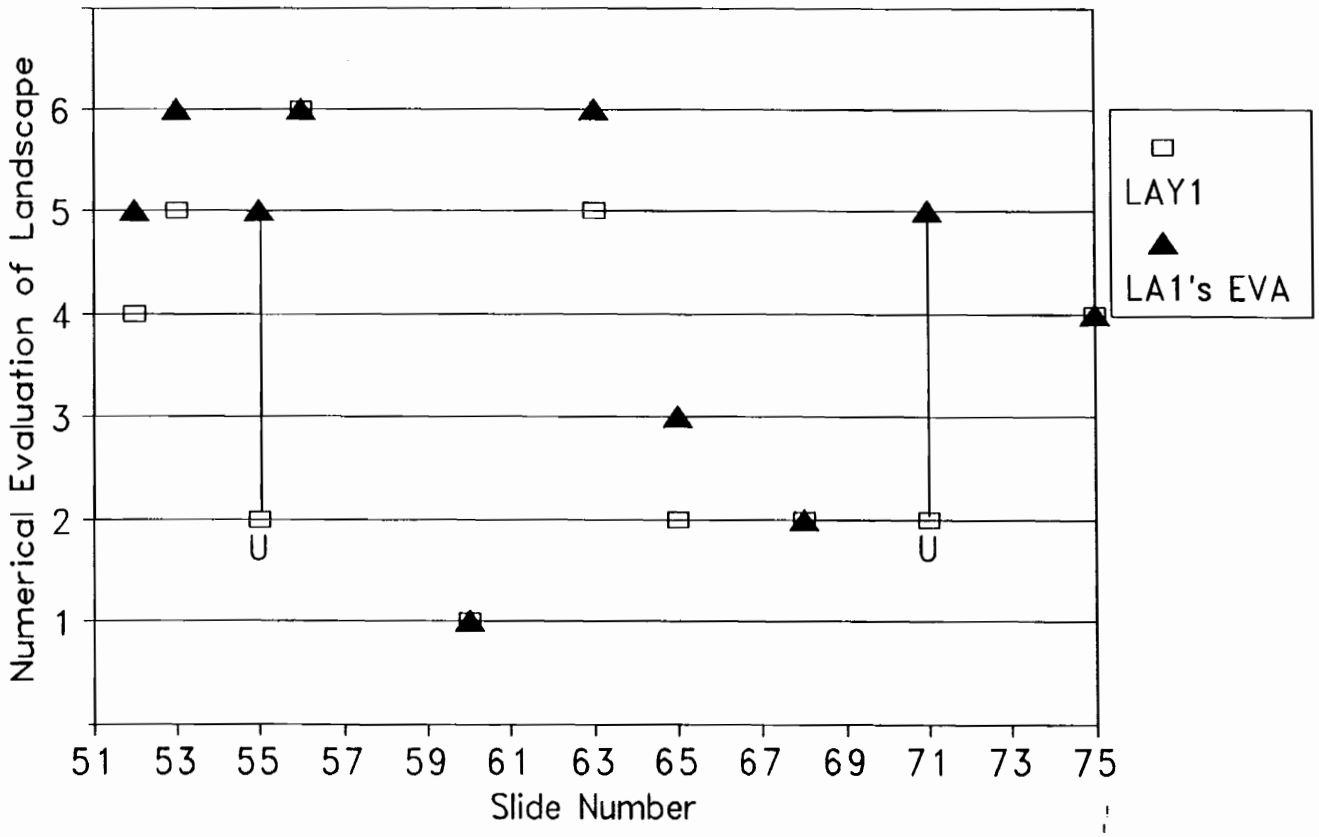

Figure 14F: Lay1 vs. LA1's EVA Time 2
Slides # 51-75

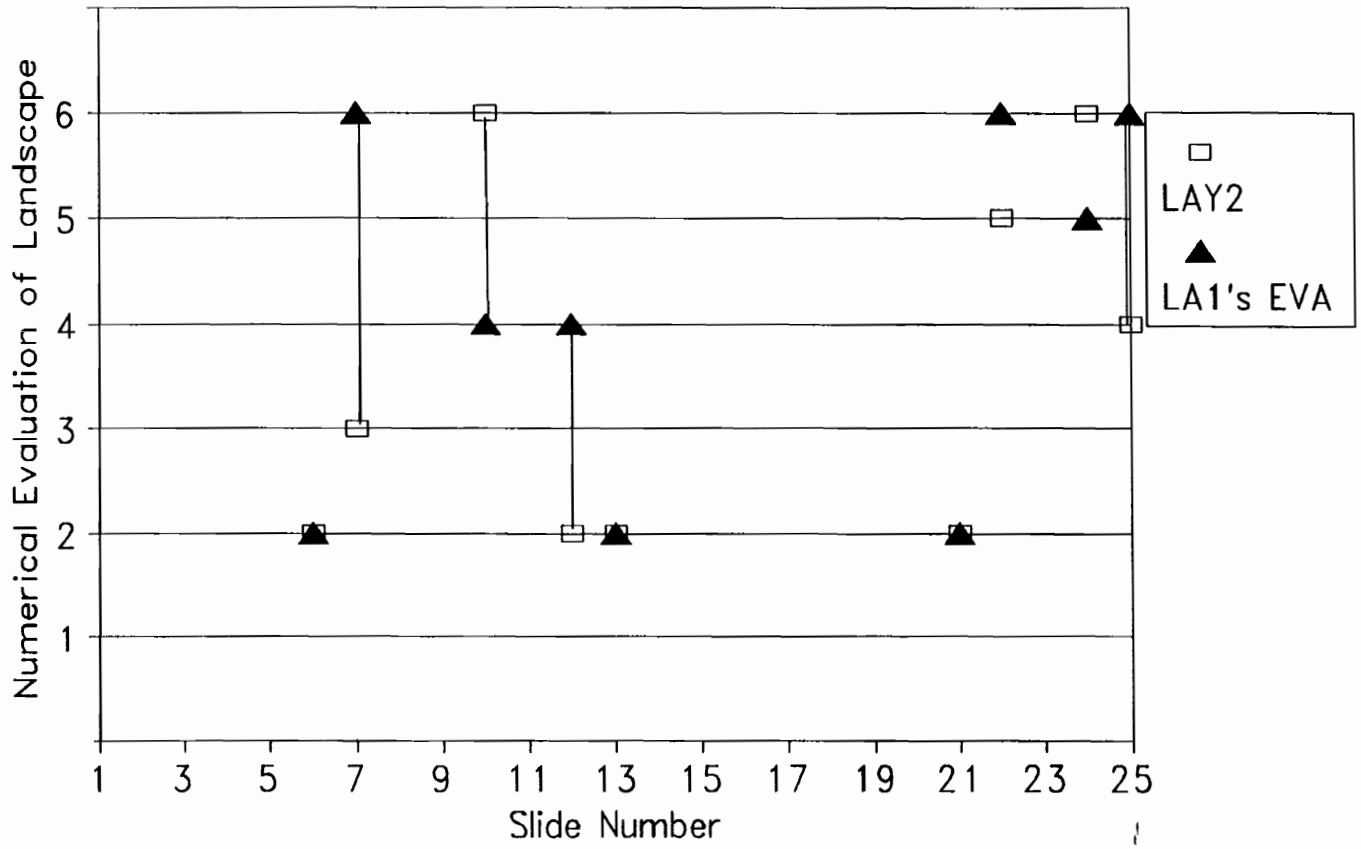

Figure 15A: Lay2 vs. LA1's EVA Time 1
Slides # 1–25

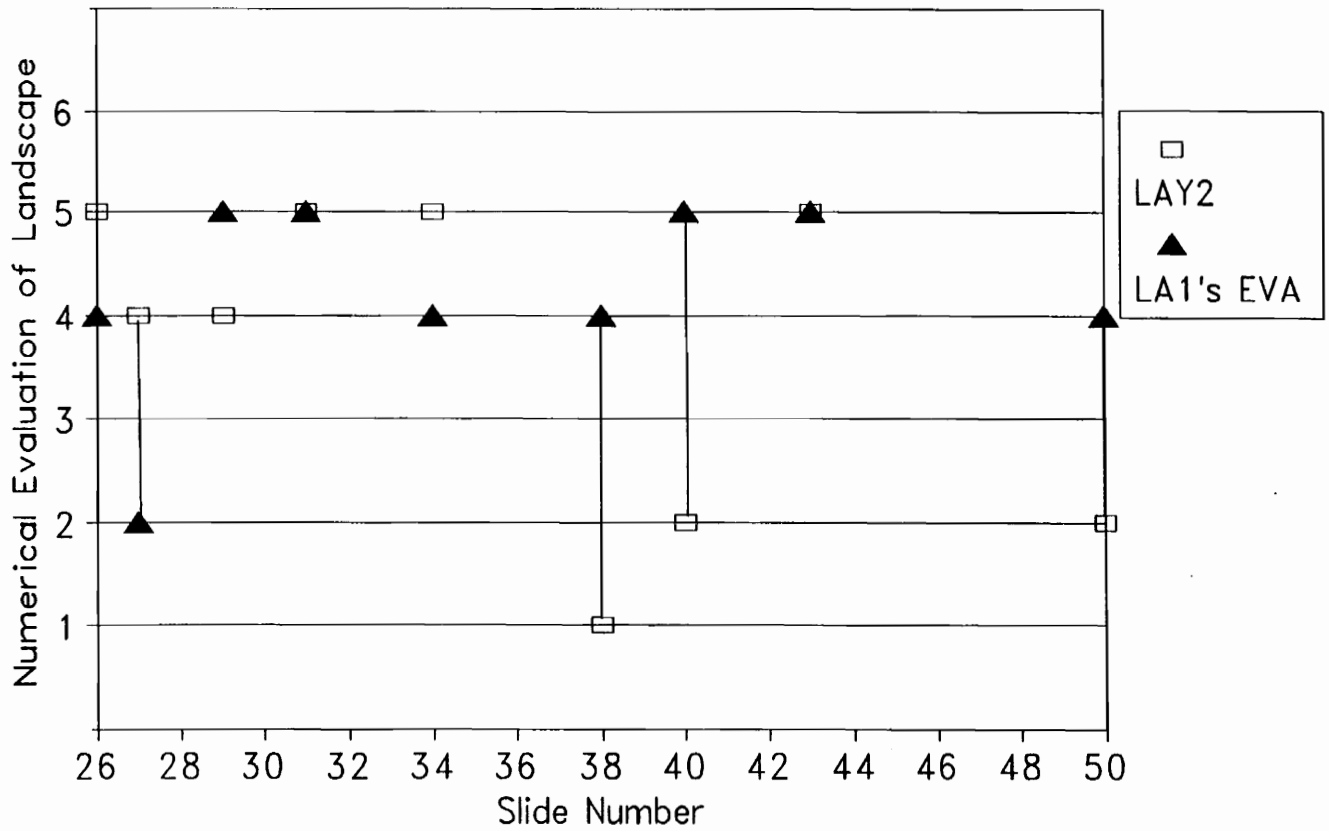

Figure 15B: Lay2 vs. LA1's EVA Time 1
Slides # 26-50

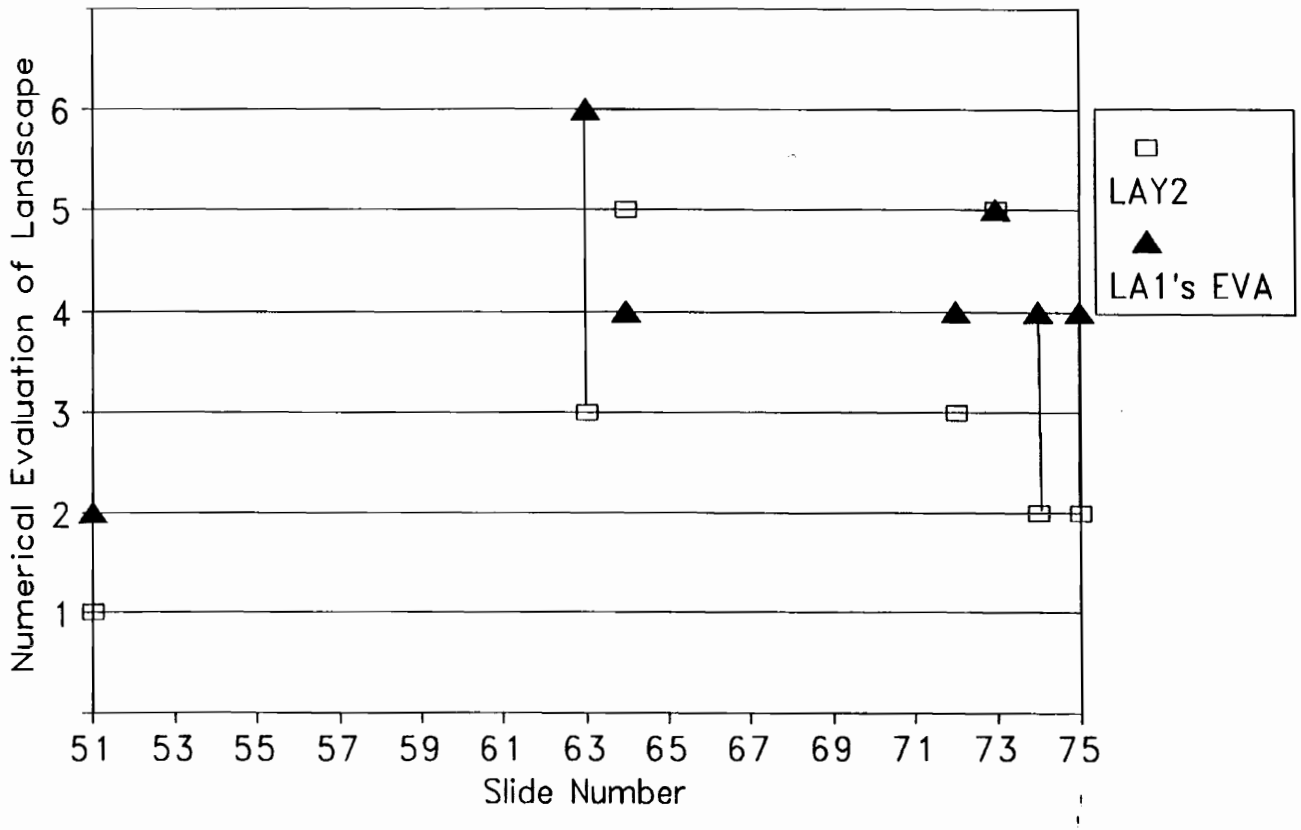

Figure 15C: Lay2 vs. LA1's EVA Time 1
Slides # 51–75

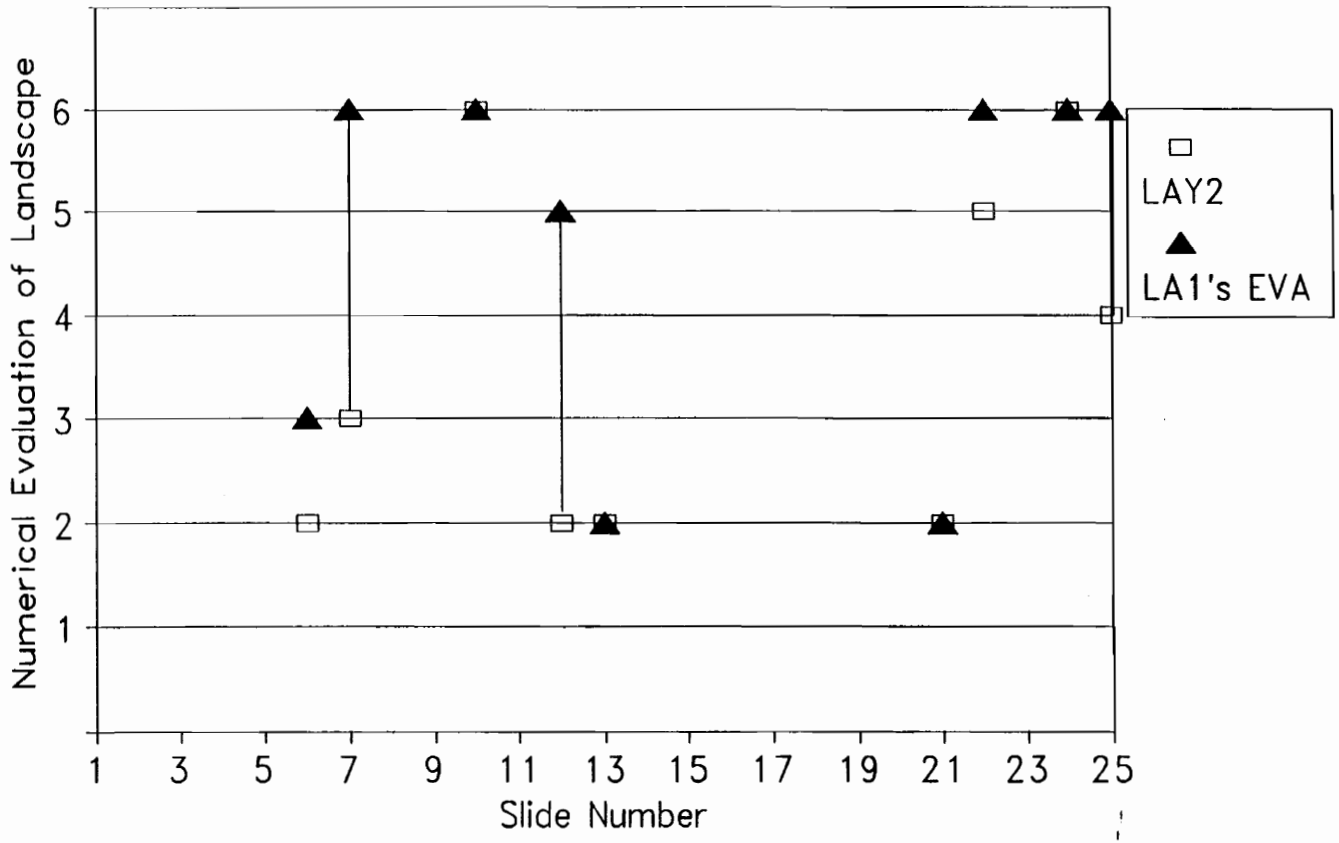

Figure 15D: Lay2 vs. LA1's EVA Time 2
Slides # 1–25

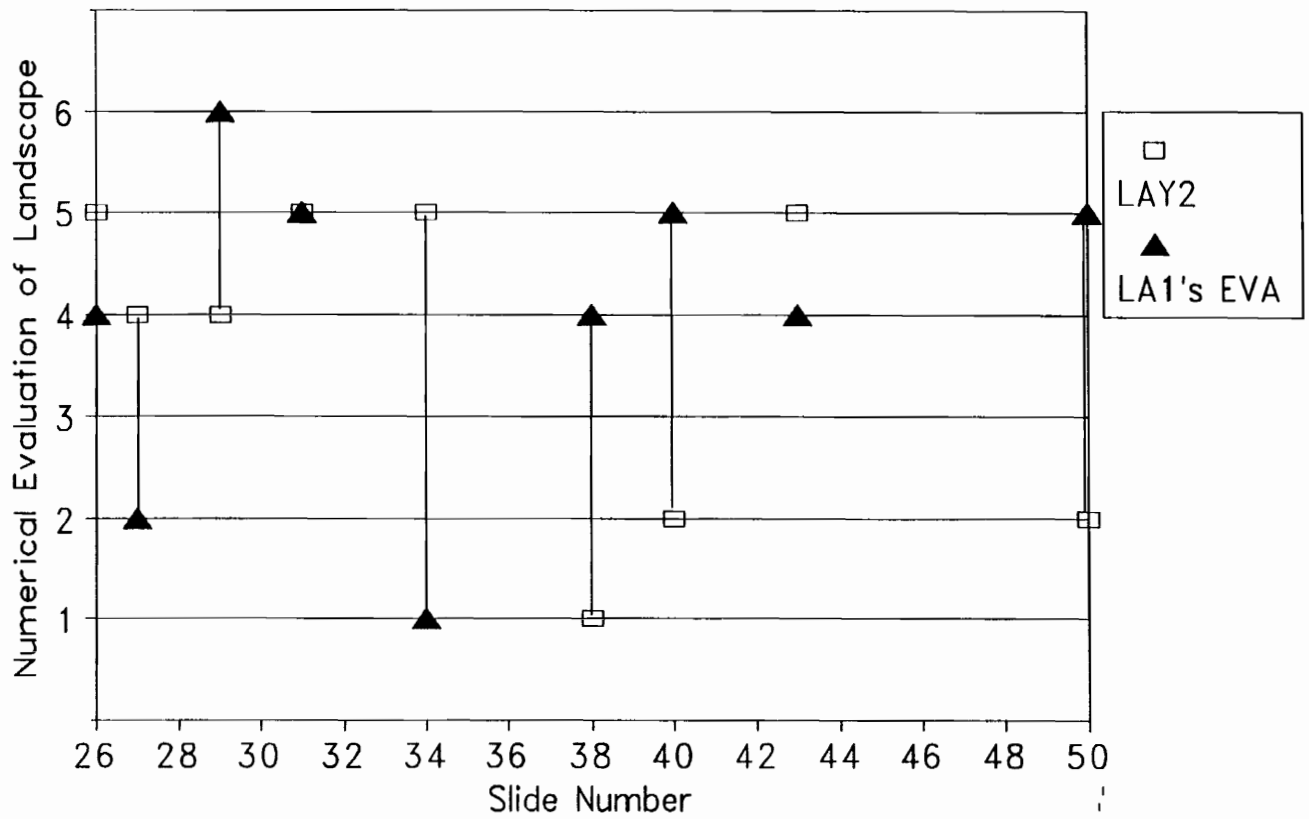

Figure 15E: Lay2 vs. LA1's EVA Time 2
Slides # 26–50

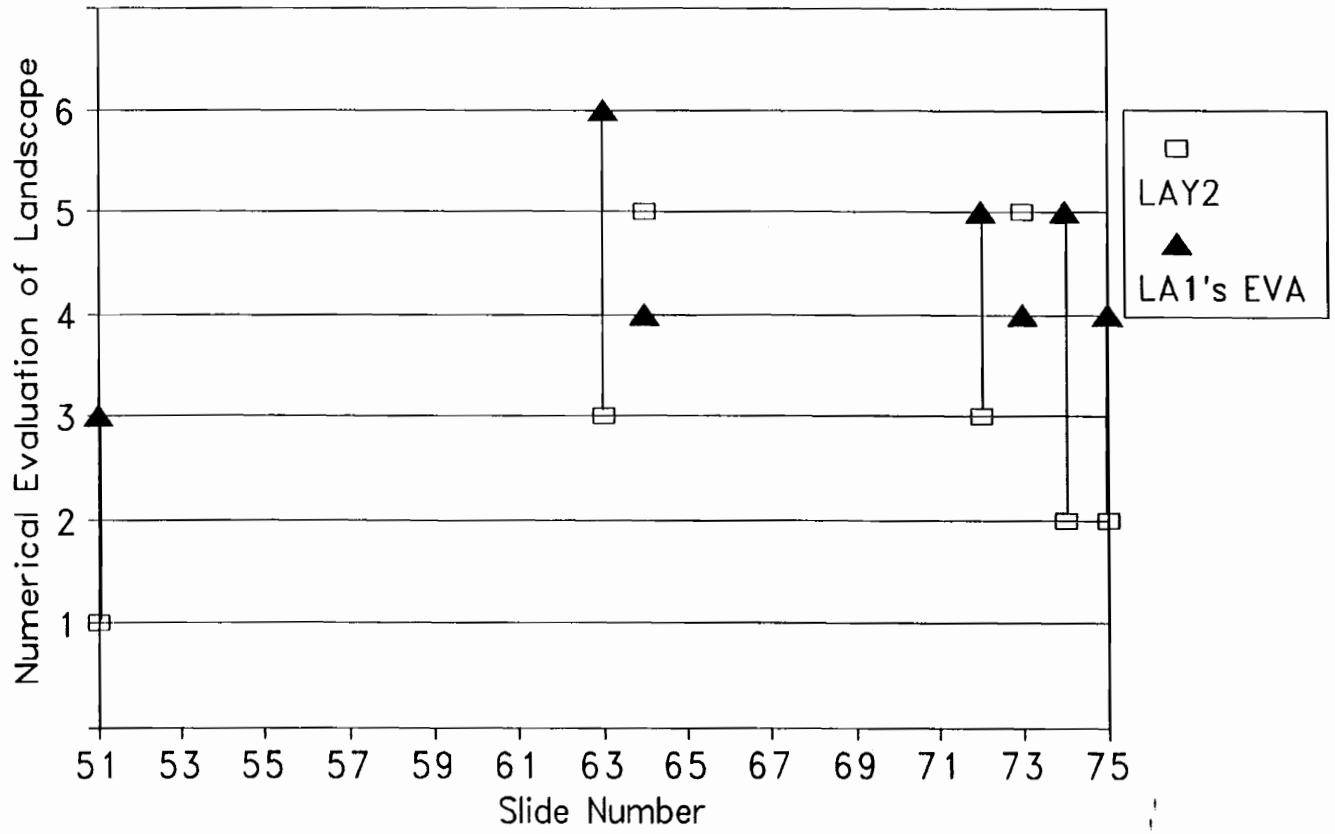

Figure 15F: Lay2 vs. LA1's EVA Time 2
Slides # 51-75

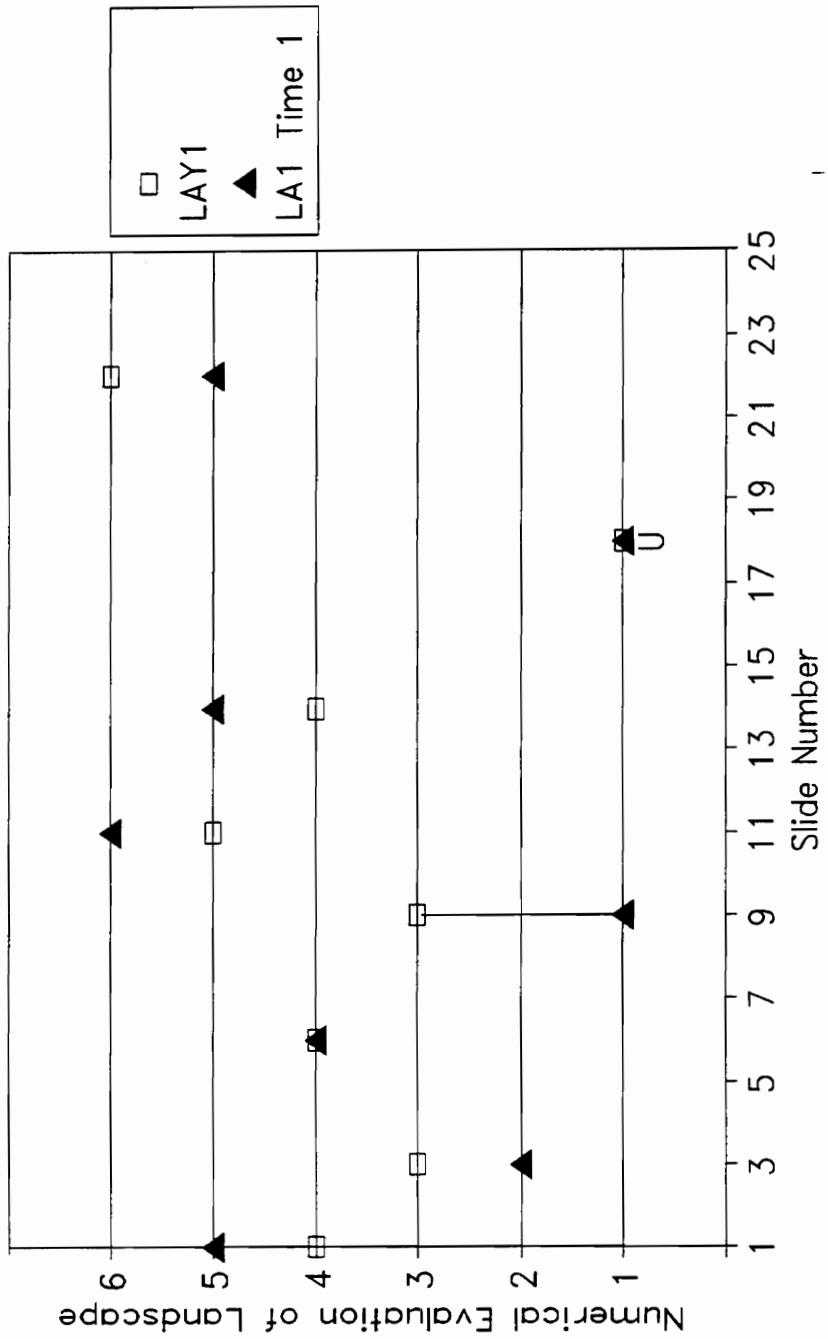

Figure 16A: Layman 1 vs. LA1 Time 1
Slides # 1–25

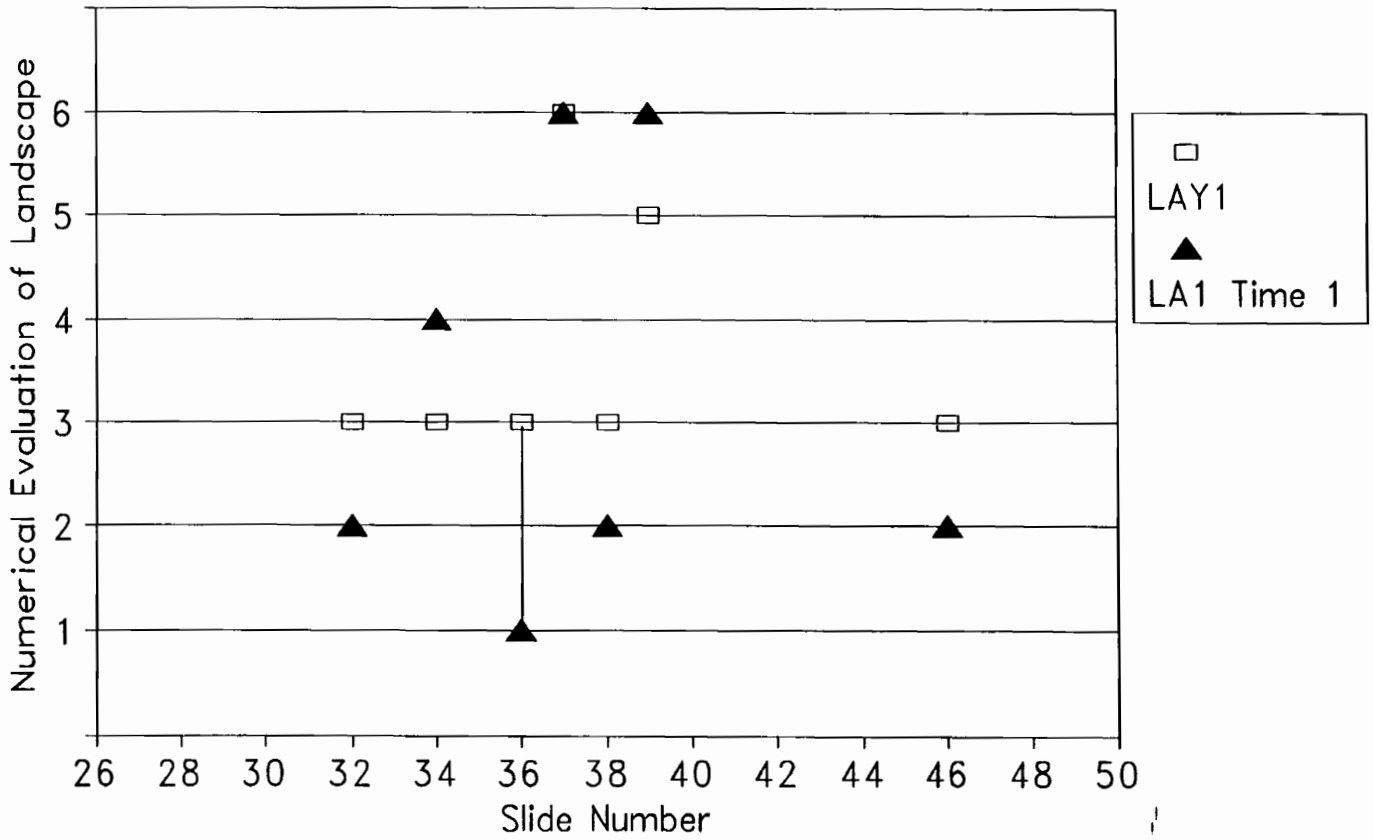

Figure 16B: Layman 1 vs. LA1 Time 1
Slides # 26-50

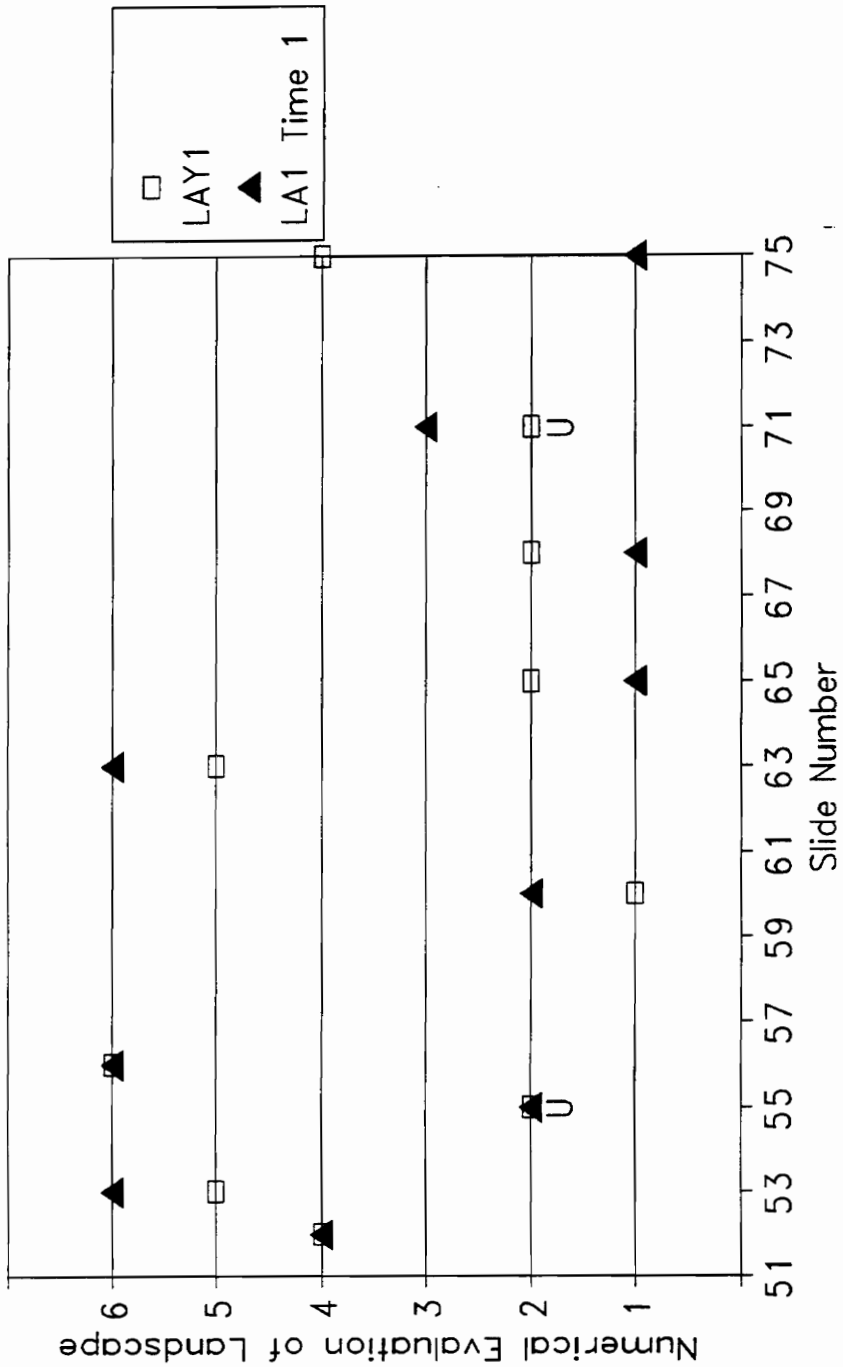

Figure 16C: Layman 1 vs. LA1 Time 1
Slides # 51–75

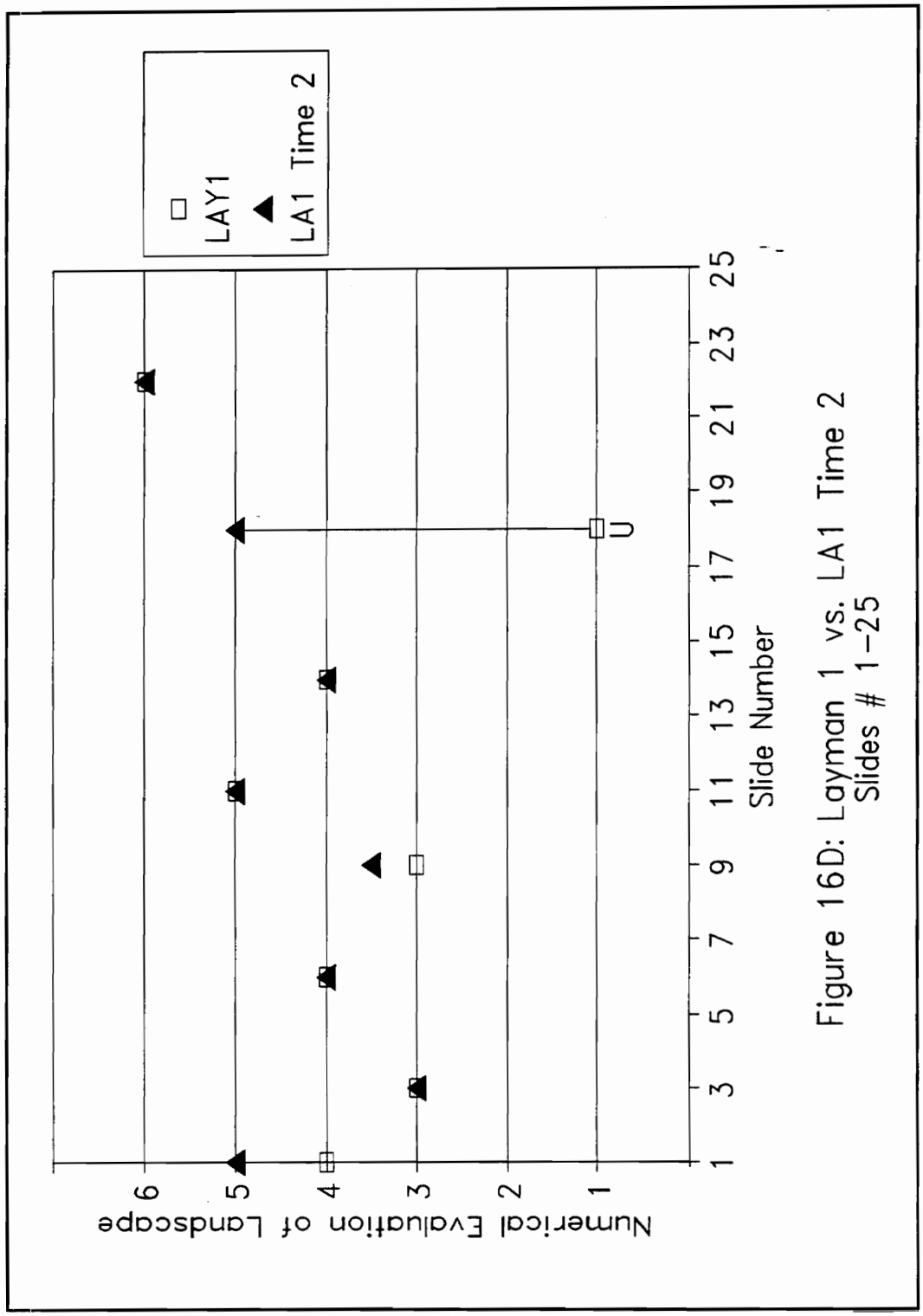

Figure 16D: Layman 1 vs. LA1 Time 2
Slides # 1–25

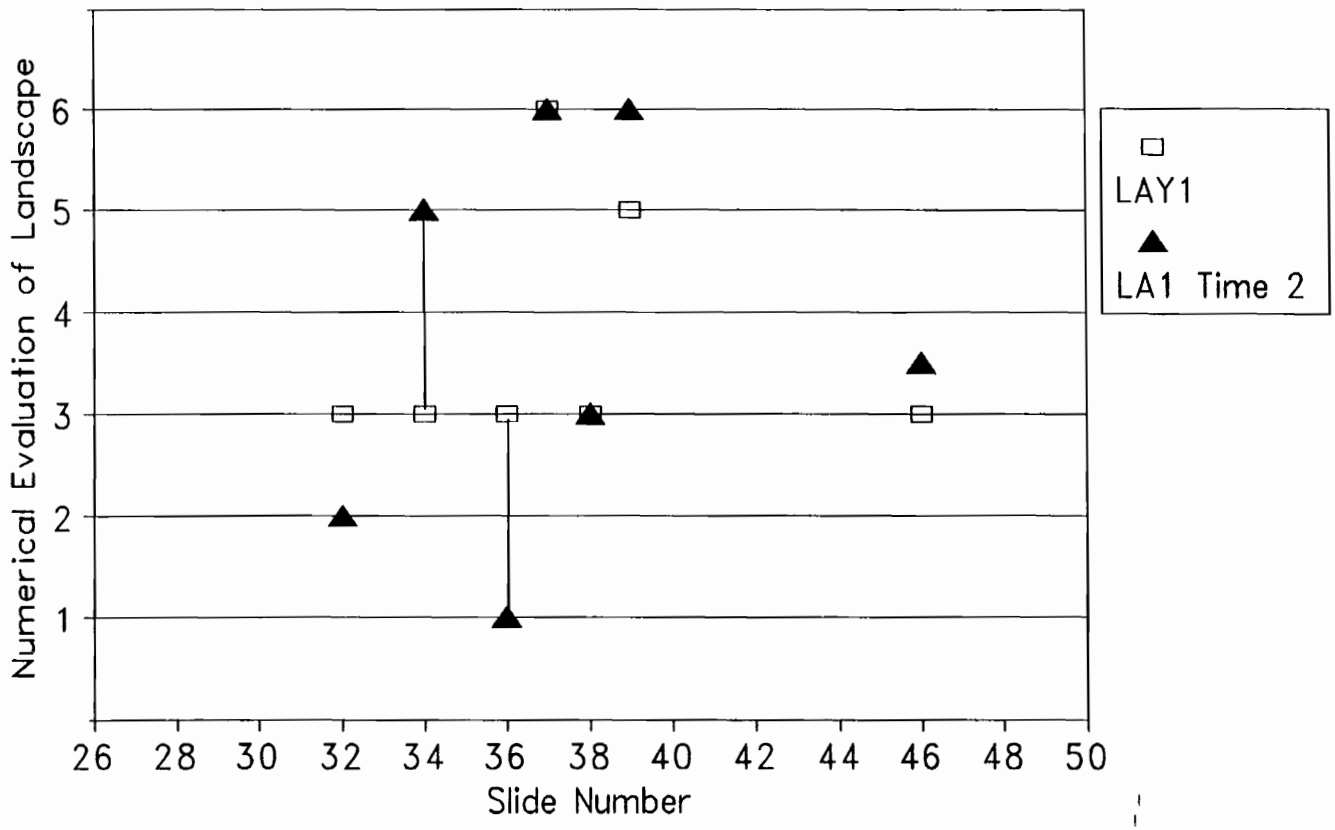

Figure 16E: Layman 1 vs. LA1 Time 2
Slides # 26–50

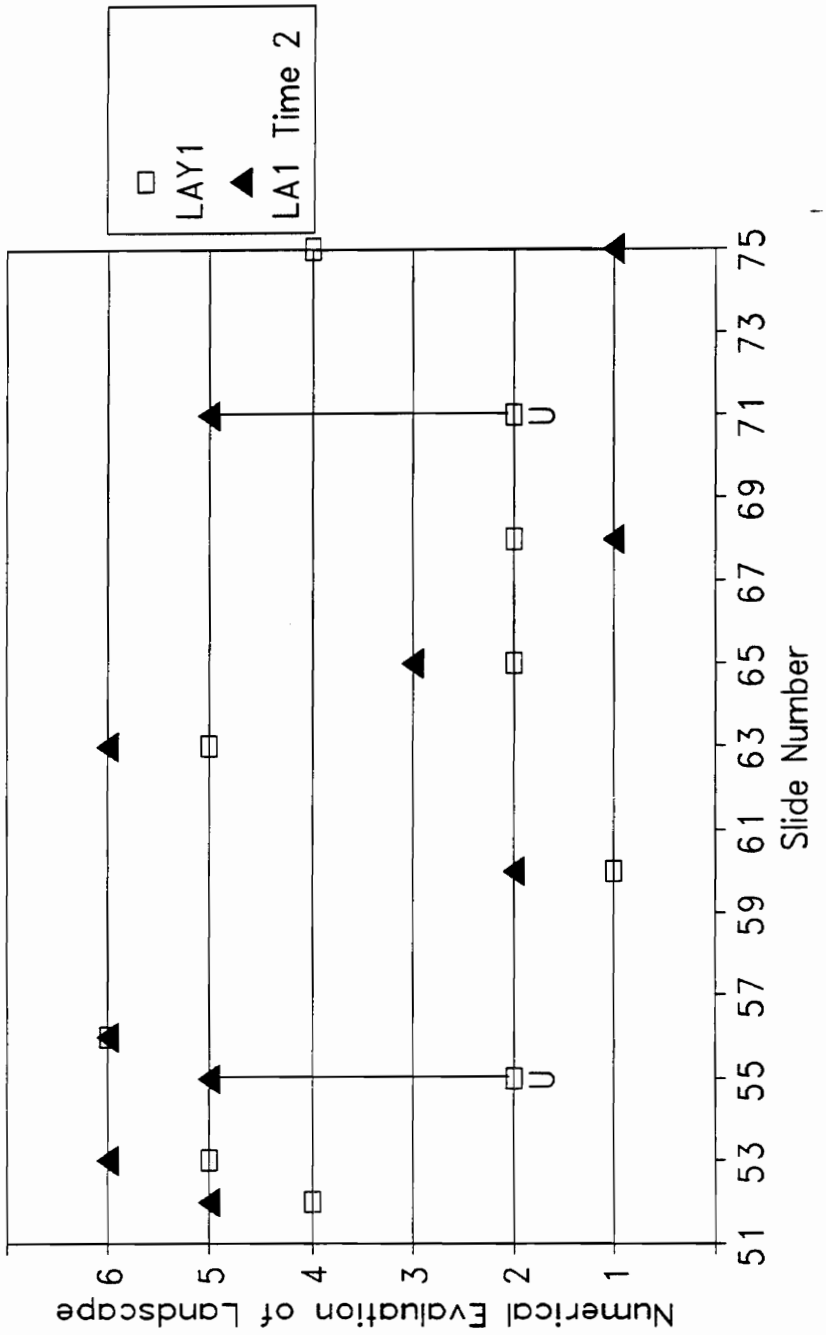

Figure 16F: Layman 1 vs. LA1 Time 2
Slides # 51–75

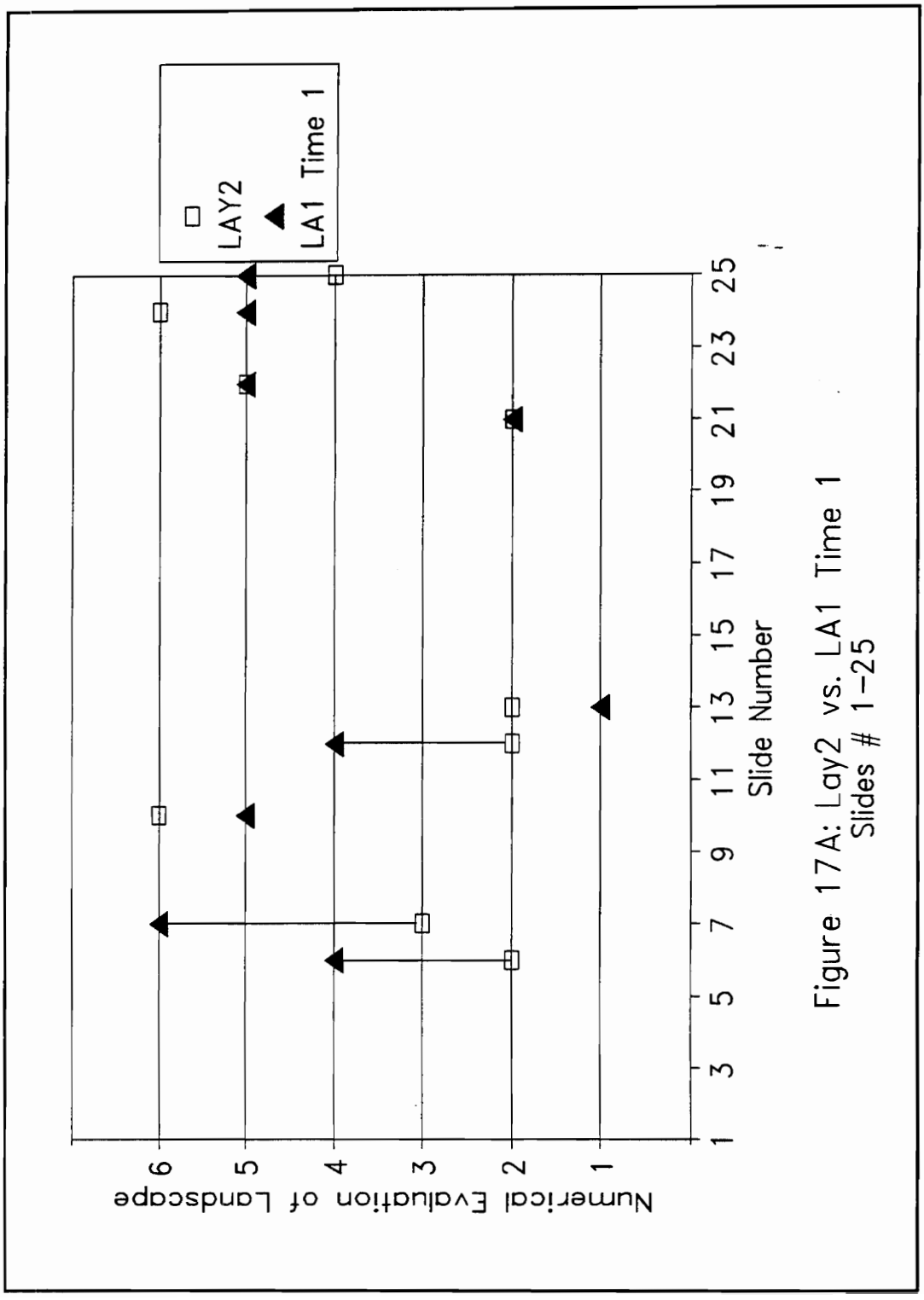

Figure 17A: Lay2 vs. LA1 Time 1
Slides # 1–25

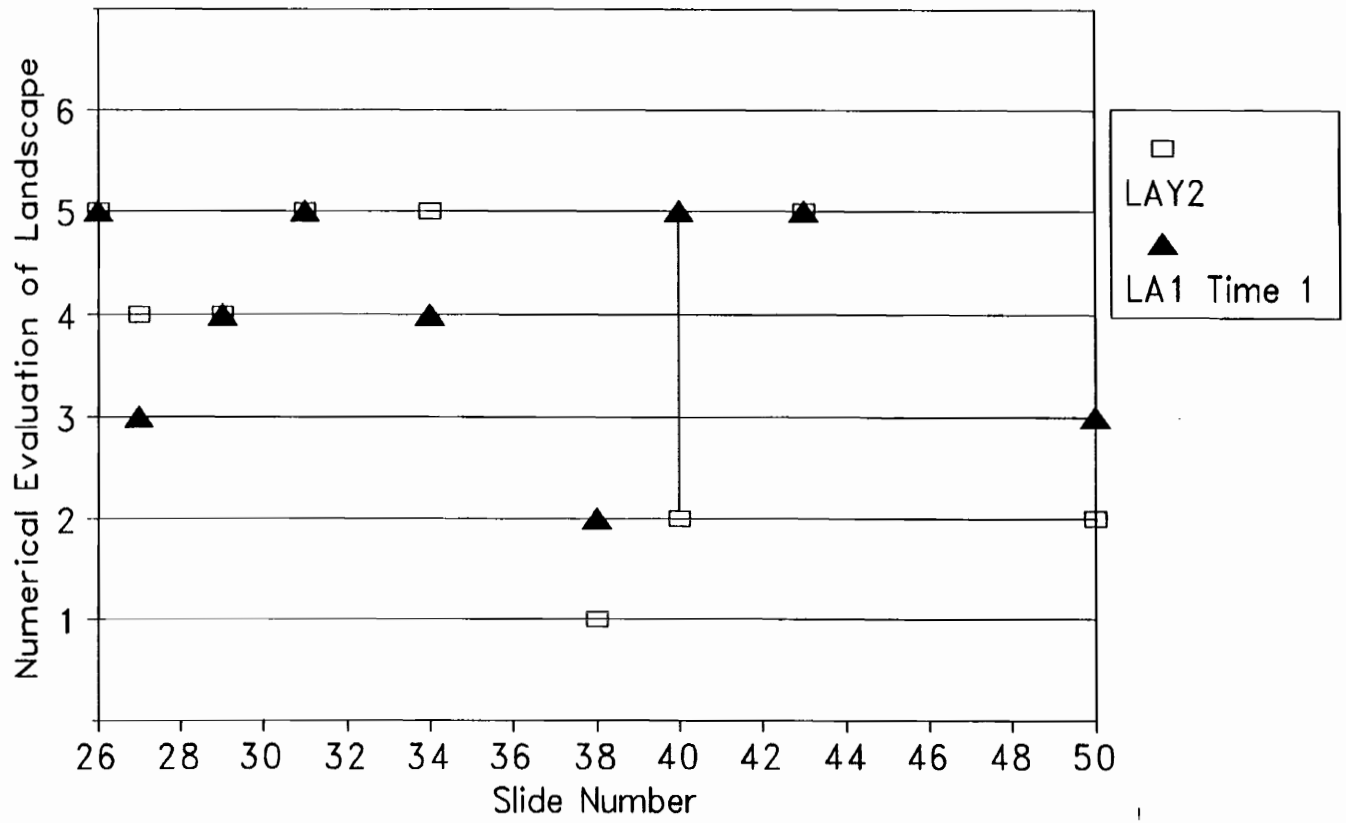

Figure 17B: Lay2 vs. LA1 Time 1
Slides # 26-50

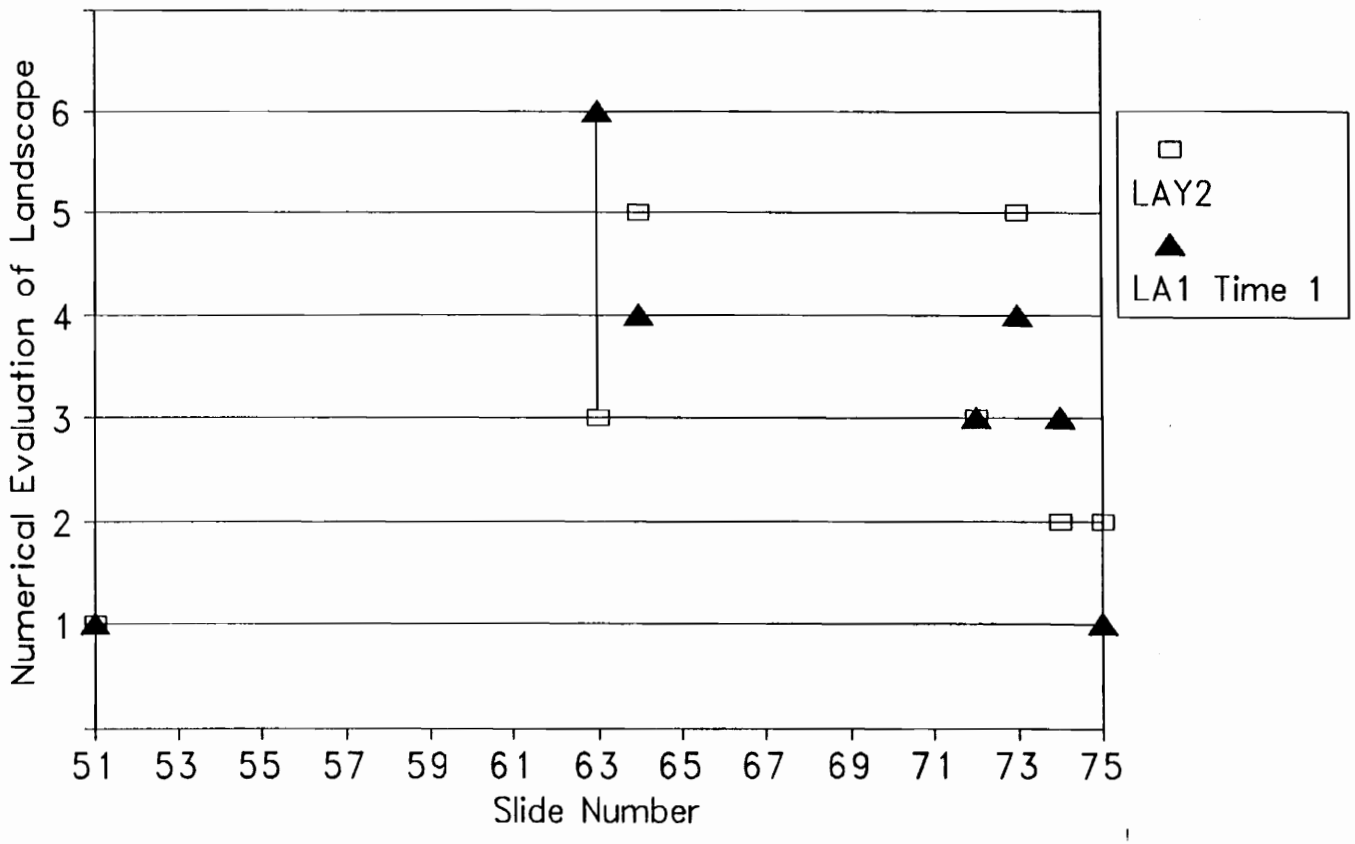

Figure 17C: Lay2 vs. LA1 Time 1
Slides # 51-75

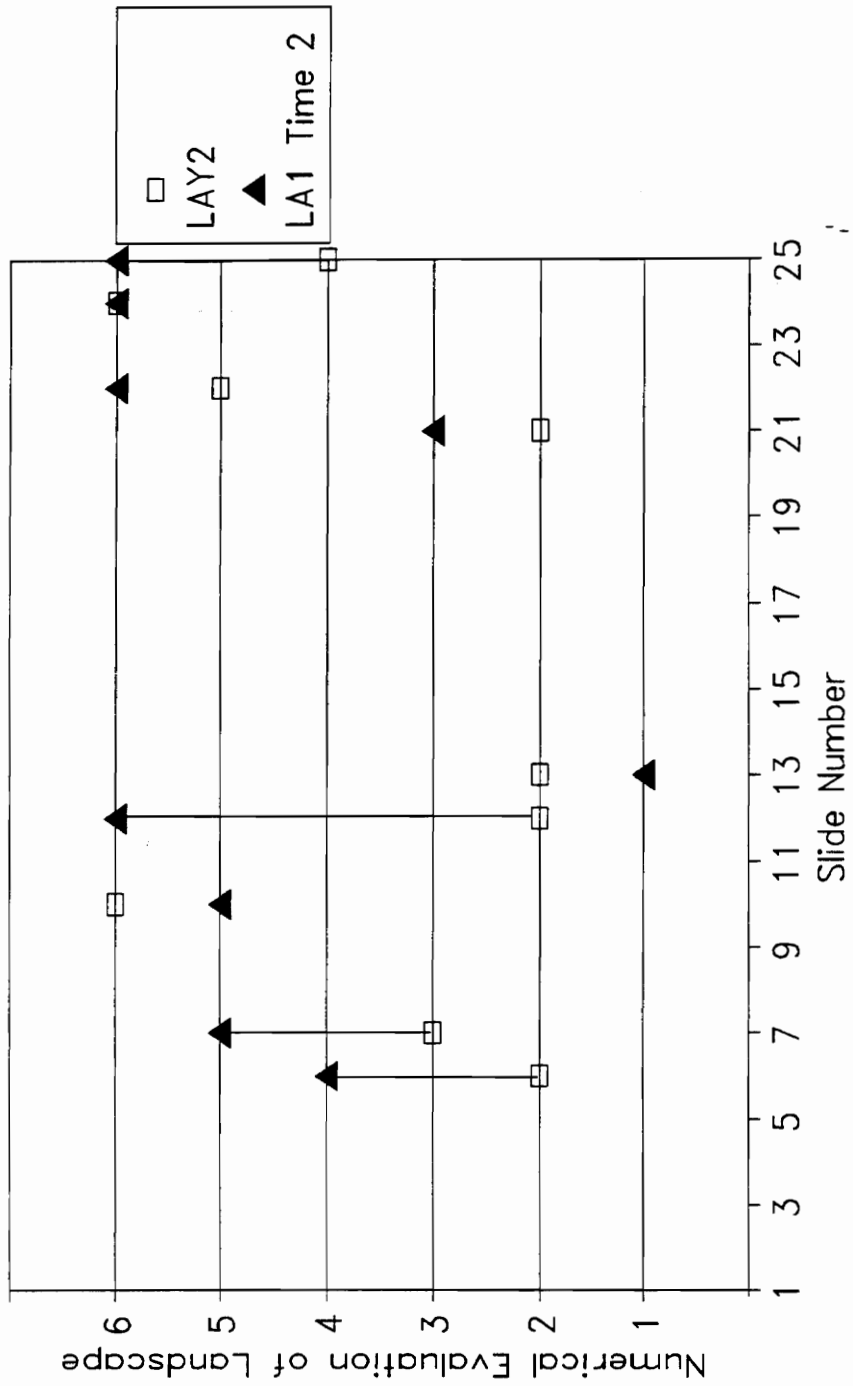

Figure 17D: Lay2 vs. LA1 Time 2
Slides # 1–25

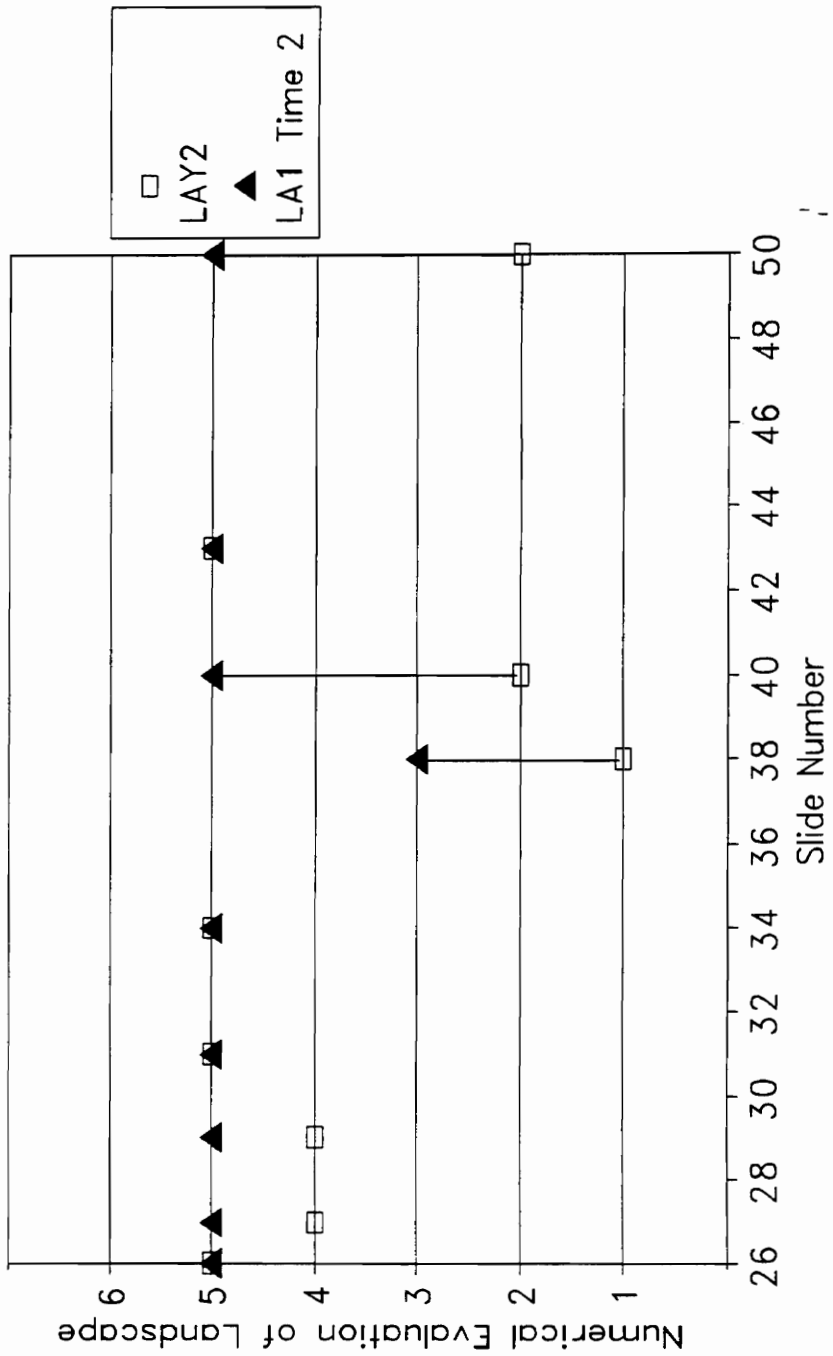

Figure 17E: Lay2 vs. LA1 Time 2
Slides # 26–50

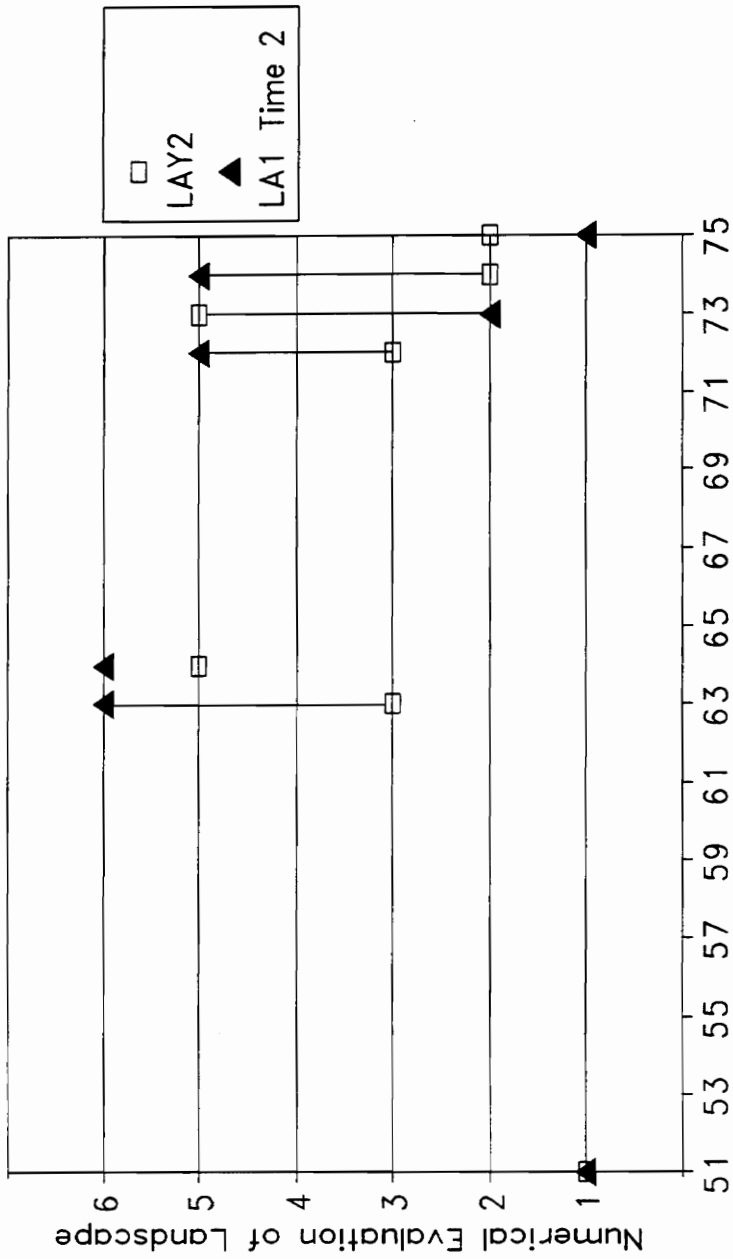

Figure 17F: Lay2 vs. LA1 Time 2
Slides # 51-75

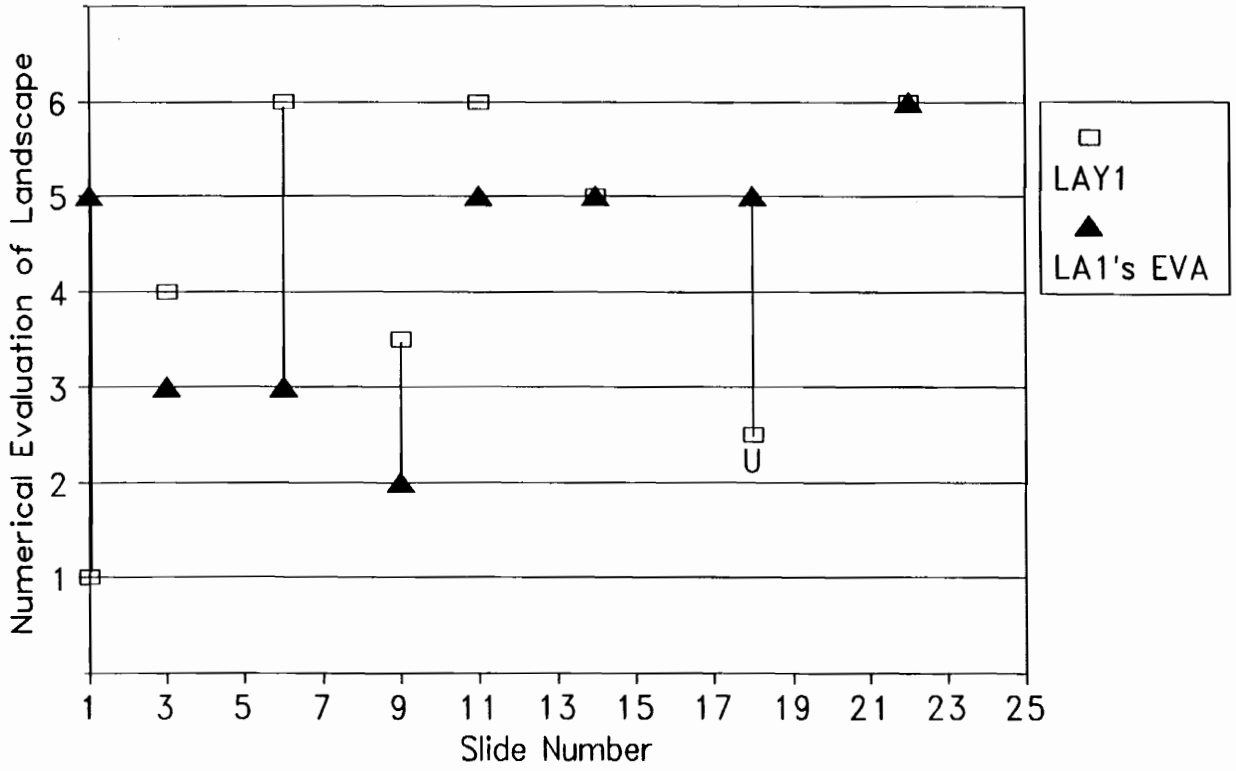

Figure 18A: Lay1 vs. LA1's EVA Time 1
Slides # 1–25

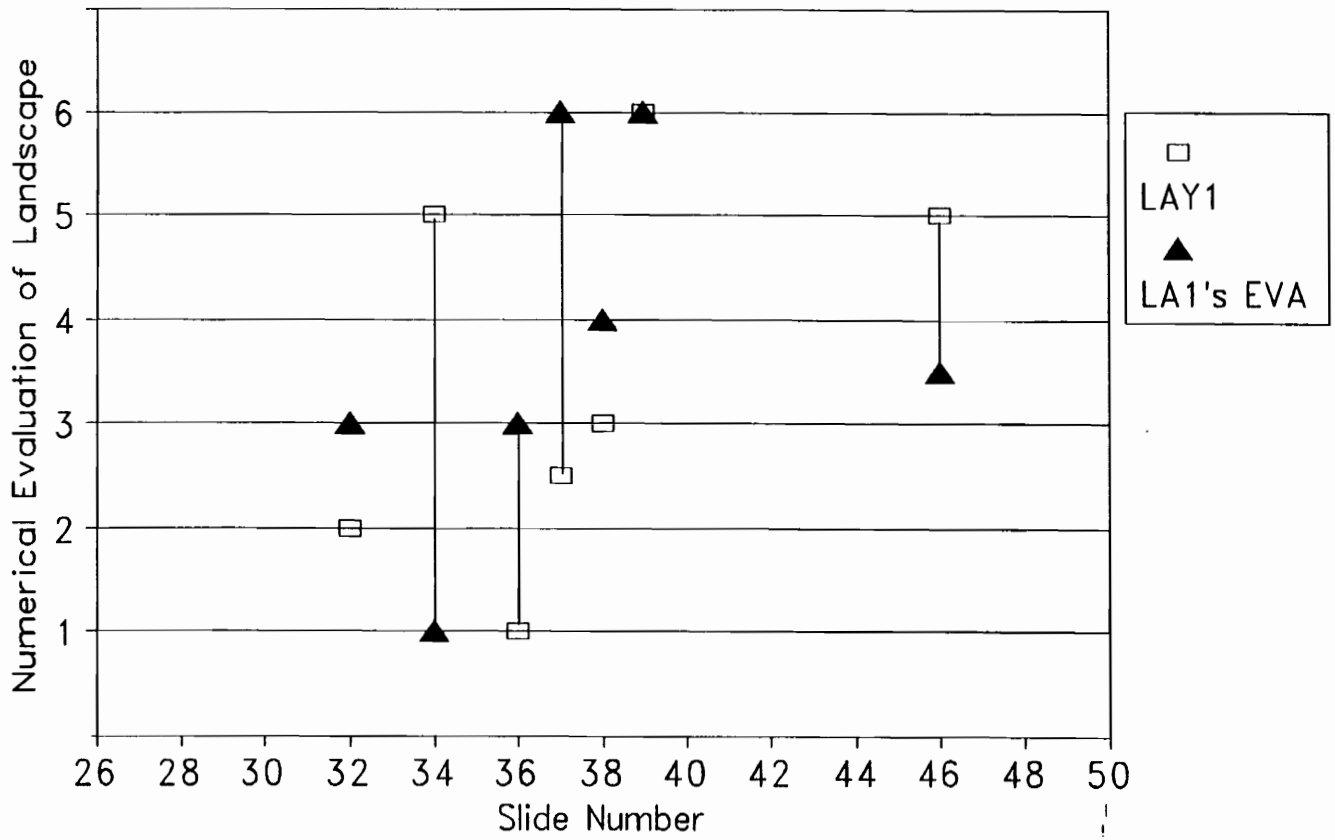

Figure 18B: Lay1 vs. LA1's EVA Time 1
Slides # 26-50

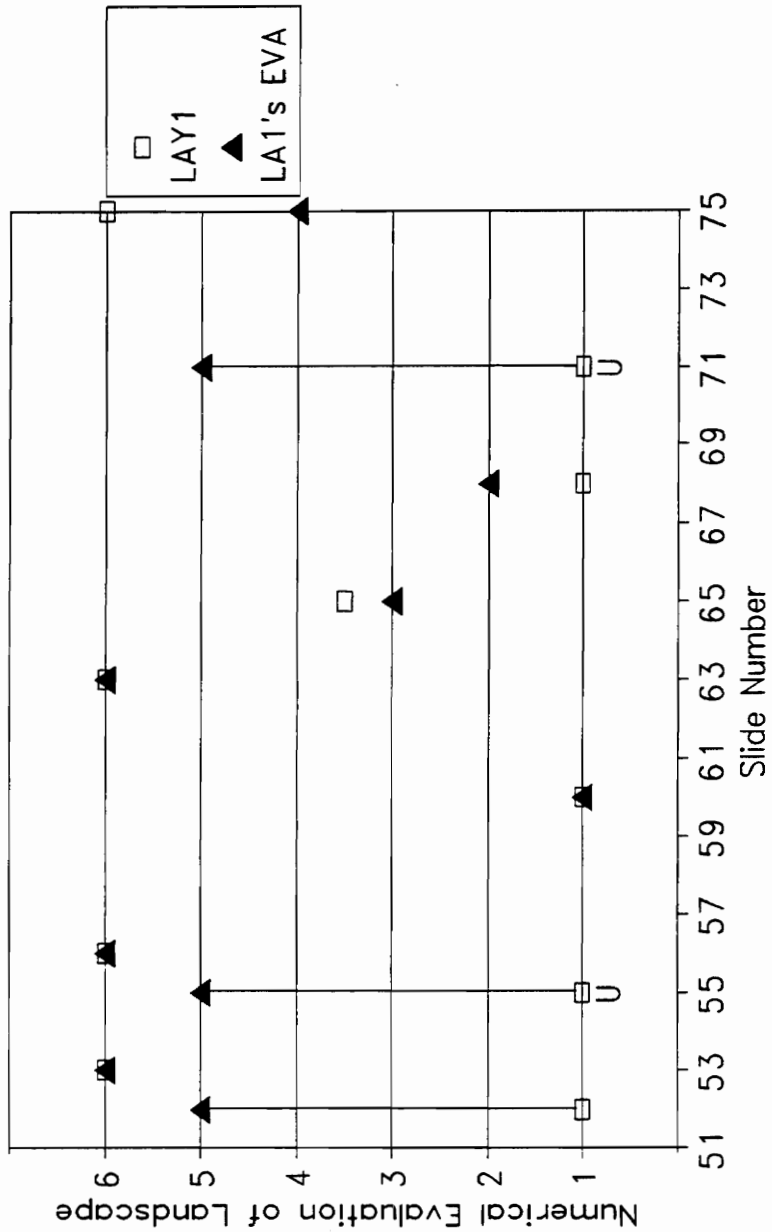

Figure 18C: Lay1 vs. LA1's EVA Time 1
Slides # 51–75

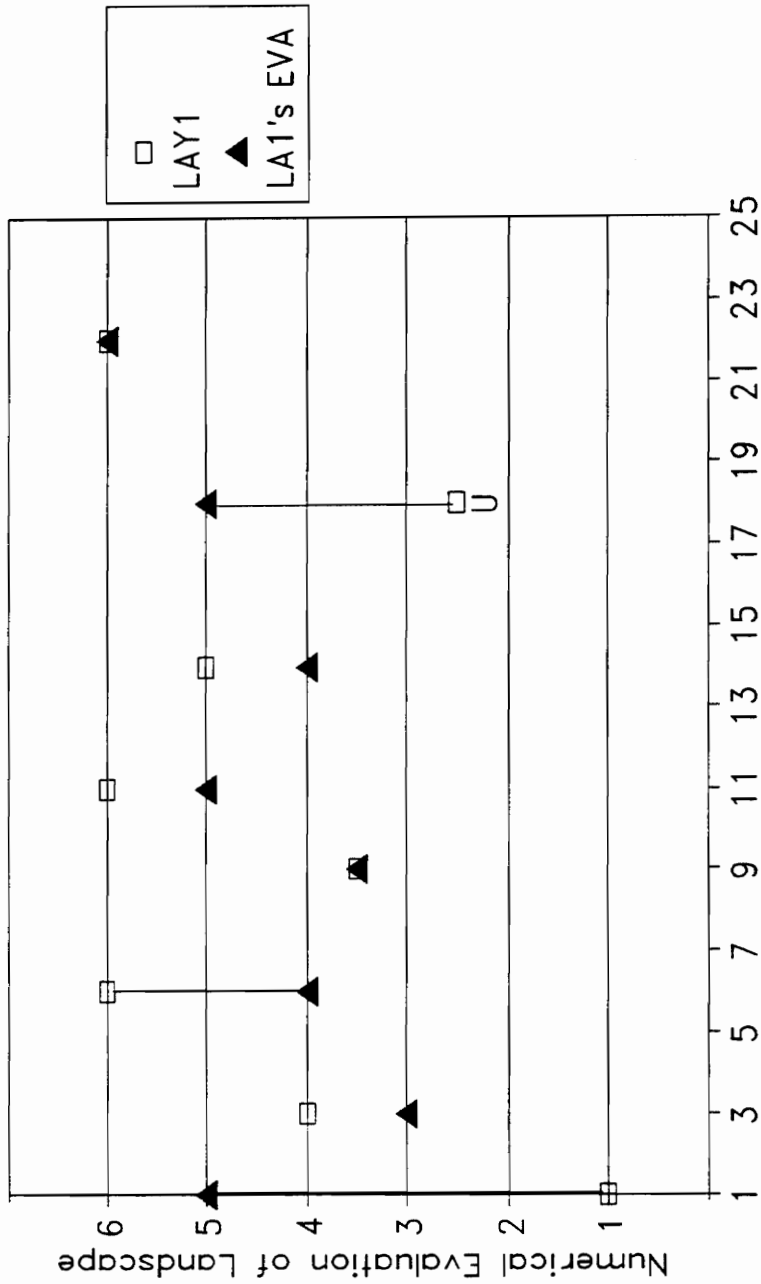

Figure 18D: Lay1 vs. LA1's EVA Time 2
Slides # 1-25

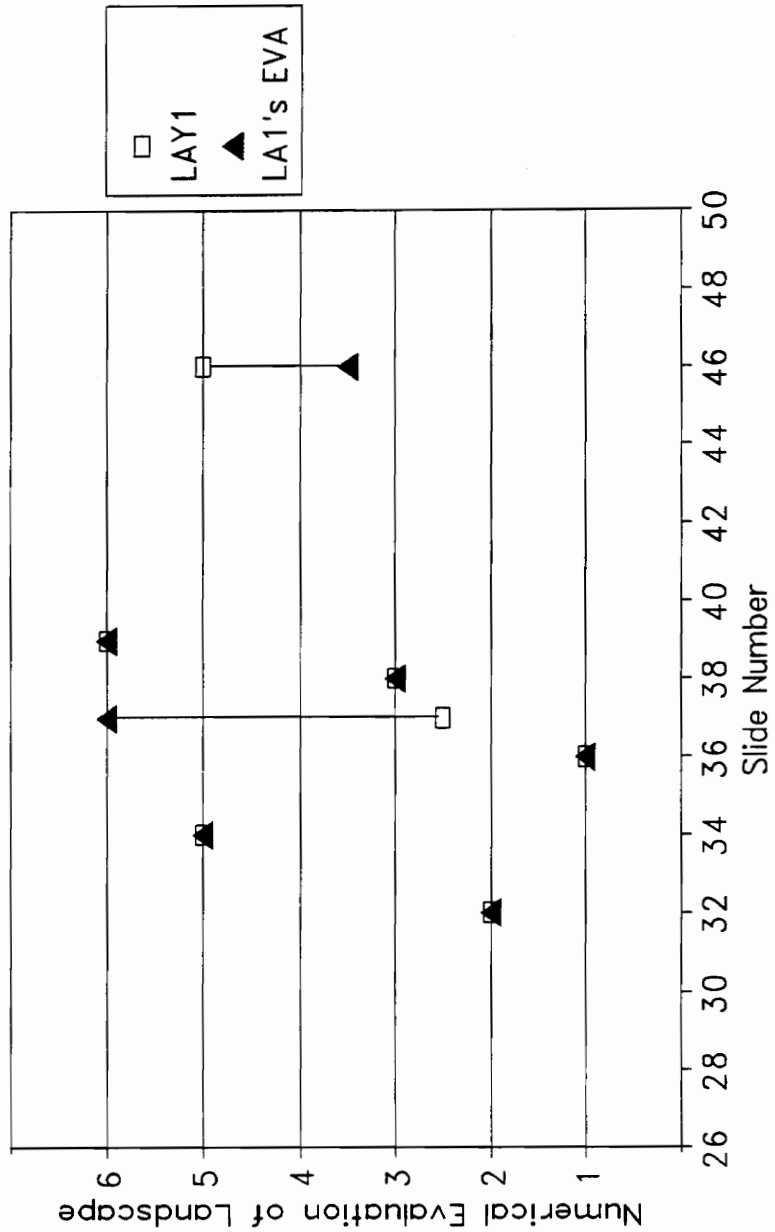

Figure 18E: Lay1 vs. LA1's EVA Time 2
Slides # 26–50

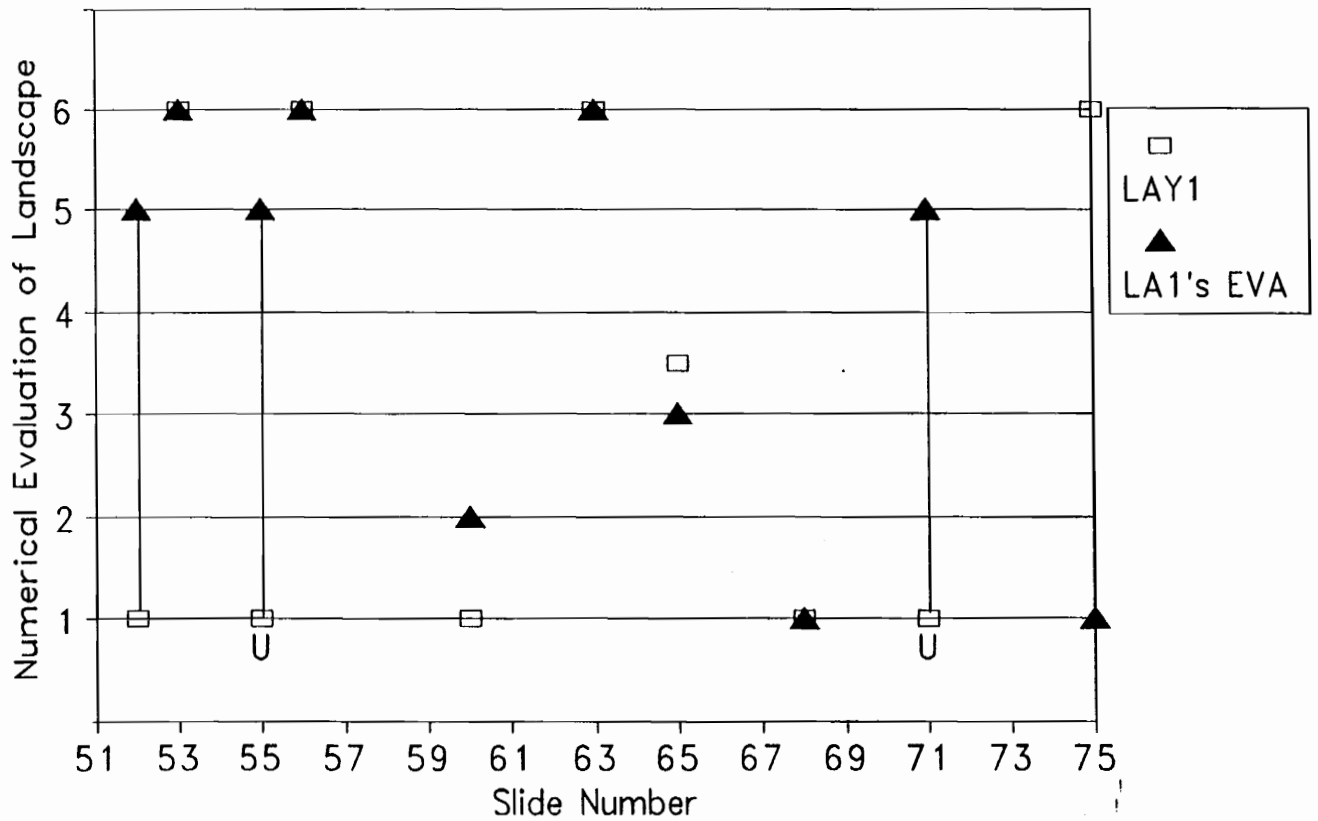

Figure 18F: Lay1 vs. LA1's EVA Time 2
Slides # 51-75

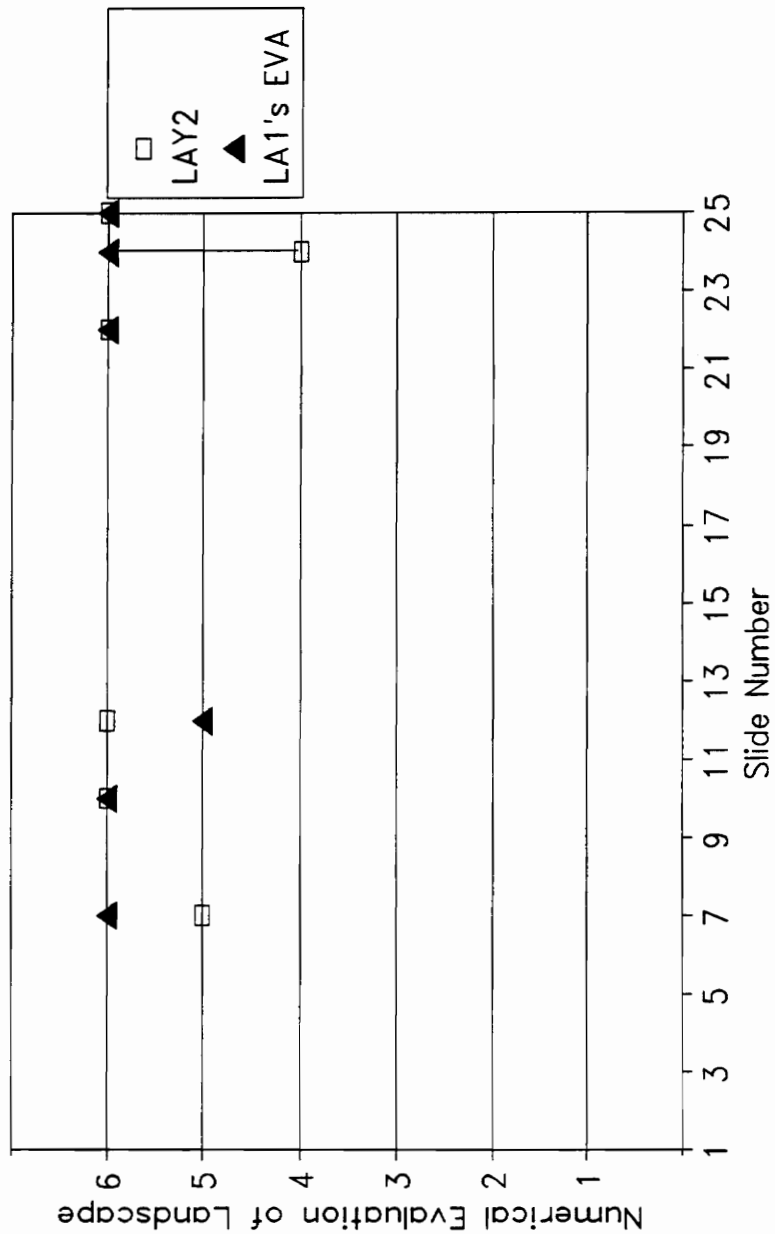

Figure 19A: Lay2 vs. LA1's EVA Time 1
Slides # 1–25

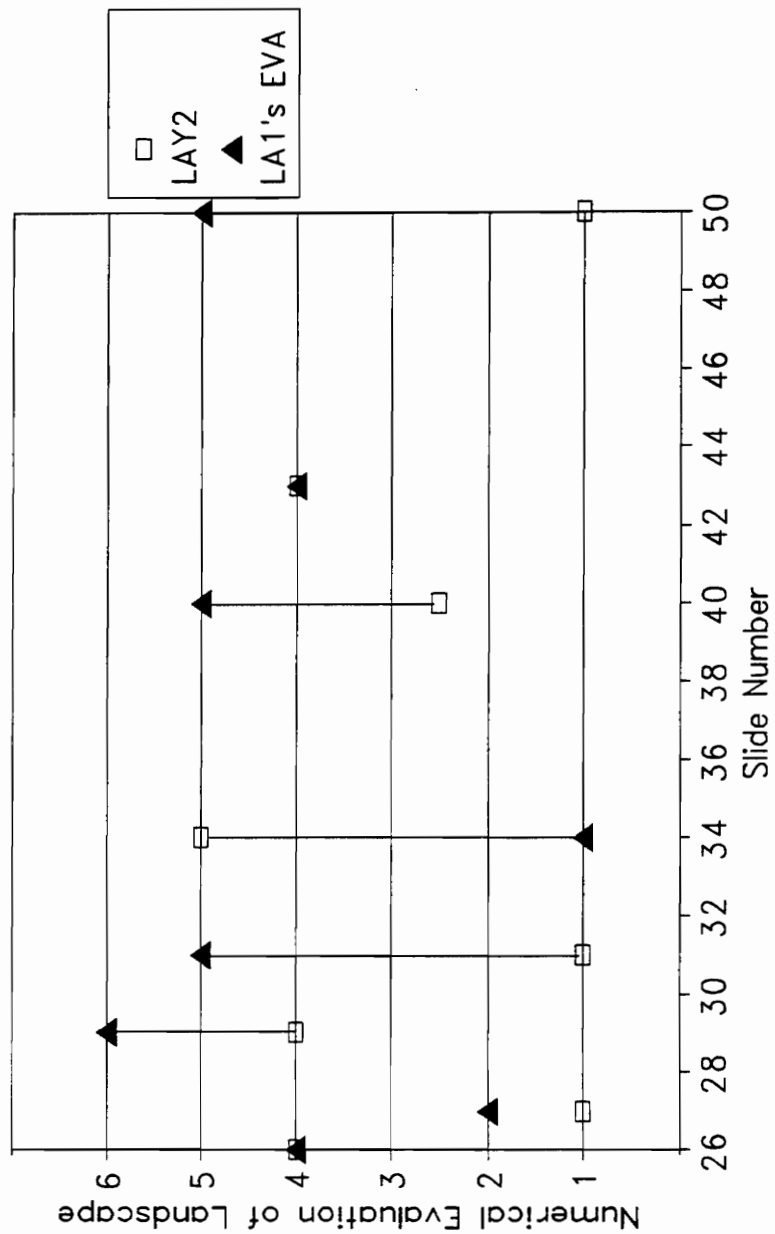

Figure 19B: Lay2 vs. LA1's EVA Time 1
Slides # 26-50

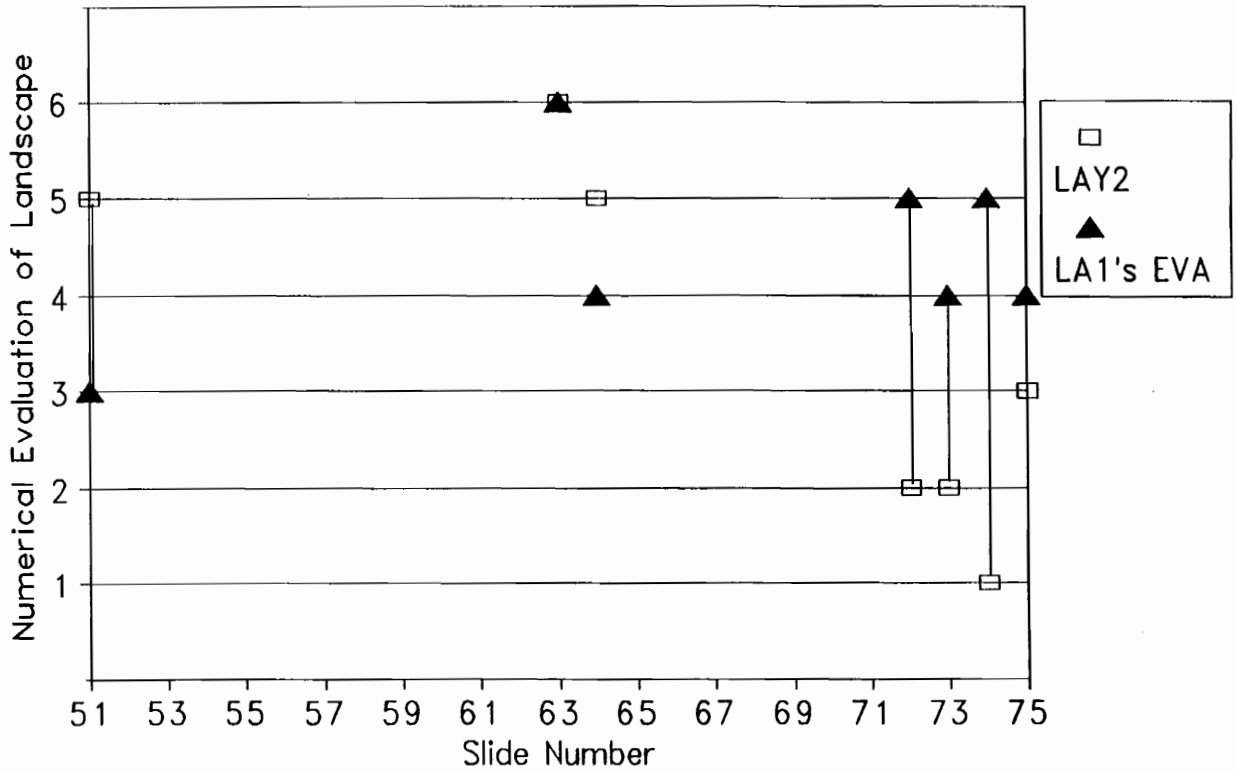

Figure 19C: Lay2 vs. LA1's EVA Time 1
Slides # 51-75

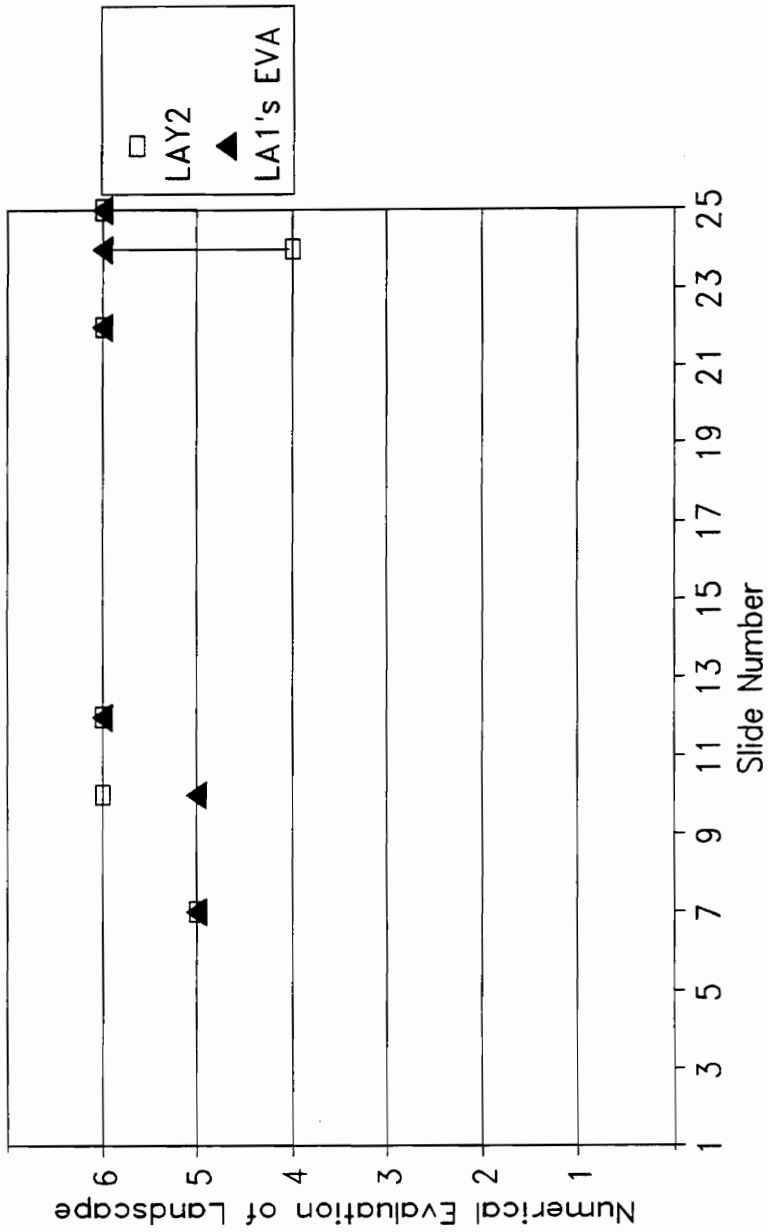

Figure 19D: Lay2 vs. LA1's EVA Time 2
Slides # 1–25

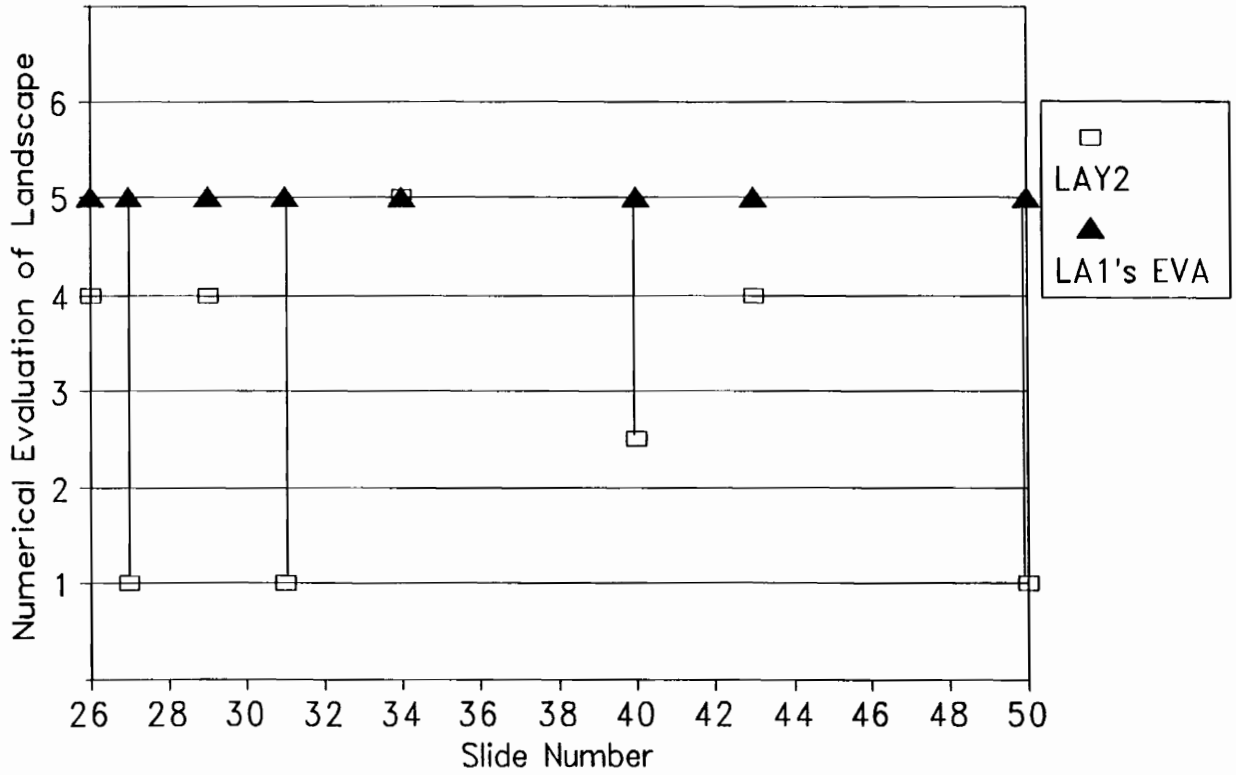

Figure 19E: Lay2 vs. LA1's EVA Time 2
Slides # 26-50

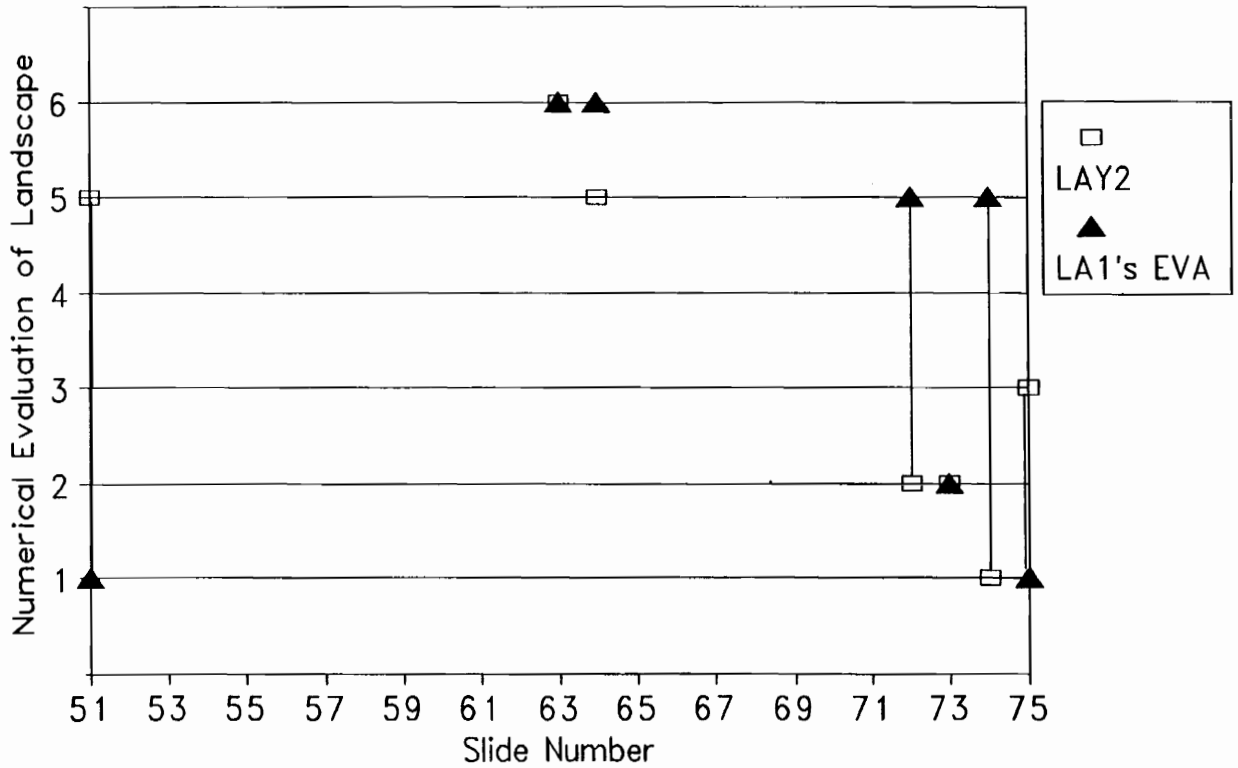

Figure 19F: Lay2 vs. LA1's EVA Time 2
Slides # 51-75

Vita

Don Schlagel was born in Longmont, Colorado on September 8, 1968. He received his B.S. in Forestry from Northern Arizona University in December of 1991 and began his MS studies at VPI in August of 1992. He will be relocating to Ames, Iowa, where his wife Deborah, has recently been accepted as a doctoral candidate in the chemistry department of Iowa State University.

Donald H. Schlagel