26
38

Comparing Two Post Occupancy Evaluation Methods with an Urban Plaza Test Case

by

Charles W. Ware Jr.

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Landscape Architecture

in

The College of Architecture and Urban Studies

APPROVED:

Dean R. Bork, Committee Chairman

Francis T. Ventre

Len Singer

Dr. Patrick Miller, Department Chairman

January 14, 1989

Blacksburg, Virginia

C.2

LD
5655
V855
1989
W373
C.2
Folio

Comparing Two Post Occupancy Evaluation Methods with an Urban Plaza Test Case

by

Charles W. Ware Jr.

Dean R. Bork, Committee Chairman

The Colege of Architecture and Urban Studies

(ABSTRACT)

Post occupancy evaluation is part of a design-evaluation-design cycle in which designers learn from their successess and mistakes and subsequently improve their designs. But, if designers want to make most effective use of information collected in such studies they must be done reliably and validly - few studies give evidence to justify such a claim. In the present study, two commonly and interchangeably used POE observation methods (direct observation and time-lapse photography) were comparatively tested in order to assess their reliability. Reliability concerns the extent to which different observers or the camera yield the same results in observing the same situation. The test case was conducted in a heavily used urban space and much of the data, from observer to observer, and observer to camera, was found unreliable. Reliability decreased as pedestrian frequency increased but not so uniformly that data from this study could be used to determine an exact number of persons that can be accurately mapped. Reliability "checks" should be made in pretesting of direct observations, also in retrieval of data from film. Direct observation and time-lapse photography can be used conjointly with the intent of using camera as an accurate basis against which to assess the relibility of direct obsservations, but with precaution taken to ensure the accuracy of camera data. Standards of reliabillity and validity, with simple tests or approaches to measuring them need to be developed in order to make it easier for researchers to "check" the reliability and validity of their findings.

Acknowledgements

Thanks to:

the Landscape Architecture Foundation, for funding in a Student Research Grant; the VPI & SU Graduate Student Assembly, for funding in a Graduate Research Development Project Award; Committee Chairman Dean R. Bork, professor of landscape architecture, whose good natured support, encouragement and reviews made for an enjoyable and rewarding thesis experience; Committee Member Dr. Francis T. Ventre, professor of environmental design and policy, whose uncompromising demand for articulation and attention to detail led to a much improved product; Committee Member Len Singer, professor of architecture, for introducing me to the field of environment and behavior studies, providing a base with which to begin such research and also inspiration to continue in the future; Dr. B. Skarpness, professor of Statistics, of the VPI & SU Statistical Consulting Center, whose complete assistance in the choice, use and interpretation of statistical tests was invaluable; Nick O'Donohue, professor of technical writing, whose guidance in the preparation of grant proposals helped secure funding; second year landscape architecture students Kevin Hall, Virginia Flanagan, and Andrea Harrison, for their time volunteered in the collection of data; Tim Korbelak, Harborplace Project Director, from Wallace Roberts and Todd, of Philadelphia, for interest in and suggestions on how the effort would be most valuable to designers; William H. Whyte, Craig M. Zimring and those others whose early insights sent me in the proper direction; all the faculty, staff and students of Virginia Tech Landscape Architecture, who made the graduate experience a great one; my family, who are always supportive and who I love very much; and Meghan, for waiting.

Table of Contents

List of Illustrations

List of Tables

Introduction

As with many emerging fields, post occupancy evaluation (POE) is beset by considerable methodological eclecticism. A great number of measurement methods are available in the man-environment field of study, but certain methods are more suitable for certain research purposes than others. Lozar (1980) recognized a need for a categorization of methods and attempted to develop a comparative analysis of them within a taxonomic structure. The taxonomy was aimed partly to allow researchers to select methods best related to their research purposes. The taxonomy represented only a review of current evaluation works; it involved no real assessment of the methods themselves. Rigorous reliability and validity testing of POE methods would allow future researchers more confidence in both the methods they choose and the results they obtain.

Friedmann et al. support the use of post occupancy evaluation as part of a design-evaluation-design cycle in which "designers learn from their successes and mistakes and subsequently improve their designs."[1] They describe the evaluation phase as one where the designer/researcher first defines the problem or "relationship of special concern," such as the impact of housing regulations on residents' satisfaction. Next one defines how the focal

---

[1] Friedmann, A., et al., *Environmental Design Evaluation* (New York: Plenum Press, 1978), p. 20.

problem may be affected by other influences - the "larger system." The larger system can consist of factors such as the design process, users, setting, social-historical context or the proximate environmental context. Definition of the focal problem and larger system then should guide the selection of appropriate measurement methods and hopefully the gathering of relevant data. Friedmann et al. propose that such a routine "can increase the quality of information."[2]

The present author proposes further development of this evaluation phase to ensure "quality information," i.e. - reliable and valid data. The figure below displays the proposed modification to the cycle, as indicated by tone.

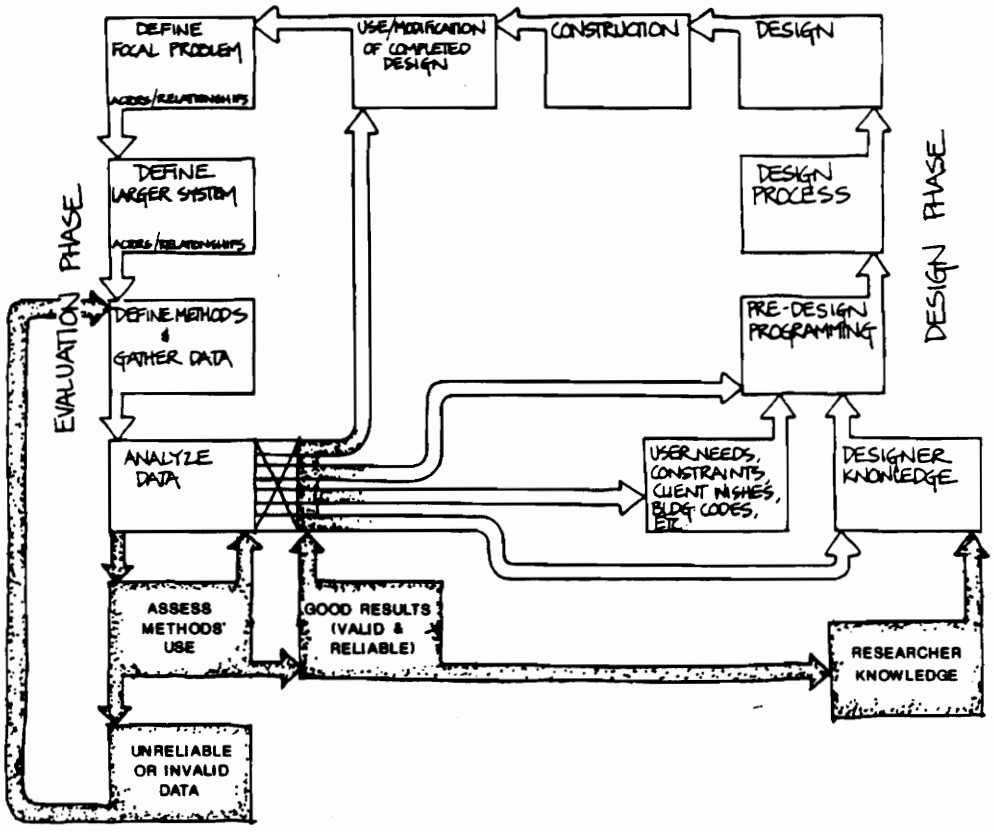


Figure 1. Modified Design-Evaluation-Design Cycle: Original Source: Friedmann, A., et al., *Environmental Design Evaluation* (New York: Plenum Press, 1978), p. 21.

---

[2] Ibid., p. 26.

Schematically, the designer/researcher follows the evaluation phase through to beginning data analyses. Then, concurrently with later analyses, the methods' effectiveness should be assessed by reliability and validity tests, such as "convergent validity," and "test-retest" and "alternate forms" for reliability. Results of data analyses should not move toward the design phase until "good results" are proven. If measurement methods are shown to be performing invalidly or unreliably their use should be redefined before gathering additional data.

The major objectives of this study are:

- To develop a POE research design for an existing urban space, addressing a relevant, site specific research question, utilizing two statistically comparable *types* of a single POE research *method*;[3]

- In applying this research design, to conduct independent tests of reliability and/or validity upon each type, providing a statistical basis to compare the effectiveness of each type in gathering behavioral information;

- To also compare necessary resources and ease of data analysis for each type in order to determine the general appropriateness of the types of method under study.

The intent is not to assert that a single method can replace a multi-method approach in providing a true perspective on the environment. Two types of only one method are chosen in order to most readily and efficiently conduct a comparative analysis.

---

[3] A *method* is considered as the most gross descriptor of measurement techniques used in environmental evaluations. Methods include observation, survey/questionnaire, interview, physical traces and document review. *Types* are kinds of each method. For instance: behavior mapping, videotaping, and descriptive note-taking are *types* of an observation *method*.

Note that a well-defined, site-specific research question shall be articulated before deciding upon types of a POE method to be compared. The choice of techniques (types) should be a response to the serving of this purpose. As an example, a researcher interested in studying movement patterns in a small, well-defined area in order to fine tune its layout might choose either a hodometer or time-lapse photography as a measurement technique. A researcher interested in studying movement patterns on a larger scale with less concern for precise findings useful for future designs might not require instrumented techniques. Instead he might choose direct observation or even an indirect method such as examination of physical traces or erosion tracks.

A relevant, site-specific research question for this study will be established in response to interests expressed by site designers. This study, however, is not aimed at application of evaluation results; it concentrates instead on a comparison of research methods. Generalizations are intended from results of such a comparison, but only to similar studies in similar settings. Data and results gathered in response to the research question can only be site-specific. This approach will ensure an emphasis on improvement of the method itself and hopefully will be of use to others attempting similar studies.

Organization of the Report

The report consists of seven major chapters. In Problem/Case Study Review (Chapter I), the subject of post occupancy evaluation is introduced by reviewing case studies in order to define a problem area with regard to recent work at the site scale. A research site for the present study is then selected and briefly described (Chapter II), resulting in a focus and research design for the evaluation itself (Chapter III). The middle section of the paper (Chapter IV) is devoted to a description of the larger study's focus, how the evaluation methods are to be assessed, culminating in a list of statistical test questions with accompanying null and alternative hypotheses. Results (Chapter V) are presented in reference to the support or rejection of each hypothesis, with a summary of other findings ending the chapter. A discussion of results includes possible explanations and interpretations (Chapter VI). Conclusions (Chapter VII) includes implications of the study to the larger body of research and recommendations derived from such implications.

For the designer or other reader not interested in the statistical basis/body of the study, he should only skim Chapters IV and V. Chapter VI, Discussion, does not exclude statistical analysis, but is in summary form; it should be read by those wishing to understand the basis of Conclusions. The reader is also referred to a bibliography which contains case studies of

evaluations completed at the site scale (exterior space) and methodological and general references on the subject of post occupancy evaluation.

I. Problem/Case Study Review

Introduction

Landscape architects and behavioral scientists have inherently different interests and attitudes that limit the effectiveness of their collaboration. But, "when researchers and designers do cooperate, each [can] use the other to do more than either can do alone; researchers to have designers use and improve their information; designers to have researchers help close the gap between them and their ... user clients."[4] One opportunity for cooperation is evaluation research of completed projects in use.

Post occupancy evaluation is the "examination of the effectiveness for human users of occupied environments. Effectiveness includes the many ways that physical and organizational factors enhance achievement of personal and institutional goals."[5] Such examinations are becoming more common, yet diverse in scope. Zimring et al. (1980) reviewed recent POE

---

[4] John Zeisel, *Inquiry By Design, Tools for Environment Behavior Research* (Cambridge: Cambridge University Press, 1984), p. 50.

[5] Craig M. Zimring and Janet E. Reizenstein, "Post Occupancy Evaluation, An Overview," *Environment and Behavior* 12, No. 4 (1980), p. 429.

work ranging from brief academic projects to well funded longitudinal studies, and included among their "researchers" untrained "users," social science consultants, academicians and practicing designers.

But designers neither sponsor nor conduct many of such studies. In a 1978 review of 165 evaluations Bechtel and Strivastara reported that among the 92% of the studies that received funding, only 3% were sponsored by design firms. Friedman et al. (1978) reported that designers comprise only 5% of the occupational groups performing government- and university-funded evaluations. Friedman et al. also reported that public spaces, such as those designed by landscape architects, were evaluated far less frequently than other settings.

Twenty-seven evaluations of site-scale public space completed since 1970 were reviewed as a basis and justification for the present study. The review is not meant to be comprehensive in terms of the number of studies considered. It is, however, deemed a representative sample of such POE work; a range of settings, from plazas to campus spaces and housing sites are included. There was no attempt made to select studies based on any criteria other than that of being a site-scale public space - the first-found 27 of such studies were reviewed. Note that the findings represent only what was documented in the report reviewed.

Table 1 (Case Study Review) lists the studies by principal author[s] (columns 1 - 27). A full citation of each entry is included in Reviewed Case Studies under Selected Bibliography. The studies are listed in rough chronological order by when the actual study was undertaken, not the publication date. If the study date was unknown, the date of publication was used.

The case studies were reviewed based on general criteria germane to the present study's purpose. Each criterion heading (e.g. - Setting, Purpose) is subcategorized (e.g. - under Setting - plaza, park) although many studies "fit" into more than one subcriterion of several headings. Where a categorization is unknown for a study, the heading is left blank. The last

column in the table represents the percentage of studies of the total (27) that were included or described by the coinciding subcriterion.

Definitions of Review Criteria

- RESEARCHER TYPE - describes those who principally conducted the study, can be described by more than one subcriterion, as with interdisciplinary teams.
  - *design academician/student* - including landscape architecture, planning, etc.
  - *sociologist/environmental psychologist* - from either private-practice or academia.
  - *private-practice designer* - of any environmental design discipline.
  - *government agency* - with design responsibilities, such as department of parks and recreation.
  - *other* - organization or independent consultant where it is not known to which subcriterion its researchers "fit," such as practices in environmental design research, including Projects for Public Spaces and People Environment Group.
- SETTING - place[s] of study focus.
  - *plaza* - for the most part, largely paved, "hard" space with no designated active recreational purpose such as baseball or jogging; usually bounded by buildings; includes public courts or squares, campus spaces and vest pocket "parks."
  - *park* - largely green open space, designated for public recreational use.

- *playground* - specially purposed for children's outdoor games and recreation, with play structure[s].

- *housing site* - courtyards and other exterior spaces found between residential units in multi-family housing communities.

- *miscellaneous* - includes zoo spaces, National Park Service visitor centers, campus quadrangles and other settings that cannot be included in the above categories.

- NUMBER of SITES

  - *single* - one site as study focus.

  - *multiple* - more than one site, either as comparative study, or with subsites forming a whole environment (campuses, downtown areas), or simply as larger scoped, multiple-site evaluation.

- LENGTH of STUDY - basically what is reported for duration of data collection, not necessarily including data analysis.

- PURPOSE - goals and/or objectives, usually stated up front, or sometimes inferred from study results.

  - *evaluate quality or success* - not only in reference to designers' objectives or assumptions, but also criteria set by evaluator or another, such as Love's (1973) evaluation of Forecourt and Lovejoy plazas where she uses Jacob's (1961) four requisites for a sucessful plaza; also includes understanding relationship between design elements and use, e.g. - "what works?"; also, simply to see how space functions or answer general questions such as: "How safe is the space?" or, "To what extent is a barrier-free environment provided?"

- *behavioral information* - focused specifically on how people act or react in an environment, more from a behavioral than design viewpoint.

- *identify open space needs* - stated specifically as such, no suggestion of translation into physical design solutions.

- *propose design guidelines* - conceptual design ideas, applicable to more than site[s] under study.

- *information for redesign* - particular suggestions to refine or redesign the existing space under study.

- *methodological insight* - at least some part of purpose devoted to examining the quality or application of methods used in the evaluation; for example, by Rutledge (1975): "The nature of [this as] a pilot study is to investigate methods, refine data-gathering instruments ...", and the present study: "... to compare the effectiveness of each type [of POE observation method] in gathering behavioral information."

- METHODS' ASSESSMENT - how evaluation methods were assessed, if they were at all, either by performers themselves or others.

  - *brief critical description* - as part of discussion, with no real ordering or predetermined method of assessment; including phrases like: "... method is less suitable for ...", "... provides a richer source of data"..., and some even in the form of recommendations: "There are a number of promising directions for methodological improvement ..."

  - *detailed objective evaluation* - study most always purposed at least partly as methodological inquiry; either lists criteria to evaluate methods, or limitations, pros and cons, advantages and disadvantages; also on more detailed level than brief

critical description, includes points such as: "The absence of a large sample leads to grouping problems and completely eliminates valid cluster analysis"; formal tests of reliability and/or validity, however, are not included.

- *quantitative tests* - of validity and/or reliability, such as "convergent" validity, "test-retest" for reliability; also "interobserver" and "alternate forms" reliability tests like those conducted in the present study.

Table 1. Case Study Review

	1 Love	2 Reynolds, Nicholson	3 Brower	4 Moore	5 Nager, Wentworth	6 Marcus	7 Palmer, Crystal	8 Rutledge	9 Kueffer	10 Cohen et al.	11 Share	12 Preiser et al.	13 Kantrowitz, Nordhaus	14 Burden (PPS)
RESEARCHER TYPE														
academician/student						•	•	•	•	•		•	•	
sociologist/env. psych.	•	•			•								•	
priv.-pract. designer		•												
government agency			•				•							
other		•		•							•			•
SETTING														
plaza						•		•		•	•			•
park	•				•				•					
playground				•										
streetscape			•											
housing site		•											•	
miscellaneous							•					•		
NO. of SITES														
single				•	•			•	•	•				•
multiple	•	•	•			•	•				•	•	•	
LENGTH of STUDY														
< four days						•				•				
4 days - 2 weeks								•						
2 weeks - 1 season	•				•		•				•	•		
> 1 season			•	•					•					•
PURPOSE														
eval. quality or success	•				•	•	•	•			•	•	•	•
behavioral information		•	•	•										
identify open space needs				•					•		•			
propose design guidelines	•	•	•				•				•	•	•	•
info. for redesign					•					•		•	•	
methodological insight								•	•					
METHODS														
observation	•	•	•	•	•	•		•	•	•	•	•	•	•
survey/questionnaire				•	•				•	•		•		
interview	•	•	•		•		•	•	•		•	•	•	•
physical traces		•	•				•						•	
document review			•		•			•						•
METHODS' ASSESSMENT														
brief critical description		•		•	•					•				
detailed objective evaluation								•	•					
quantitative tests														

Table 1. Case Study Review

	15 Miles, et al.	16 Allor, Murphy	17 Dozio, et al.	18 Van Valkenburgh	19 Kaplan	20 Allor, Howe	21 Aguar	22 Francis, Girot	23 Martin, O'Reilly	24 Im	25 Brown, Burger	26 Anderson	27 Chidister	PERCENT of ROW
RESEARCHER TYPE														
academician/student		•	•	•		•	•	•	•	•		•	•	63
sociologist/env. psych.		•		•	•	•								30
priv.-pract. designer														04
government agency			•											11
other	•													19
SETTING														
plaza	•	•	•		•	•							•	41
park														11
playground				•							•			11
streetscape							•	•						11
housing site												•		11
miscellaneous									•	•				15
NO. of SITES														
single		•	•		•	•			•					41
multiple	•			•			•	•		•	•	•	•	59
LENGTH of STUDY														
< four days		•	•										•	19
4 days - 2 weeks	•													07
2 weeks - 1 season				•	•		•	•	•		•			41
> 1 season						•				•		•		26
PURPOSE														
eval. quality or success	•		•		•	•	•			•	•		•	63
behavioral information		•	•			•			•		•			30
identify open space needs	•												•	19
propose design guidelines	•	•		•					•					44
info. for redesign								•						19
methodological insight						•	•	•		•		•		26
METHODS														
observation	•	•	•	•		•	•	•	•		•	•	•	89
survey/questionnaire				•	•				•	•		•		37
interview	•	•				•	•		•			•		63
physical traces											•			19
document review	•			•			•					•		30
METHODS' ASSESSMENT														
brief critical description						•							•	22
detailed objective evaluation							•	•				•		19
quantitative tests				•						•	•	•		15

Discussion

Researchers conducting the 27 evaluations reviewed represented similar occupational backgrounds to those in the aforementioned reviews (pp. 7-8). Reinforcing the findings of Friedman et al. and Bechtel and Strivastara, practicing designers took part in only 4% of the studies reviewed. Meanwhile, design students and/or academic groups were involved in the majority of the evaluations (63%). Designers can make better use of such studies by becoming more frequently and directly involved in their formulation and administration.

Settings included a fairly even distribution of parks, playgrounds, streetscapes and miscellaneous sites, but plazas were evaluated four times as frequently as any other. This report does not propose why this may be so. Although this distribution of setting types is representative of POE work at the site scale, a much greater number of POE studies address other kinds of environments, such as building systems, interior spaces and planned unit developments. In other words, landscape architectural site designs appear to be evaluated far less frequently than other settings, as also reported by Friedman et al. (1978).

Just over half of the reviewed studies evaluated multiple sites. This is not viewed as significant, although only three of these were completed in less than two weeks. Note that over one-quarter of the studies reviewed were successfully completed in less than two weeks, although only one of these was done since 1978. This trend toward longer studies is perhaps a reflection of a coinciding concern for greater validity and generalizability, and generally, better results that might be achieved by greater amounts of data gathered over a greater range of conditions. In the most recent section of the table (since Van Valkenburgh, 1978) most (80%) studies of length greater than two weeks involved some degree of methodological assessment.

Important in justifying this study were the purposes of the 27 evaluations reviewed. Ninety-three percent were at least partly concerned with an issue relevant to a design implication; either to refine or redesign existing conditions, to make recommendations or suggest design guidelines for future work, to understand behavior in relationship to a design element[s] or simply to evaluate the "success" of the design.

Only one of the studies was purposed solely as an investigation or assessment of evaluation methods. A total of 26% were at least partly concerned with some degree of methodological insight. Although a fair number (22%) reported a brief critical description of their methods' performance, fewer (19%) went as far as to objectively order their effectiveness, and even fewer (15% or four studies) conducted quantitative tests, if only checks of reliabilty and/or validity. This lack of reflective research testifies to the relative nascency of the field. Also, there was little evidence in any of the studies that methods were chosen in response to reliability or validity given to methods previously tested in similar settings with similar research purposes.

As reported earlier, however, a recent trend is for greater concern for more effective use of methods, as further evidenced by the following:

1. Of the reviewed studies conducted since 1978, 63% were intended partly for methodological insight, as were only 11% before 1978.

2. Of the reviewed studies conducted since 1978, 80% reported at least a brief critical description of their methods' use, while only 24% conducted before 1978 did the same. Moreover, 60% of the studies conducted since 1978 reported no less than a detailed objective evaluation of methods used, while only 12% of those conducted before 1978 did the same. Also, three of the four studies that used quantitative tests of reliability and/or validity were conducted since 1983.

Anderson (1986) supports this trend by declaring that although "few studies examine the effectiveness of methods in POE ... this is a recent issue of concern to those in environmnetal design research and application, as illustrated by the number of EDRA 16 [(1986)] sessions devoted to some aspect of the method."

Conclusion

The ultimate purpose of this review has been to define a problem with regard to post occupancy evaluation of site-scale public space. The problem is well synopsized with the observation that

> ... the [methods] used in environmental evaluations are ahistorical and site specific. They rely little on formats which have been used and tested previously. As a result, these [methods] rarely meet basic psychometric standards for reliabilty and validity. Thus, reliability (replicabilty of results) and validity (meaningfulness of information) is rarely known to the researchers.[6]

Landscape architecture needs a methodologically sound basis for testing designs in use. Landscape architects who do use tests to improve design rely mainly on tacitly acknowledged criteria, having a weak tradition of explicitly shared tests. They make empirically testable assertions and then are unsure what methods are best to test them. They need assurance they are taking advantage of the best possible evaluation methods to produce the most useful results.

With such needs in mind, it serves well to restate the major objectives of the present study:

---

[6] Richard E. Wener, "Standardization of Testing in Environmental Evaluations" in Proceedings, *EDRA 13*. (College Park, Maryland: n.p., 1982), p. 78.

- To develop a POE research design for an existing urban space, addressing a relevant, site-specific research question, utilizing two statistically comparable types of a single POE research method;

- In applying this research design, to conduct independent tests of reliability and/or validity upon each type, providing a statistical basis to compare the effectiveness of each type in gathering behavioral information;

- To also compare necessary resources and ease of data analysis for each type in order to determine the general appropriateness of the types of method under study.

II. Research Site

Selection

The listed criteria were used to select a site for this study.

The site shall:

- be a site-scale public space, similar to those in the studies reviewed as a basis for this work, with a clear spatial definition, but of urban character;
- be well used, suitable for the collection of adequate, but manageable data;
- have been completed (built) at least five years ago and then regarded as outstanding in its solution, preferably an ASLA award winner;
- have been designed by an accessible design team interested in a particular research question addressing the site.

Based upon such criteria the Harborplace on the Inner Harbor in downtown Baltimore was selected for the study. The site design has received numerous awards, including a 1981 ASLA Honor Award for Urban Design. Landscape architects for the project were Wallace, Roberts and Todd (WRT) of Philadelphia.

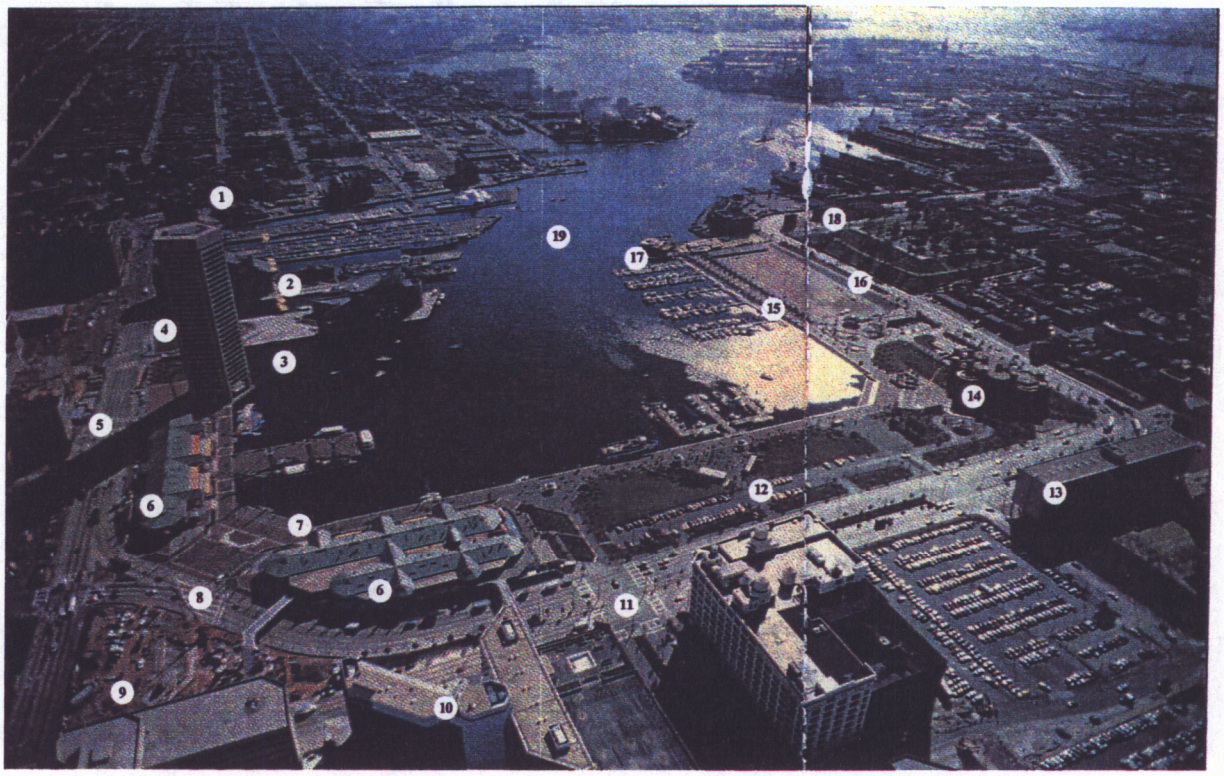


Figure 2. Site Context: View across Inner Harbor toward southeast shows Chesapeake Bay at top. Central business district is off lower left corner. Going counter-clockwise from 11 o'clock: 1) Little Italy; 2) [Fun Factory]; 3) Aquarium; 4) World Trade Center; 5) Pratt Street; 6) Harborplace showing pavilions; 7) "Constellation"; 8) "the Crotch"; 9) McKeldin Plaza; 10) Hyatt Regency Hotel; 11) Light Street; 12) Sam Smith Park; 13) Christ Lutherin Church Apartments; 14) Md. Science Center; 15) Joseph Rash Memorial Field and Public Marina; 16) Federal Hill; 17) Rusty Scupper restaurant; 18) Bethlehem Steel shipyards; and 19) site of unbuilt highway bridge. Source: Clay, G., "The Roving Eye: On Baltimore's Inner Harbor," *Landscape Architecture Magazine*, November 1982, pp. 48-49.

Description

The intersection of Pratt (5) and Light Streets (11) is referred to by Baltimoreans as "the Crotch" (8) and marks the main entrance to Harborplace, the "centerpiece of Baltimore's downtown revitalization."[7] Harborplace acts as a focal point for all activity in the Inner Harbor area. ("Inner Harbor" has come to mean the entire district around the water.) "Harborplace" consists of an L-shaped public plaza and promenade about 1500 feet long, framed by a pair of large (75,000 sq. ft.) two-story pavilions sheltering 144 shops and eating places (6).

Harborplace was opened in 1980 by The Rouse Company. Since then, literally millions of people have visited the site to shop, walk the promenade and participate in the festive atmosphere and events which are now an integral and easily accessible part of Baltimore's central business district. Only two blocks uphill is Charles Center, 42 acres of new skyscrapers, dozens of older towers, offices, a civic center and a new public library that draws thousands of daily city-goers.

Visual and walking access to the harbor has been carefully planned and managed. Nearly two miles of the Inner Harbor are now accessible to, and owned by, the public. Points of interest on the Inner Harbor include the 30-story World Trade Center (4), the National Aquarium (3), historic Federal Hill (16), the Fun Factory (2), the Maryland Science Center (14) and other parks and urban spaces. Most attention lately has been focused particularly on Harborplace - the observation site for this research consists of a large portion of its exterior space.

---

[7] Jury, 1981 ASLA Awards, "The 1981 ASLA Awards' New Meanings," *Landscape Architecture Magazine*, September 1981, p. 600.

Figure 3. Study Area: Jury, 1981 ASLA Awards, "The 1981 ASLA Awards' New Meanings," *Landscape Architecture Magazine*, September 1981, p. 601.

Most of the site is paved with brick, edged and subdivided by concrete bands. Furnishings include various seating elements (low walls, benches, and movable chairs), tables, custom lighting, informational kiosks and flagpoles. A promenade forms the lower-most level adjacent to the harbor; it is both simple and large enough to accomodate the constantly large crowds. A central plaza visually links the city and draws such crowds to the harbor. A small but suitably scaled ampitheatre sits in the center of the plaza. Here the city's convention and tourist bureau stages shows and exhibits. Impromptu "sidewalk" performers are also welcome. In either case, it is not rare for large crowds to overflow the ampitheatre steps.

Above the ampitheatre and promenade level and surrounding the pavilions are elevated areas, also paved in brick, with split granite paving denoting major entries to the buildings. This level also features lighting, seating, steps and ramps, with concrete band accents. Sep-

arating and defining these two levels is a series of planted areas which provide relief from the paving and act partly as a low seatwall.

Many site design details have also been utilized on the adjacent boulevards and open spaces to unify and establish the identity of the project as part of the overall downtown plan. The landscape design is intended to "orient all users of the space to downtown and coordinate a complex visual environment ..., clarify[ing] a kaleidoscope of visual elements, making Harborplace highly successful [and meaningful for the pedestrian]."[8]

[8] Ibid.

III. Research Design

Evaluation Focus

Even with the major goals of this study to assess evaluation methods, a POE addressing relevant, site-specific research questions is necessary in order to effect such an assessment.

Based partly upon the large numbers of pedestrians that use the site, and discussions with the Harborplace project director (of Wallace, Roberts and Todd), the following questions were formulated to guide the POE research design:

1. *Where* are the most and least used seating/gathering areas?

2. *Where* are the centers of activity?

Note that only *where* pedestrian activity occurs is important; when, why and which persons (type) are gathered or seated is beyond the scope of this study. But discerning locational patterns alone is instructive nevertheless. For instance, consider deJong's use of simple

density diagrams to discover a human attraction for edges in parks.[9] The discovery not only raised the generalized prospects of edge gravitation but eventually influenced design thinking.

Other, less renowned but valuable revelations have been made through the use of density diagrams. Rutledge (1981) describes a plaza study in which a density diagram showed clustering of users along stairway wingwalls and handrails that bisected the steps. Originally overlooked, but later reexamined through slides, it was found that sitters gravitated to the edge in order to clear "pedestrian channels" through the middle of the steps, and that this "mid-channel" clearing repeated itself on a regular basis. The clearing effect was also found to take place by those gathered in other areas of the plaza. Important to plaza designers was their original concern that the site layout would not "serve its through-traffic function well with all the milling about that was also expected to take place. The density diagram proved [them] wrong."[10]

---

[9] "Applied Hodology," in Albert J. Rutledge, *A Visual Approach To Park Design* (New York: Garland STPM Press, 1981), p. 146.

[10] Rutledge, p. 146.

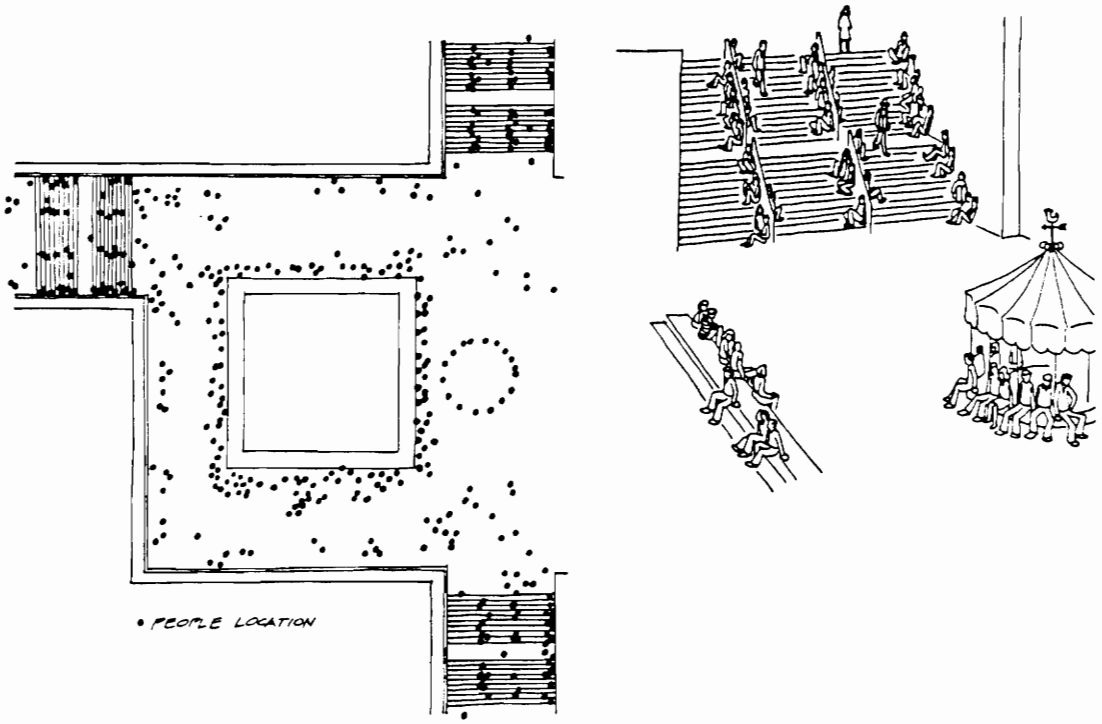


Figure 4. Density Diagram: "A density diagram may reveal unique patterns (left) and prompt a second look to determine their cause." Source: Rutledge, p. 146.

NPS (1982) notes examples how "activity mapping" can be used in evaluating park design. For instance, an activity map (density diagram) may show that a picnic area next to a playground is always heavily used, while another picnic area is seldomly used. Such information can then be used to make decisions about the placement and design of these and future facilities. NPS also describes how data from activity mapping concerning what parts of the park are used or not used, at what time, and what activity types occur can be used to develop a maintenance schedule to identify where and when maintenance is needed.

In an evaluation of the First National Bank Plaza in Chicago Rutledge (1975) took 35mm slides at 15-minute intervals to map densities of people. He found edges of steps and planter walls to be used heavily for seating, but found the overall site design to be "essentially a 'number of aisles' with no 'subspaces' to accomodate ... [basic] activities [such as relaxing]."[11] Such

[11] Friedmann, p. 151.

interpretations, together with those derived from other sources, including interviews, could be used to refine or redesign the space.

In a 1976 study of an outdoor space at the University of Massachusetts, pedestrian density maps were used by a landscape architect to develop a redesign for that space. Here, areas of constriction and congestion were identified. The problem area was subsequently attributed partly to lack of visual access into the space, caused by large, low-branching trees and elevation changes. The problem was addressed by raising the existing tree canopy and "simplifying and clarifying" the ground plane paving to provide a better sense of direction and flow.

These examples illustrate several ways an understanding of where activity takes place in an environment can be used to improve the quality of that or other environments. Admittedly there are more categories of information that are usually necessary to fully understand environmental behavior, including what exactly the people are doing, who or what type of people are involved, and with whom they are involved, when these activites are occuring, and perhaps most importantly, why people are doing what they are doing.

This study, however, is not aimed at a full understanding of activity at Harborplace but instead will focus on one category of information in order to most readily compare the effectiveness of methods chosen to gather such data.

Instrumentation

It was found appropriate to use observation methods for this study based upon the nature of the evaluation focus (research questions) to be addressed. Lozar (1979) identified two distinct types of observation methods - instrumented and direct observations. In instrumented ob-

servation the observer uses some mechanical support devices such as timelapse film, still photography or videotape to acquire data about behavior in a setting. Direct observation, on the other hand, is unaided by mechanical means but includes a variety of techniques, including proxemic and personal space methods, time-sampling and mapping.

With further regard for the research questions addressed, and also the physical conditions and potential use numbers for the setting chosen, this study utilized 1) time-lapse photography and 2) direct mapping as comparative method types for the research design. In order to most directly compare such measurement techniques, and to minimize variable influences, both types were conducted during the same days and hours, on the same territory.

Data Collection

In order for the POE research design to be useful for designers with minimal time attempting similar studies, data were collected for only three days - a consecutive Friday, Saturday and Sunday in October of 1986. Weather was pleasant; temperatures averaged in the low sixties, skies were partly to mostly clear. No holidays or large-scale events concurred with these dates. As such, these data are viewed as typical or representative of ordinary fall days in 1986.

Data were collected each day from 10:20 A.M. to 4:20 P.M. Eighteen observations were made each day, at twenty minute intervals. For each observation three observers simultaneously mapped positions of persons either seated or in stationary positions. Landscape architecture students were chosen as observers. This would allow assessment of the usefulness of the method for designers with little experience in the collection of behavioral data. The observers were positioned on a second story terrace overlooking the site (figure 5). No correspondence

was allowed between observers during observations although a brief group training session was given prior to actual data collection to ensure a generally consistent approach. A simple dot was used to denote each person (see Appendix D). Moving (walking) persons were not included; circulation studies are more appropriately conducted by other techniques such as videotape.

Time-lapse photographs were made simultaneously with direct observations, on the same territory. A camera was stationed on an historic ship permanently docked in the harbor (the USS Constellation), overlooking the research site (figure 2, #7). A 17 mm wide angle lens was used to frame the entire study area in a single shot - the camera position was maintained undisturbed on the ship's deck for the three days of data collection.

Figure 5. Camera View

As the intent was for the camera data to be collected simultaneously with those of direct observations in order to compare them, a simple method of sychronization was used. At the instant that each photograph was made, a handwave was given to those making direct observations by the camera attendent. (The two were in a direct sight line of one another, elevated above the pedestrian crowd.) Observers were then each allowed an undetermined

amount of time to make and record their observations. No instruction was given as to a particular area of the plaza to begin with in making each observation. Observations were, however, directed at progressively larger areas of the plaza, with size increased each of the three days (see Appendices A-C).

Data Retrieval and Input

After data collection the entire study area was spatially subdivided into fairly homogeneous areas in order to provide a framework to compare data collected by each of the two observation types and to determine where the most and least used areas of the plaza occurred. The resolution of the subdivisioning was determined subjectively, but with concern for what would be the smallest area or degree of spatial accuracy usable by a researcher interested in translating such data into design decisions, for a site-scale urban space. For example, it was asked whether it was sufficient to say that x number of persons gathered or sat in the area of one ampitheatre step, or better to subdivide each step into smaller, human-sized sections. Partly in hope of acquiring sufficient data (persons) in each area for computer analysis, a resolution more like the the earlier was chosen. Each subdivision was coded alpha-numerically.

Data retrieval was fairly simple for direct observations. An acetate spatial subdivisioning plan was used to overlay each raw data sheet (see Appendix E). Dots (persons) were then counted as they occurred in each area, for each observation. Similarly, for photo data, a large perspective drawing of the same spatial subdivisioning was constructed by overlay on a slide projection of the study area. Because each photograph was made from exactly the same vantage, with undisturbed camera settings, it was then possible to project slides upon the

subdivisioning perspective to retrieve data. Persons were then counted as they occurred in each subdivision, as was done with direct observations.

Both sets of data were organized, coded and input into a computer file under five column headings, for each observation.

```
4   2   18   LL    12
4   2   18   LL1   03
4   2   18   MM    04
4   2   18   MM1   05
4   2   18   NN    00
4   2   18   NN1   03
1   3   01   A     00
1   3   01   B     00
1   3   01   C     01
1   3   01   D     00
1   3   01   E     01
1   3   01   F     00
1   3   01   F1    00
1   3   01   G     01
1   3   01   G1    00
1   3   01   H     00
1   3   01   H1    00
1   3   01   I     00
1   3   01   J     00
1   3   01   J1    01
1   3   01   K     00
1   3   01   L     00
1   3   01   L1    00
```

Figure 6. Coded Data File

Column one represents the *observer* number (1-4): 1-3 are the human observers; 4 is the camera. Column two represents the observation day or *period* (1-3): 1 is Friday; 2, Saturday; and 3, Sunday. Column three represents the observation *time* for each day (1-18): 1, the first observation, was made at 10:20 A.M. and 18, the last observation of each day, was made at 4:20 P.M. Observations were made at twenty minute intervals. Column four represents the *location* (subdivision) where counts were ascribed (A-NN1 (56 total)), for each observation. The last column represents the number of persons counted, from either mapped or photo data, for each observation, at each location. A total of 20,875 persons were observed and input for analysis.

IV. Method's Assessment

Reliability and Validity

The quality of any measurement method is evaluated with respect to to two technical considerations: *reliability* and *validity*. Carmines and Zeller (1979) describe these as the most valuable qualities of any measurement method or instrument. In order to assess either quality, one first must understand what they mean. Carmines and Zeller (1979) propose the following interpretation:

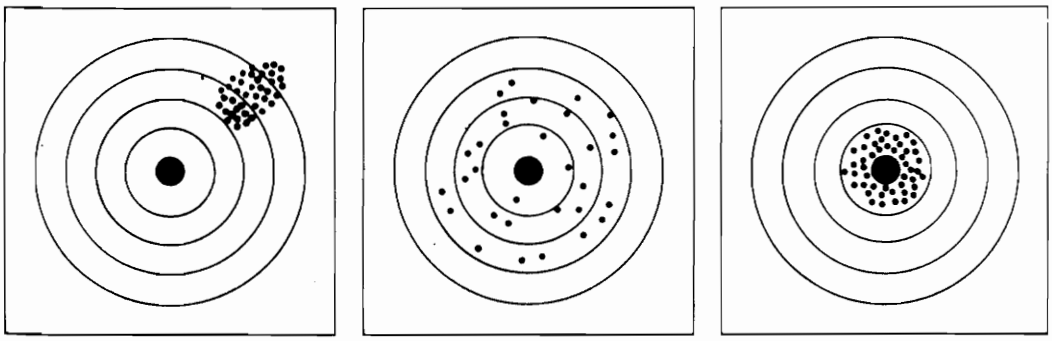
> At the most general level, there are two basic properties of empirical measurements. First, one can examine the reliability of an indicator. Fundamentally, *reliability* concerns the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials. The measurement of any phenomenon always concerns a certain amount of [random] error. ["Random error is the term used to describe all those chance factors that confound the measurement of any phenomenon."] The goal of error-free measurement - while laudable - is never attained in any area of scientific investigation. Instead, the amount of [random] error may be large or small, but it is universally present to some extent. Two sets of measurements of the same features of the same individuals will never exactly duplicate each other. Because repeated measurements never *exactly* equal one another, *unreliability* is always present to at least a limited extent. But while repeated measurements of the same phenomenon never precisely duplicate each other, they do tend to be consistent from measurement to measurement. The person with the highest blood pressure on a first reading, for example, will tend to be among those with the highest reading on a second examination given the next day. And the same will be true among the entire group of patients whose blood pressure is being recorded. Their readings will not be exactly the same from one measurement to another but they will tend to be consistent. This tendency toward consistency found in repeated measurements of the same phenomenon is referred to as *reliability*. The more consistent the results given by repeated

measurements, the higher the reliability of the measuring procedure; conversely, the less consistent the results, the lower the reliability.

But an indicator must be more than reliable if it is to provide an accurate representation of some abstract concept. It must also be valid. In a very general sense, any measuring device is valid if it does what it is intended to do. An indicator of some abstract concept is valid to the extent that it measures what it purports to measure. For example, the California F Scale (Adorno et al., 1950) is considered a valid measure to authoritarian beliefs to the degree that it does measure this theoretical concept rather than reflecting some other phenomenon. Thus, while reliability focuses on a particular property of empirical indicators - the extent to which they provide consistent results across repeated measurements - validity concerns the crucial relationship between concept and indicator. This is another way of saying that there are almost always theoretical claims being made when one assesses the the validity of social science measures. Indeed, strictly speaking, one does not assess the validity of an indicator but rather the use to which it is being put. For example, an intelligence test may be valid for assessing the native intellectual potential of students, but it would not necessarily be valid for other purposes, such as forecasting their level of income during adulthood (Nunnally 1978).

Just as reliability is a matter of degreee, also is validity. Thus, the objective of attaining a perfectly valid indicator - one that represents the intended, and only the intended, concept - is unachievable. Instead, validity is a matter of degree, not an all-or-none property. Moreover, just because an indicator is quite reliable, this does not mean that it is also relatively valid.

For example, Babbie (1986) includes a target shooting analogy to illustrate the "tension" between reliability and validity.



Reliable but not valid | Valid but not reliable | Valid *and* reliable

Figure 7. An Analogy to Reliability and Validity: Earl Babbie, *The Practice of Social Research*, 4th ed. (Belmont, Calif.: Wadsworth Publishing Co., 1980), p. 113.

Suppose the shots were fired from a well-anchored rifle. Notice that the shots don't necessarily need to be concentrated about the target to be considered reliable; they only need to be consistently placed. "A highly reliable indicator ... is one that leads to consistent results on repeated measurements [(shots)] because it does not fluctuate greatly due to random er-

ror."[12] Conversely, unreliability or presence of random error is indicated (middle) by shots that were as likely to hit above the target as below it or as likely to hit to the right of the target as to its left.

Nonrandom error is the second type of error that affects empirical measurement; it is related to validity. Unlike random error, nonrandom error has a systematic biasing effect on measuring instruments. Notice in the invalid illustration (left) that the shots aimed at the bull's eye hit approximately the same location but not the bull's eye - some form of nonrandom error affected the targeting of the rifle. Similarly, a thermometer that always indicates a temperature 5 degrees lower than it should is evidencing nonrandom measurement error and thus is performing invalidly. If the observer, however, is aware of the bias, then he can "compensate" accordingly.

Reliability Assessment

Although "research in the area of environmental design evaluation has been limited by the lack of [both] tested reliability and validity of research instruments,"[13] reliability assessment has been chosen as the more appropriate focus of this study, as explained below.[14]

The nature of the evaluation focus and instrumentation inherently promote a substantial degree of "face validity."[15] In other words, the simplicity of the research design provides a

---

[12] Edward A. Carmines and A. Richard Zeller, *Reliability and Validity Assessment* (Beverly Hills: Sage Publications, 1979), p. 13.

[13] Wener, p. 77.

[14] This is not to renounce the claim that for any measurement procedure to be scientifically useful, it must lead to results that are both reliable and valid.

[15] Babbie, p. 112.

clearly logical basis for determining that the instrument is measuring what it is supposed to. Observation (mapping) of persons seated or gathered is hardly contestable as not accurately representing where the most and least used seating/gathering areas occur. It can easily be contended, however, that such indicators possibly do not provide a valid picture beyond the scope of the data collected. Many factors can contribute to such invalidity - weather, time of year, current downtown activities, etc. These factors or "nonrandom error[s] lie at the very heart of validity [assessment]."[16] They do, however, contribute to a wider scoped study than is reported here.

Hence, *reliability assessment*, or the extent to which indicators (observations) contain random error is the focus of this study. In this case, random errors can include observer fatigue, inattention or data overload. The *type* of random error is of little importance to this study; only the *amount* will be determined. "The amount of random error is inversely related to the degree of reliability in a measuring instrument."[17]

Statistical Tests

Two tests were used to assess the reliability of the gathered data: the chi-square test of significance (or test of independence), and the analysis of variance (ANOVA) statistic for row mean score differences. The general purpose and use of each statistic will be described briefly; their particular applications in the testing of this study's hypotheses are described in Results.

---

[16] Carmines and Zeller, p. 14.

[17] Ibid., p. 13.

The Chi-Square Test of Significance

Parametric tests of significance are part of a body of inferential statistics usually used to decide whether samples randomly drawn from a population accurately reflect that population or are simply products of sampling error - i.e, an unrepresentative sample. The statistical significance of a relationship observed in a set of sample data is expressed in terms of probabilities. Significant at the .05 level ($p < .05$) means that the probability of a relationship as strong as the observed being attributable to sampling error alone is no more than 5 in 100.

With *chi-square*,[18] levels of significance are derived from a model that assumes statistical independence. That is, there is no association between the variables in the population. Then, given the *observed* distribution of values on the two or more separate variables, the marginal totals are tabulated and under the assumption that the two or more variables are independent an *expected* frequency is calculated. The result is a set of expected frequencies for all cells in the contingency table. The expected frequencies are then compared with the observed frequencies in order to determine the probability that a possible discrepancy could have resulted from sampling error alone.

In this study, the chi-square probability value was adapted for use as a measure of *consistency*, or lack of discrepancy, between the observations made directly by the observers and between observers and camera. An example, as the statistic was used in this study, will illustrate the procedure.

---

[18] Earl Babbie, *The Practice of Social Research*, 4th ed. (Belmont, Calif.: Wadsworth Publishing Co., 1980), pp. 416-425 passim.

SAS

PERIOD=1

TABLE OF OBSERVER BY LOC

OBSERVER	LOC							
FREQUENCY PERCENT ROW PCT COL PCT	AA	BB	CC	DD	E	R	S	TOTAL
1	9 1.43 5.92 40.91	15 2.39 9.87 24.19	25 3.98 16.45 28.09	2 0.32 1.32 28.57	3 0.48 1.97 12.00	4 0.64 2.63 26.67	8 1.27 5.26 25.00	152 24.20
2	3 0.48 1.97 13.64	16 2.55 10.53 25.81	19 3.03 12.50 21.35	1 0.16 0.66 14.29	7 1.11 4.61 28.00	4 0.64 2.63 26.67	11 1.75 7.24 34.38	152 24.20
3	5 0.80 3.14 22.73	14 2.23 8.81 22.58	25 3.98 15.72 28.09	3 0.48 1.89 42.86	7 1.11 4.40 28.00	5 0.80 3.14 33.33	8 1.27 5.03 25.00	159 25.32
4	5 0.80 3.03 22.73	17 2.71 10.30 27.42	20 3.18 12.12 22.47	1 0.16 0.61 14.29	8 1.27 4.85 32.00	2 0.32 1.21 13.33	5 0.80 3.03 15.63	165 26.27
TOTAL	22 3.50	62 9.87	89 14.17	7 1.11	25 3.98	15 2.39	32 5.10	628 100.00

Table 2. Data to Illustrate Chi-Square Procedure

Given the above observed data, the expected frequencies for each cell are first computed. This is done by using the marginal (row and column) totals. For example, in the cell described by the intersection of observer 1 and location AA, the observed frequency is 9 and the column total for all of "variable" AA is 22. The expected frequency for the cell is computed by multiplying 22 by row one's (observer one) total percentage of all frequencies (24.20). The product (expected frequency) is about 5.3. To compare the observed frequency with the expected frequency, chi-square is computed, for each cell by the formula: (observed — expected)^2÷expected. Hence, for this cell, the chi-square value is 2.58.

But the real value of the chi-square test, as in most all statistics, is as a summary of all the data. Consequently, this procedure is carried out for each cell in the table and all the results are added together. The final sum is the value of chi-square; 30.929 in the table above. This value is the overall discrepancy between the observed distribution in the sample and the distribution that should be expected if the variables were unrelated to one another. Normally, the magnitude of the chi-square value is used to estimate the probability that the variables are indeed related, as the discrepancy represents a countering of the original assumption of

independent variables. In other words, the higher the chi-square value, the more probable that the value could be attributed to sampling error alone.

In this study, higher chi-square values[19] are interpreted as not only more probably attributable to sampling error, but more importantly, less probably attributable to nonsampling or random error. Random errors may include erroneous recordings by observers caused by fatigue, overload, etc. Although the extent to which observations contain random error is inversely related to the degree of reliability or *consistency* in the measurement, the degree of consistency in the table is only referential to the expected, or marginal table totals. In other words, the chi-square statistic tells if observations are made in a *consistent* fashion to one another, with respect to the expected values. For instance, if observer one records x persons in a location, how accurately can it be predicted that observer two will record y persons for that same location?

It is important to note, however, that the chi-square tests does not state the probability that observers one and two will record the *same* number of persons in their observations (uniformity). For this test, the chi-square statistic is supplemented with the ANOVA statisic for row mean differences.

The ANOVA Statistic for Row Mean Differences

Analysis of Variance, abbreviated ANOVA, applies the same basic logic of statistical significance as in the Chi Square Test. The level of significance of an observed association is reported in the form of the probability that that association could have been produced merely by sampling error. To say that an association is significant at the .05 level is to say that an

---

[19] Note that chi-square values must be translated into probability values by use of standard chi-square tables and the computation of degrees of freedom.

association as large as the observed one could not be expected to result from sampling error more than five times out of 100.

ANOVA[20] is based simply on comparing variations between and within groups. Fundamentally, the cases under study (observations) are combined into groups representing an independent variable, and the extent to which the groups differ from one another is analyzed in terms of a dependent variable.

In this study, the mean number of observations, from observer to observer, or observer to instrument, was used as the dependent variable. Uniformity, or the degree to which observation totals are the same, was measured with ANOVA through the use of variance. In other words, the variance of a distribution, of observation totals, represents the extent to which the observations are clustered close to the mean or range very high and low from it.

General Use of Statistical Tests

To reiterate: In assessing reliability, the chi-square and ANOVA tests will be used conjointly, but with the respective purposes of:

- Measuring *consistency* from observer to observer, or observer to instrument, i.e.- the degree to which one can predict individual observations based on proportional relationships of total daily observations;

- measuring *uniformity* from observer to observer, or observer to instrument, i.e.- the degree to which daily observation totals are the same.

---

[20] Babbie, pp. 440-442 passim.

With the particular use of such tests it is not enough simply to say that observations are being made either consistently or uniformly. For instance, observers may be making observations proportionally consistent to one another, but not necessarily with the same number of observations. Contrarily, observers may make the same or similar number of total observations, but only by coincidence, i.e.- by means of inconsistent observations from observer to observer that happen to add to the same daily totals.

Consequently, both tests will be applied to the data in order to assess their reliability. Although more specifically described with their accompanying hypotheses, the tests will proceed generally as follows, for each day:

1. Conduct "interobserver" reliability tests for direct observation data to determine reliability from observer to observer.

2. Conduct "alternate forms"[21] reliability tests (compares results of two different techniques of the same type of tool where one technique considered reliable is tested against the other) for direct observation data versus time-lapse film data (deemed reliable) to assess reliability from observer to camera.

Research Question

In addition to tests of reliability, the following question will be addressed:

At what point, either in terms of increasing size of space observed or frequency of pedestrian

---

[21] Zeisel, p. 79.

activity, is reliability forfeited, for observations directed at persons either seated or in stationary position?

In other words, how large a space, or how many people, in stationary or in seated position can an untrained observer[s] accurately map?

Hypotheses

In *User Analysis: An Approach to Park Planning and Management*, the National Park Service (1982) describes methods used to examine park use in terms of the situations in which they are applicable, the type of information that can be obtained, and their limitations. Data collection and analysis techniques are also included, but more importantly is a discussion of reliability, for each method reviewed.

In support of the suitability of the methodological comparison made in the present study, the NPS proposes that some of the same techniques done using direct observation, such as counting, can also be done using a camera. In describing reliability for each of these two measurement techniques (counting or locational mapping by both direct observation and photo) the following points are made:

1. "Counting [by direct observation] is a reliable method of data collection because accuracy does not depend on observers judgements. [However], interobserver reliability check[s] should be made ... until all the observers agree 75 to 80 percent of the time.

2. The nature of the film medium is that it allows an indisputable record to be made of the use occurring in a given space and is thus a very reliable technique."[22]

Such premises suggest a good amount of reliability or absence of random error in the use of each method. In fact, interchangeable use of either method to gather the same data is proposed. Partly contrary to such beliefs, the present author proposes the following hypotheses for the present study. Each includes a null ($H_{(o)}$) and alternative ($H_{(a)}$) hypothesis, preceded by a statement of what is to be tested. Each test will be conducted for each of the three days. Each major hypothesis is subdivided to test both consistency and uniformity, each of these in repect to *location* and *time*. Location refers to the spatial subdivisions used to count and compare data. Time refers to total persons observed at each observation interval; there were 18 each day.

Hypothesis 1 (Interobserver Reliability)

Hypothesis $1a_{(loc)}$ consistency - observer to observer by location (spatial subdivision)

Test Question: Are the three observers' data *consistent* for each *period* (day) in terms of the numbers of persons observed for each *location* (spatial subdivision)?

$H_{(o)}$: The three observers gave consistent results in terms of the numbers of persons observed and their locations, for each period (day); (chi-square $p > 0.5$).

$H_{(a)}$: The three observers gave inconsistent results in terms of the numbers of persons observed and their locations, for each period (day); (chi-square $p < 0.5$).

---

[22] National Park Service, *User Analysis: An Approach to Park Planning and Management* (Washington, D.C.: ASLA, 1982), p. 47.

Hypothesis $1b_{(loc)}$ uniformity - observer to observer by location (spatial subdivision)

Test Question: Are the three observers' data *uniform* for each *period* (day) in terms of the numbers of persons observed for each *location* (spatial subdivision)?

$H_{(o)}$: The three observers gave uniform results in terms of the numbers of persons observed and their locations, for each period (day); (ANOVA $p > 0.5$).

$H_{(a)}$: The three observers gave non-uniform results in terms of the numbers of persons observed and their locations, for each period (day); (ANOVA $p < 0.5$).

Hypothesis $1a_{(tme)}$ consistency - observer to observer by time (total observed at each interval)

Test Question: Are the three observers' data *consistent* for each *period*(day) in terms of the numbers of persons observed for each *time* interval?

$H_{(o)}$: The three observers gave consistent results in terms of the numbers of persons observed for each time interval, for each period (day); (chi-square $p > 0.5$).

$H_{(a)}$: The three observers gave inconsistent results in terms of the numbers of of persons observed for each time interval, for each period (day); (chi-square $p < 0.5$).

Hypothesis $1b_{(tme)}$ uniformity - observer to observer by time (total observed at each interval)

Test Question: Are the three observers' data *uniform* for each *period* (day) in terms of the numbers of persons observed for each *time* interval?

$H_{(o)}$: The three observers gave uniform results in terms of the numbers of persons observed for each time interval, for each period (day); (ANOVA $p > 0.5$).

$H_{(a)}$: The three observers gave non-uniform results in tems of the numbers of persons observed for each time interval, for each period (day); (ANOVA $p < 0.5$).

Hypothesis 2 (Alternate Forms Reliability)

Hypothesis $2a_{(loc)}$ consistency - observers to camera by location (spatial subdivision)

Test Question: Are the observers' and camera data *consistent* for each *period* (day) in terms of the numbers of persons observed for each *location* (spatial subdivision)?

$H_{(o)}$: The observers and camera gave consistent results in terms of the numbers of persons observed and their locations, for each period (day); (chi-square $p > 0.5$).

$H_{(a)}$: The observers and camera gave inconsistent results in terms of the numbers of persons observed and their locations, for each period (day); (chi-square $p < 0.5$).

Hypothesis $2b_{(loc)}$ uniformity - observers to camera by location (spatial subdivision)

Test Question: Are the observers' and camera data *uniform* for each *period* (day) in terms of the numbers of persons observed for each *location* (spatial subdivision)?

$H_{(o)}$: The observers and camera gave uniform results in terms of the numbers of persons observed and their locations, for each period (day); (ANOVA $p > 0.5$).

$H_{(a)}$: The observers and camera gave non-uniform results in terms of the numbers of persons observed and their locations, for each period (day); (ANOVA $p < 0.5$).

Hypothesis $2a_{(tme)}$ consistency - observers to camera by time (total observed at each interval)

Test Question: Are the observers' and camera data *consistent* for each *period* (day) in terms of the numbers of persons observed for each *time* interval?

$H_{(o)}$: The observers and camera gave consistent results in terms of the numbers of persons observed for each time interval, for each period (day); (chi-square $p > 0.5$).

$H_{(a)}$: The observers and camera gave inconsistent results in terms of the numbers of of persons observed for each time interval, for each period (day); (chi-square $p < 0.5$).

Hypothesis $2b_{(tme)}$ uniformity - observers to camera by time (total observed at each interval)

Test Question: Are the observers' and camera data *uniform* for each *period* (day) in terms of the numbers of persons observed for each *time* interval?

$H_{(o)}$: The observers and camera gave uniform results in terms of the numbers of persons observed for each time interval, for each period (day); (ANOVA $p > 0.5$).

$H_{(a)}$: The observers and camera gave non-uniform results in terms of the numbers of persons observed for each time interval, for each period (day); (ANOVA $p < 0.5$).

V. Results

Introduction

Data were collected over three days (*periods*), with *observations* made over 18 *time* intervals each day by four *observers*, with observer four being the camera. Data were collected each day over an increasingly larger area: day one covered approximately 2,800 square feet or .06 acre; day two, 28,800 or .66 acre; and day three, 43,100 or .98 acre. Correspondingly, observations made increased each day: 628 total observations were made on day one; 8,174 on day two; and 12,073 on day three. An *observation* is the siting and recording of a single person.

Data were analyzed/compared compositely by *location* and *time*, from observer to observer. In other words, total observations (persons seated or stationary) were summed for each spatial subdivision (*location*), and for each time interval, for each observer, for the course of each day. It would not be enough to assess reliability based on only location or time. That is to say, reliability based on daily total observations made by each observer, for each location, disregards the time interval that they were made, and vice versa. Together, though, they form a complete assessment.

Table 3. Exemplary Data Totals

PERIOD=1

TABLE OF OBSERVER BY LOC

OBSERVER / LOC FREQUENCY PERCENT ROW PCT COL PCT	AA	BB	CC	DD	EE	FF	GG	TOTAL
1	9 1.43 5.92 40.91	15 2.39 9.87 24.19	25 3.98 16.45 28.09	2 0.32 1.32 28.57	3 0.48 1.97 12.00	4 0.64 2.63 26.67	8 1.27 5.26 25.00	152 24.20
2	3 0.48 1.97 13.64	16 2.55 10.53 25.81	19 3.03 12.50 21.35	1 0.16 0.66 14.29	7 1.11 4.61 28.00	4 0.64 2.63 26.67	11 1.75 7.24 34.38	152 24.20
3	5 0.80 3.14 22.73	14 2.23 8.81 22.58	25 3.98 15.72 28.09	3 0.48 1.89 42.86	7 1.11 4.40 28.00	5 0.80 3.14 33.33	8 1.27 5.03 25.00	159 25.32
4	5 0.80 3.03 22.73	17 2.71 10.30 27.42	20 3.18 12.12 22.47	1 0.16 0.61 14.29	8 1.27 4.85 32.00	2 0.32 1.21 13.33	5 0.80 3.03 15.63	165 26.27
TOTAL	22 3.50	62 9.87	89 14.17	7 1.11	25 3.98	15 2.39	32 5.10	628 100.00

Table 3 shows observations made on day one, by each of of the four observers, for locations AA-S (x axis). *Frequency* represents the number of persons observed for the entire day, by the corresponding observer. In descending order, percentages are given for each cell's frequency, as part of 1) the overall total for the entire day, by all observers; 2) the row total, or number of observations made by the corresponding observer for all locations, for the entire day; and 3) the column total, or number of observations made by all observers, for the corresponding location, for the entire day. Exemplary data totals by *time* are not illustrated.

As results will be reported in summary form, it is only important here to note that data were analyzed and compared as daily totals for each location and time, for each observer. Such totals (observations) were necessary in order to validly conduct the statistical tests. A need for adequate total frequencies (observations) also made it necessary, in several cases, to combine adjacent spatial subdivisions (locations) where a location had too few observations for valid statistical analysis. Appendices E and F display these changes.

Results of data analysis will be presented in reference to each hypothesis addressed in this study. For each of the two major hypotheses ("interobserver" and "alternate forms" reliability) the following format will be used:

1. restatement of test questions

2. presentation of results for each sub-hypothesis, for each day

    degrees of freedom (d.f.)

    chi-square or ANOVA value

    probability value (p)

    decision as to maintenance or rejection of null hypothesis (dec.)

3. interpretations of sub-hypotheses

4. summary/conclusion regarding major hypothesis

Recall that hypotheses' tests results are based on the following:

if $p < 0.5$ then results are SIGNIFICANT and INCONSISTENT or NON-UNIFORM

if $p > 0.5$ then results are INDEPENDENT and CONSISTENT or UNIFORM

A response to the Research Question will follow the hypotheses' tests results. A summary of other findings, including a qualitative assessment of the techniques used and a response to research questions posed by site designers concerning the site, will conclude Results.

Hypothesis 1 (Interobserver Reliability)

Reliability by location

Test Questions

Test Question $1a_{(loc)}$: Are the three observers' data *consistent* for each *period* (day) in terms of the numbers of persons observed for each *location* (spatial subdivision)?

Test Question $1b_{(loc)}$: Are the three observers' data *uniform* for each *period* (day) in terms of the numbers of persons observed for each *location* (spatial subdivision)?

Table 4. Test Results of Interobserver Reliability by Location (Spatial Subdivision)

	consistency				uniformity			
	hypothesis $1a_{(loc)}$				hypothesis $1b_{(loc)}$			
	d.f.	chi-sq.	p	dec.	d.f.	ANOVA	p	dec.
day 1	26	20.576	0.763	maint.	2	1.264	0.531	maint.
day 2	72	111.318	0.002	reject	2	3.305	0.192	reject
day 3	100	317.767	0.000	reject	2	14.432	0.001	reject

Interpretations

Observers gave consistent and uniform, i.e.- reliable results by location, only on day one; on days two and three they were neither consistent nor uniform. In other words, location of individuals observed is independent of observer on day one; on days two and three there are

significant discrepancies between observations made from observer to observer, although day two is closer to having uniform results ($p=0.192$) than day three ($p=0.001$).

Reliability by time

Test Questions

Test Question $1a_{(tme)}$: Are the three observers' data *consistent* for each *period* (day) in terms of the numbers of persons observed for each *time* interval?

Test Question $1b_{(tme)}$: Are the three observers' data *uniform* for each *period* (day) in terms of the numbers of persons observed for each *time* interval?

Table 5. Test Results of Interobserver Reliability by Time (Total Observed at Each Interval)

	consistency				uniformity			
	hypothesis $1a_{(tme)}$				hypothesis $1b_{(tme)}$			
	d.f.	chi-sq.	p	dec.	d.f.	ANOVA	p	dec.
day 1	28	34.743	0.177	reject	14	0.272	1.000	maint.
day 2	34	30.345	0.647	maint.	17	16.147	0.513	maint.
day 3	34	113.0	0.000	reject	17	7.232	0.007	reject

Interpretations

Observers gave consistent and uniform, i.e.- reliable results by time, only on day two; on day three they were neither consistent nor uniform, and on day one they were only uniform. In other words, time of individuals observed is independent of observer on day two; on days one

and three there are significant discrepancies between observations made from observer to observer with respect to total daily observations by time.

Summary/Conclusion

In assessing "interobserver" reliability for direct observations made by the three observers, the data give enough evidence to reject null hypotheses of reliability for all three days. Tests of reliability were based on consistency and uniformity of numerical data, in reference to where observers mapped/recorded individuals (*location*) and during which observation (*time*) they were mapped. Data must be *both* consistent and uniform, with respect to *both* location and time, to be considered reliable.

In examining chi-square and ANOVA probability values as conjoint indicators of reliability, for each of the three days, it is apparent that probabilities of reliability (independence) generally decrease for each successive day, although quite sharply. In other words, as the size of the space and level of pedestrian activity increases, reliability probability decreases, as the tables below comparatively illustrate by p values. An investigation into the more specific determination of where reliability is lost is addressed under Research Question of this section.

Table 6. Interobserver Reliability by Location

	p values	
	consistency	uniformity
day 1	0.763	0.531
day 2	0.002	0.191
day 3	0.000	0.001

Table 7. Interobserver Reliability by Time

	p values	
	consistency	uniformity
day 1	0.177	1.000
day 2	0.647	0.513
day 3	0.000	0.007

As a supplement to results of significance testing, simple measures of interobserver consistency were calculated as the percentage of similarity from the low to high observation. For example: 4, 5 and 6 would be found 80% consistent by dividing 4 into 6. The percentages represent average (mean) interobserver consistencies for each day, by time and location.

Table 8. Simple Measures of Interobserver Consistency

	location	time	mean
day 1	61%	97%	79%
day 2	69%	80%	75%
day 3	55%	76%	65%

Here, interobserver "reliability" (>75% consistency) was not achieved for any day by location. By time, reliability levels were found generally acceptable although they can only be considered conjointly with location - data from no day was found consistent by both location and time by simple measures of consistency. Also, decreasing levels of "reliability" were found for location and time mean consistency rates. Consequently, these findings support statistical test results of unreliable interobserver data for all three days, with generally decreasing levels of reliability as the size of the space and level of pedestrian activity increases.

Hypothesis 2 (Alternate Forms Reliability)

Reliability by location

Test Questions

Test Question $2a_{(loc)}$: Are the observers' and camera data *consistent* for each *period* (day) in terms of the numbers of persons observed for each *location* (spatial subdivision)?

Test Question $2b_{(loc)}$: Are the observers' and camera data *uniform* for each *period* (day) in terms of the numbers of persons observed for each *location* (spatial subdivision)?

Table 9. Test Results of Alternate Forms Reliability by Location (Spatial Subdivision)

	consistency				uniformity			
	hypothesis $2a_{(loc)}$				hypothesis $2b_{(loc)}$			
	d.f.	chi-sq.	p	dec.	d.f.	ANOVA	p	dec.
day 1	39	30.929	0.818	maint.	3	2.353	0.502	maint.
day 2	108	393.456	0.000	reject	3	3.668	0.300	reject
day 3	150	573.602	0.000	reject	3	22.184	0.000	reject

Interpretations

Observers and camera gave consistent and uniform, i.e.- reliable results, by location, only on day one; on days two and three they were neither consistent nor uniform. In other words, location of individuals is independent of observer and camera on day one; on days two and three there are significant discrepancies between observations made from observers and the

camera, although day two is closer to having uniform results ($p=0.3$) than than day three ($p=0.0$).

Reliability by time

Test Questions

Test Question $2a_{(tme)}$: Are the observers' and camera data *consistent* for each *period* (day) in terms of the numbers of persons observed for each *time* interval?

Test Question $2b_{(tme)}$: Are the observers' and camera data *uniform* for each *period* (day) in terms of the numbers of persons observed for each *time* interval?

Table 10. Test Results of Alternate Forms Reliability by Time (Total Observed at Each Interval)

	consistency				uniformity			
	hypothesis $2a_{(tme)}$				hypothesis $2b_{(tme)}$			
	d.f.	chi-sq.	p	dec.	d.f.	ANOVA	p	dec.
day 1	42	42.817	0.436	reject	14	5.82	0.971	maint.
day 2	51	57.971	0.234	reject	17	28.622	0.038	reject
day 3	51	276.627	0.000	reject	17	105.537	0.000	reject

Interpretations

Observers and camera gave consistent and uniform, i.e.- reliable results, by time, on none of the three days. In other words, on each day there are significant discrepancies between observations made from observers and the camera, although days one and two are closer to having consistent results ($p=0.436, 0.234$) than day three ($p=0.0$).

Summary/Conclusion

In assessing "alternate forms" reliability for direct observations made by the three observers versus the camera, the data give enough evidence to reject null hypotheses of reliability for all three days. Once again, tests of reliability were based on consistency and uniformity of numerical data, in reference to both where observers mapped individuals (*location*) and during which observation (*time*) they were mapped. Data must be *both* consistent and uniform, with respect to *both* location and time to be considered reliable. Results concerning the testing of consistency and uniformity as measures of "alternate forms" reliability were the same as those found in testing "interobserver" reliability (see Table 13).

Also similar were chi-square and ANOVA probability values for tests of both sets of hypotheses, as illustrated in the table below. Here, in "alternate forms" reliabilty testing, probabilities of reliability (independence) decrease for each successive day, almost as sharply as from observer to observer.

Table 11. Comparison of Reliability "Coefficients" (p Values) for Interobserver and Alternate Forms Tests

	interobserver				observer vs. camera			
	location		time		location		time	
	consist.	uniform.	consist.	uniform.	consist.	uniform.	consist.	uniform.
day 1	0.763	0.531	0.177	1.000	0.818	0.502	0.436	0.971
day 2	0.002	0.192	0.647	0.513	0.000	0.300	0.234	0.038
day 3	0.000	0.001	0.000	0.007	0.000	0.000	0.000	0.000

As a supplement to results of significance testing, simple measures of observer to camera consistency were calculated as the percentage difference between the observers' mean observation and the camera observation. The percentages represent mean observer to camera consistencies for each day, by time and location. The mean of location and time is also given.

Table 12. Simple Measures of Observer to Camera Consistency

	location	time	mean
day 1	73%	79%	76%
day 2	66%	87%	77%
day 3	70%	79%	75%

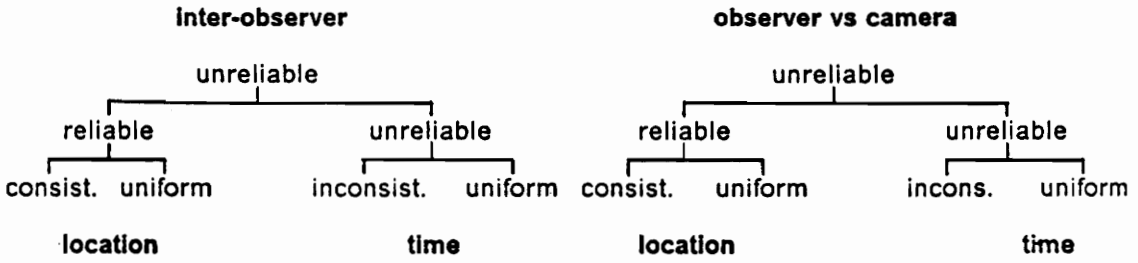
Here, in support of statistical test results, observer to camera "reliability" (>75% consistency) was not achieved for any day by location. By time "reliability" levels were found generally acceptable although they must be considered conjointly with location - data from no day was found consistent by both location and time by simple measures of consistency. Also, decreasing levels of reliability were not found for either time or location.

Consequently, these findings support statistical test results of unreliable observer to camera data for all three days although decreasing levels of reliability were not proven as the size of the space and level of pedestrian activity increased.

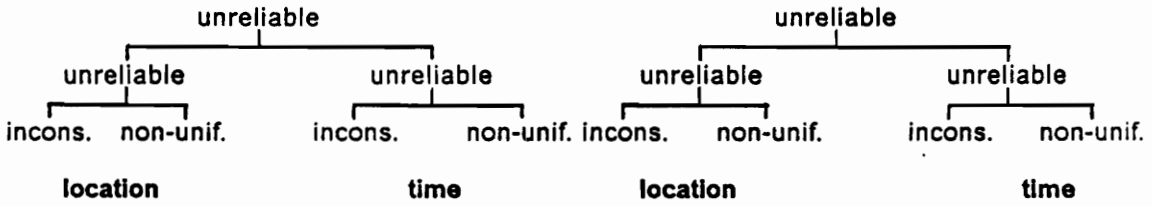
As simple measures of consistency serve mainly to supplement and cross-check statistical tests, the findings of the chi-square test of independence as a measure of consistency, and the ANOVA row mean score differences test as a measure of uniformity, are summarized below as the foundation of Results. Their practical significances are analyzed in Discussion of Results.

Table 13. Summary of Statistical Tests

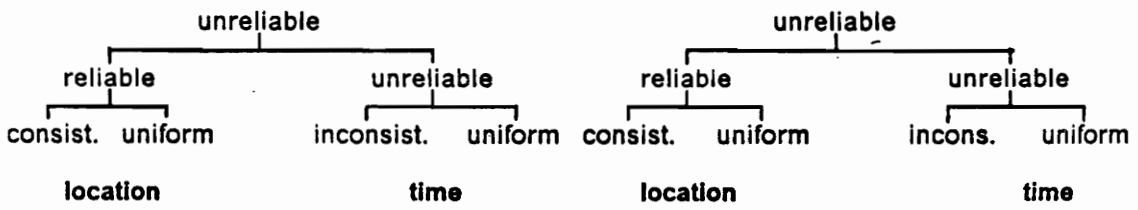
day 1



day 2



day 3



Response to Research Question

To restate the question: At what point, either in terms of increasing size of space observed or frequency of pedestrian activity, is reliability forfeited for observations directed at persons either seated or in stationary positions?

In other words, how large a space, or how many people, in stationary or in seated position, can an untrained observer[s] accurately map?

Although reliability was statistically determined to have been lacking for all three days of data collection, it is felt due much more to the increased frequency of pedestrian activity than the increased size of the space observed. This is based on the fact that when observing modest numbers of people, on either the smallest or largest observation area, similar levels of consistency were achieved. In other words, people can accurately map the locations of persons virtually as far as they can see (in a site-scale space), and it is the *number* of persons to be observed that becomes the more significant limiting factor in maintaining reliability. It is plausible to think, though, that mapping a large number of persons is less difficult in a smaller, more confined space where people are less dispersed. This is not viewed as a major issue at the site-scale.

An issue of critical importance in responding to this research question concerns the term *accuracy*, or validity. It has been noted that reliability does not necessitate accuracy. Accuracy refers to the truth, or what actually took place. In this study, camera data was considered the veridical picture of what observers attempted to record and was to be used as a basis to compare and determine the validity of observers' data. Due to unforseen limitations listed in Discussion (p. 71), such a preconception will be found dubious. Yet, in order to respond to this question, the assumption of accurate camera data is maintained. As such, these results should be viewed judiciously.

A less statistically elaborate approach was used to determine how many persons can be accurately mapped by crudely ordering a relationship between total numbers of persons mapped and consistency levels, for each time interval. For all three days (disregarding size of the observed space) each time interval (1-18 for each day) was first ranked in terms of the number of total observations made by all three observers. For each of these *times*, the difference from the low to high observation was calculated in percentage form and given as a

simple measure of consistency. In the table below the number of persons mapped is reported as an average for all three observers and given in ranges (e.g. 1-10, 11-30). Interobserver consistency rates were calculated as the percentage similarity from the low to high observation. The average number observed is then compared with the number "observed" by the camera and given also as a percentage of consistency.

Table 14. Ranked Frequencies (Numbers Observed) with Consistency Rates by Time

avg. persons mapped	inter-obs. consist.	obs.-camera consist.
1-10	96%	83%
11-30	84%	85%
31-50	84%	81%
51-75	81%	89%
76-100	80%	94%
101-150	80%	94%
*151-200	66%	76%
201-300	78%	83%
301-350	77%	84%
600-650	72%	54%

* Observations at only two time intervals totaled between 150 and 200 persons; at one of these time intervals an unusually inconsistent observation abrubtly affected consistency rates.

As measured simply by degree of numerical consistency, the data show a remarkable degree of congruence, with consistency levels generally not falling below 70%, even with the mapping of several hundred persons. The consistencies appear both in interobserver and observer versus camera data.

Although a trend for decreasing consistency was found as persons observed increased; this trend is only evident in interobserver consistency. For instance, for observations where between 10 and 100 persons were observed, an 82.5% average interobserver consistency rate was found. Where more than 100 and less than 600 average persons were observed/mapped,

there was found a 75% average consistency rate, which is still considered generally acceptable among POE researchers such as the NPS (1982).

For each time interval (persons observed) there were significant differences between inter-observer and observer versus camera consistency levels, averaging about nine points difference in either direction. There were, however, generally higher levels of consistency between observers' average and camera observations than between observers alone. In other words, inconsistencies between observers were balanced somewhat in relation to camera data by averaging.

At any rate, the acceptable consistency rates (>75%) derived in this table oppose the rejections of reliability arrived at by earlier tests in this study. However, for such results to have greater meaning, consistency rates for numbers of persons mapped would need to be computed not only by total persons mapped at each time interval, but also where, or at what locations these persons were ascribed. It is not enough to say that a consistent number of persons were mapped from observation to observation; they must also be mapped in consistent locations. In reference to this table, untrained observers are generally capable of accurately mapping the locations of up to and possibly more than several hundred people. But based upon the generally unreliable findings of statistical tests (see Table 13) and simple measures of consistency (see Tables 8 and 12) where reliability is considered by both time *and* location, it is conjectured that much of the data here would also be judged unreliable if considered by location also. This means that for observations made at each interval, measures of consistency would need to be made from location to location rather than as only totals for each location for each day. Such computations were not completed for this study. Consequently, an exact number of persons, in stationary or seated position that an untrained observer can accurately map was undetermined.

Summary of Other Findings

Qualitative Assessment of Techniques

Lozar (1974) describes direct observation (mapping) and time-lapse filming as equally suitable techniques for the collection of "overt" behavioral data, with either "social" or "physical" emphasis. NPS (1982) suggests that activity mapping or counting by direct observation can also utilize time-lapse film. The present study even recommends conjoint use of both types of observation methods to cross-check reliability of data collected simultaneously. Both techniques are "primarily used for gathering numerical data to test hypotheses concerning the relationships between activities and the physical setting in which they occur."[23] There are, however, several qualities that differentiate each and should be recognized in their use.

Zeisel (1984) presents several qualities of observing environmental behavior as a general research method: it is "empathetic and direct, deals with dynamic phenomena, and allows researchers to vary their intrusiveness in a research setting."[24] While Zeisel describes these qualities as they apply generally to obsservation, it is useful to briefly examine their applicability to the two types used in the present study.

Direct observation allows for an *empathetic* involvement with and feeling for the character of the situation under study, more so than by photographic means. Being on the spot allows researchers to understand nuances that users of the setting feel or even why people behave

---

[23] National Park Service, p. 29.

[24] Zeisel, p. 112.

as they do (or at least to think they understand; interpretations can differ). At the least, these "participant observed"[25] insights are useful for further study.

It is more difficult to project oneself into a place-situation via film, especially time-lapse or time-interval photography. By no means a *direct* replication, film destroys much of the sensory experience. In examining film frames it is not possible to overhear conversations, follow or zoom in on an individual or group, sense changes in weather or smell.

Although most telling, direct observation is matched by time-lapse photography in its ability to address certain research focuses, such as locational patterns in the present study. Here, there was little need for a more empathetic or direct means of data collection. With the questions phrased specifically enough, the camera can be set up to collect the needed information, provided an adequate sample is taken.

Sample size depends not only on length of data collection, but at what intervals photos are taken or observations are made. In a time-interval approach where the interval is greater, the *dynamism* of activity is potentially lost. Patterns and chains of events are less easily recognized and significant events can be missed all together. With time-lapse photography not supplemented by direct observation, "gaps" between photos can contain information that might not be overlooked by an observer who continually views the study area. While those making direct observations do so at periodic intervals, they can be trained to "look carefully at events, [to] continually question whether they see the whole event, whether they see all the participants, and whether something significant has been missed."[26]

When the focus of the observation is particularly on interaction *sequence*, neither time-lapse photography or time-intervaled direct observation is an appropriate method. In "sequence

---

[25] Zeisel (1984) defines a participant observer as ...

[26] Zeisel. p. 115.

sampling"[27] all behaviors under study are recorded in order of sequence. The sample continues until the interaction sequence ends and then begins with another sequence of interactions. When the focus is not on dynamic sequence but in the number or location of certain behavior[s] observed at regular intervals, both time-lapse photography and direct observation can provide the needed information.

Researchers should also consider the *intrusive* qualities of these observation types in order not to affect the observed the situation. Depending upon the setting and circumstances under study each can be performed without much notice (if desirable). In the present study participant observers posed as restaurant patrons on an outdoor terrace overlooking the study area. The camera was set up in an inconspicuous location aboard a ship in the harbor. Neither set-up seemed to cause people to react differently. Nevertheless, the choice of vantage point or level of intrusiveness can have a dramatic effect upon the way subjects act, as in the now classic Hawthorne experiment where those who knew they were being observed changed the way they acted (Roethlisberger and Dixon, 1939).

Not every decision concerning the use of a particular observation type will have a dramatic effect upon the study's outcome but qualities such as these considered (empathy, directness, dynamism and intrusiveness) can help to choose a most appropriate technique. In regard to these four qualities as they apply to the two observation types discussed, the present author supports Zeisel's statement that photographic observation "removes the observer from the scene of action, depriving the method of a large part of its research potential."[28]

---

[27] Altmann, J., "Observational Study of Behavior: Sampling Methods," *Behaviour* Vol. No. (49): page unknown.

[28] Zeisel. p. 116.

Necessary Resources and Ease of Data Analysis

The following discussion is not meant as a technical critique or procedural guide to the use time-lapse photography or direct observation mapping. Rather, as a reordering of notes taken and lessons learned during the process of conducting this "POE" points are offered only for consideration; they should be considered by anyone conducting a similar study. More importantly, this author highly recommends the consulting of a professional with experience in such work (e.g.- Project for Public Spaces, Inc. and Partners for Livable Places, Inc.) in order to ensure meaningful and reliable results. Observation research can be deceptively simple.

Equipment/Materials Needed: Those making direct observations need only detailed basemaps of the study area and a pencil and eraser (mismarkings are inevitable but observers should be encouraged to correct them). The basemap should include all features in the space recognizable by the observer, including pavement joints or scoring lines. The map should also designate a space for the observer's name, time period, weather conditions and other notes relevant to the observation period. Similar notes should be kept for time-lapse photos.

A zoom lens 8mm movie camera is suitable for time-lapse photography although a 35mm camera can be used in a time-sampling approach where photos are taken less frequently by a camera attendent such as in the present study. Here, the camera can be fitted with a wide-angle lens to enframe larger study areas. For time-lapse filming an intervalometer is necessary to set photo intervals and engage the shutter. Also, a projector with a variable film speed control is needed to view and analyze the film. A slide projector can be used for photos taken with a 35mm camera.

Staff/Training: One does not have to be an expert to observe behavior, but it must be done accurately. Training very much depends on the purpose and complexity of the evaluation although staff should be trained by an on-site consultant, the cost of which outweighs that of

collecting useless data. He/she should be present at least until "reliability checks" have proven successful (i.e.- data is judged reliable).

Some additional points to consider:

- Staff should be well-trained although prior experience is not necessary; it is essential that all observers record information correctly and consistently.

- Although the number of observers depends upon both the size of the space observed and the number of potential users, no less than three staff are necessary in order cross-check reliability and perhaps average data.

- Although possibly more expensive to send three or more employees into the field to collect data than a single camera operator, the time necessary to retrieve and summarize data from film adds to make direct observation more expedient.

- With direct observation, staff researchers must understand the purpose of the observation study (exactly who or what is being counted and why this information is needed) in order to know what to observe and to collect only the kind of information needed. For example, in the present study observers needed to be consistent in deciding what represented stationary activity. A clear definition of what is to be recorded must be given. With time-lapse work there is advantage in having a complete record of activity although decisions must still be made as to size of the space observed and consequent level of photographic detail.

Data Analysis: Comparatively speaking, with proper set-up and preparation, data is more easily *gathered* via time-lapse photography. "Work" involves basically only the setting up and reloading of the camera. However, in terms of data analysis conducted in the present study, it is far simpler to retrieve and consolidate data from direct observation mapping than to re-

trieve the same data from film. Transfering data from film to a data base can take 4-5 times that of tallying raw data from direct observation maps.

Once data are entered into computer data bases, their manipulation is essentially the same for both observation types in translating to summary tables and charts. Once data are in summary form, comparing them with original hypotheses and interpreting them are the same for each.

Some additional points to consider:

- Data analysis should be conducted in reference to original hypotheses, i.e.- in summary form (graphs, charts, etc.) that is most applicable to these hypotheses. The implications of some formats are often more evident than others when examining certain hypotheses. However, data should be examined in more than one format to ensure unbiased results.

- During data analysis, new hypotheses can be formulated. Data may then need to be re-evaluated and additional data should be gathered.

- Data analysis should fit into the overall context of the study; counts/tabulations are often meaningless when used alone.

Conclusion: The major limitation of both observation types is the amount of time it takes to organize the field work, form hypotheses, collect a representative sample of data (i.e.-during different times of day, week, year), and finally to analyze data and make conclusions.

While time-lapse work involves the training and actual staff time of only one person for data collection (as compared to three for direct observation), retrieval of data from time-lapse photos can take as much as four times that of direct observation (with the particular evaluation focus of this study).

In addition to time, other qualities that have been reviewed distinguish direct observation mapping as more appropriate for this type of evaluation:

- Direct observation allows for a more empathetic, dynamic and potentially less intrusive involvement with the situation under study.

- Direct observation requires less expensive equipment and materials.

- Direct observation does not necessarily require more training or skill development time than time-lapse photography.

- Ease of data analysis is not an advantage for time-lapse photography.

- Time-lapse photography has not been proven a more reliable source of data.

Response to Evaluation Focus

Harborplace site designers expressed interest in examining what areas of the central plaza (see Figure 3) were "most and least used [as] seating/gathering areas" and where the "centers of activity occurred." Such questions were kept on a general level so not to distract or detract attention and effort from the larger study's focus as an assessment of the observation methods themselves.

Nevertheless, the defining of an addressable research problem, answerable research questions, or testable hypotheses with anticipated results are required before before any methods of data collection or analyses are chosen. "It must be clear exactly what information is expected to [be] record[ed], in what form it will be recorded, and how it will contribute to the solution of the problem [(evaluation focus)]. It is extremely easy either to gather a large

amount of detailed information, interesting in itself, but irrelevant in making design decisions, or to gather information too general or too obvious to be of real value,"[29] such as may have been the case in the present study.

If the present study had been rigorously conducted as a veritable post occupancy evaluation aimed at using results in one way or another, questions would have been developed more specifically. For instance, the research problem might have been narrowly defined to test a hypothesis that the ampitheatre steps are used significantly less than other seating elements or areas, aside from during performances. A question might then have been asked: Within these other subareas, what percentage of people are eating, socializing or people watching? The research design could then be structured to address this and other potential questions and/or hypotheses.

At any rate, data collected via direct observation was consolidated in order to discern *where* pedestrian activity occurred. Moving persons were not included. Figure 8 represents the summary of user density over 1 day (observations at 18 time intervals) from 10:20 A.M. to 4:20 P.M. The map or density diagram is based on the observations of one of the three observers.

During this particular day (Sunday) a series of afternoon performances took place in the central ampitheatre. Aside from an occasional water taxi on the north side of the study area's harbor edge, no other boarding or deboarding of boats is represented. A dot indicates a person in stationary position long enough to be recognized by the observer.

---

[29] National Park Service, p. 44.

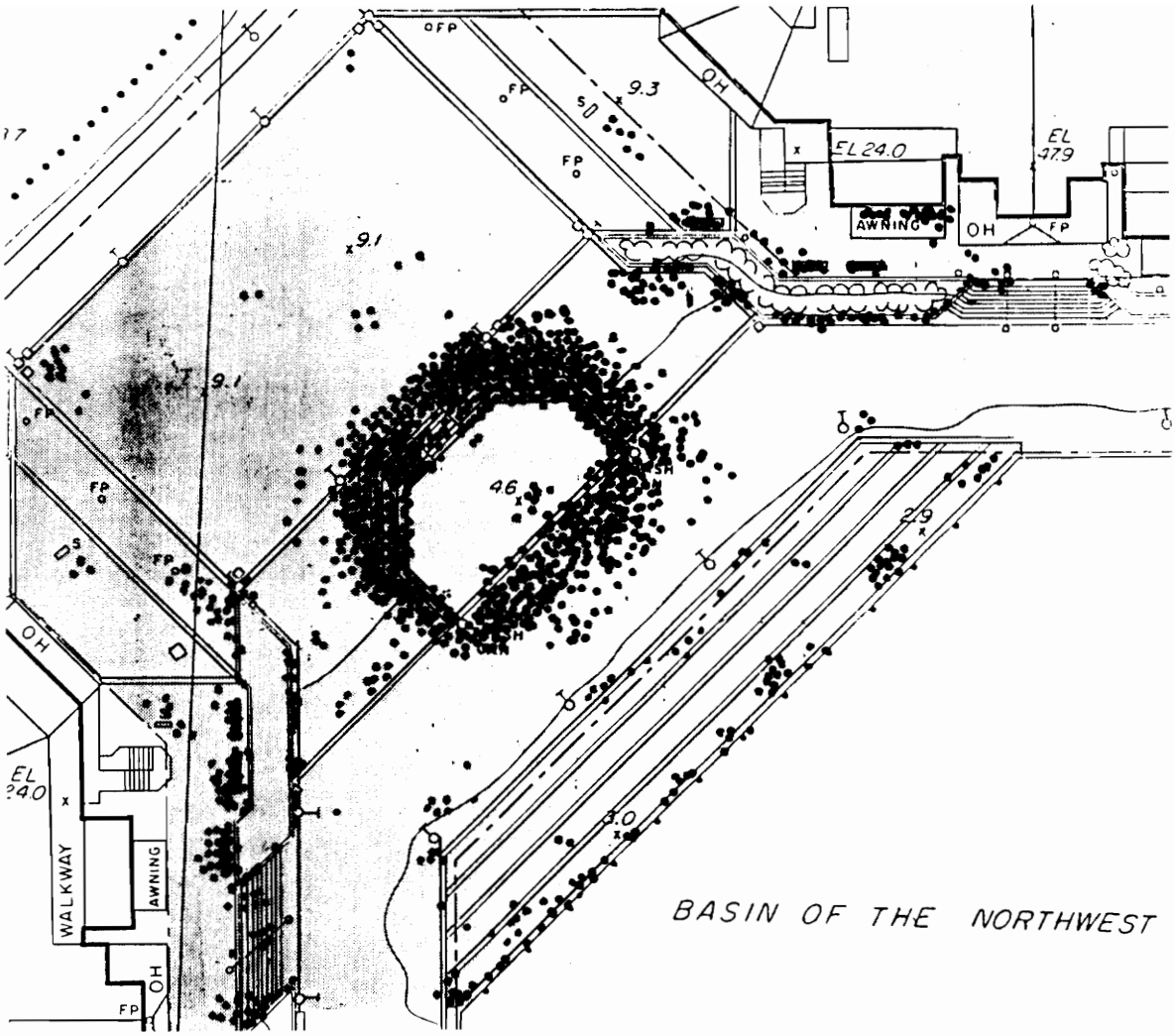


Figure 8. Study Area Density Diagram

The following were interpreted from density diagrams such as figure 8, percentages of use as they occurred in subdivided plan (see Appendix G, User Density), and insights and anecdotal information related by researchers in the field. The results are only representative of activity during a Friday, Saturday and Sunday of Fall, 1986. During these three days regular performances took place in the ampitheatre. No attempt will be made to analyze the results for design implications.

- There is very little attraction to the street forming the northwest edge of the site. The street side of the plaza, almost to the ampitheatre edge, receives little pedestrian use aside from as a means of traversing one way or another.

- The ampitheatre steps and an area several feet beyond its perimeter, including its harbor-side edge, are by far the most heavily used seating/gathering area. No particular ampitheatre steps were found to be used significantly more or less than any others, aside from the central "walking" steps leading down to the stage space, which are used less as seating elements. The ampitheatre area, however, only predominates as a center of activity during performances; crowds gather and then leave quickly as people retreat to less central fringe areas.

- Perimeters of the two major planting areas separating the upper and lower levels of the plaza receive relatively consistent use. People use the seatwall on ampitheatre-side of the planter although the upper, pavilion-side is used more heavily - it is lined with several benches and provides a better overlook. Of the two terrace levels the east side is used considerably more; the east pavilion is the major food source.

- During non-performance periods people are less likely to make their way out to the ampitheatre to eat, people-watch or relax, but remain in closer proximity to the pavilions. Attention is diverted to the promenade which once again becomes the center of activity, with the harbor as its backdrop.

- The ampitheatre's stage space is rarely intruded upon as a place to gather.

- The promenade is used minimally as a gathering space; pedestrians are reluctant to stop here, except in the vicinity of the ampitheatre's open end where it can get quite congested after a performance begins.

- The water's edge is not heavily used for seating; people are reluctant to sit right on the edge although individuals and small groups occasionally stand gathered close to it.

- Steps up to the pavilions' terrace levels are not used heavily for seating although people occasionally sit near their wingwalls.

VI. Discussion of Results

In order to discuss the meaning of hypotheses' tests results it is important to highlight how they were derived.

Tests of statistical significance are used to determine the probability that an observed relationship could have been produced by chance alone. Significant at the .05 level ($p < .05$) means that the probability of a relationship as strong as the observed one being attributable to chance is no more than 5 in 100. Conversely, p values greater than 5% suggest the likelihood that the relationship is due only to chance, and that variables are unrelated or statistically independent. (Independence is often assumed before testing.) In this study, tests of independence were adapted to represent consistency and uniformity (reliability), or the probability that that data differed from "variable" to "variable," i.e., observer to observer or observer to camera (see Statistical Tests, pp. 30-34).

Babbie (1986) reports that "although tests of significance provide an objective yardstick against which to estimate the significance of associations between variables ... the researcher should be wary of ... dangers in their interpretation ..."[30] The most important "danger" in re-

[30] Babbie, p. 424.

gard to this study is that *statistical significance* is easily misinterpreted as *substantive significance*, or the degree to which the "observed association is strong, important, meaningful or worth writing home to your mother about."[31] Babbie warns against the risk of basing the importance of findings only on a significant (unreliable) relationship at the .05 level.

Such a warning does not preclude the use of results from this study's tests of significance; it only means their importance must be examined legitimately and logically, especially in light of their purported significance to a designer/researcher.

Is it legitimate, for instance, to support rejections of reliability for all data collected on all three days? The answer depends very much on the *level* of reliability achieved for data from observer to observer or observer to camera, i.e.- the strength of their association. Such a measure is not provided by significance probabilities.

Simple measures of numerical consistency, however, generally supported (although sometimes contradicted) results of statistical significance. Possible explanations are provided for these results.

Chi-square and ANOVA tests gave evidence to reject hypotheses that observers gave consistent and uniform results between themselves and versus the camera, for each day (see Table 13). Recalling that reliability does not ensure accuracy and vice versa, these results can be explained by any one of the following:

1. The observer's data are inaccurate, but the camera data are accurate;

2. The observer's data are inaccurate, as are the camera data;

3. The observers' averaged (mean) data are accurate, but the camera data are not.

---

[31] Ibid., p. 426.

The individual observers' data cannot be accurate because they differ from observer to observer; only one can be accurate. Also, the observers' and camera data cannot both be accurate because they differ. But there is no way of knowing which, if any data represent what actually took place, especially because the film data had been misjudged as an "indisputable record" (NPS, 1982). The following limitations, as became evident during the collection of this study's data, support the potential inaccuracy of the camera data.

1. imperfect resolution or clarity of data slides, making persons in distance or on fringe slightly out of focus and difficult to interpret;

2. obstructions, by both masses of people and site elements, partially created by low camera vantage;

3. difficulty deciphering "stationary" from moving persons;

4. superimposition of spatial-subdivisioning perspective on slide of dubious alignment to plan data (direct observation data).

For instance, during heavy pedestrian use, when problems like obstructions (people) might have become more prevalent, camera totals were found significantly lower than those of observers, reflecting non-retrieved film data. This and other listed limitations are judged likely to have had significant impact on the accuracy of the camera data.

Although it is difficult to determine the relative accuracies of either direct or camera observed data, it has been found that there are similar levels of discrepancy from observer to observer and observer to camera. Since the camera is now considered a somewhat inaccurate source of data, it is less reasonable to use it as a base against which to measure the reliability of direct observations. But, nevertheless, interobserver reliability tests and consistency measures alone give sufficient evidence to disprove the reliability of data from observer to observer for the majority of the data collected.

Large amounts of data were indeed collected by each observer. The mean number of total persons observed for each time interval (18 were made each day) was about 105 for all three days; 10 for day one; 117 on day two; and 190 on day three.

It is questionable how data from day one could be collected unreliably, with such low pedestrian volume. In actuality, such a determination is partly due to misrepresentation of data. In order to explain the misrepresentation it is more legitimate to focus only on interobserver reliability since camera data is considered an inaccurate data source. Firstly, by simple measure of interobserver consistency (see Table 8) day one was determined only 61% consistent by location. This is explained as a fault in simple consistency measure. For instance, suppose observations were made of 2, 3 and 3 persons by the three observers. A consistency measure would find the percentage difference between the lowest and highest as the rate of consistency - in this example, only 66% consistent. Sixty-six percent does not seem like a fair description of consistency between these data - they are actually very similar. Such misrepresentation of data is typical for this type of measure at low frequencies, such as day one in this study.

Secondly, in looking at results of statistical tests of reliability for data from day one, they are in fact judged reliable by location (see Table 13). By time they are considered uniform but inconsistent. Uniformity represents the similarity of data by totals for each time interval, and inconsistency suggests differences like what was misrepresented by simple measures of consistency. Consequently, data from day one is probably more reliable than the tests show.

A large increase in the frequency of use for days two and three explains the general decreases in probabilities and levels of reliability in their data. In addition to data overload, random errors made by observers probably include carelessness and mismarking of observations out of boredom, fatigue or pure mistake, and the use of different observation "approaches." For instance, by beginning to record data at opposite "ends" of the study area data could easily be collected inconsistently from observer to observer (because people

move, etc.). The amount of random error attributable to particular causes is not of particular importance here; the implications of its presence by any means is discussed in Conclusions.

VII. Conclusions

Implications of Results

Although days two and three were ultimately judged as having unreliable data between observers, and between observers and camera, this finding must be put in context - this is a very heavily used urban space. (The Inner Harbor attracts more annual visitors than Disneyworld.)[32]

Data from this study were collected in mid-Fall when use numbers had substantially subsided from peak season. Nevertheless, pedestrian volume was generally large. Observers mapped, on average, 105 persons per observation. Recall that only persons seated or in stationary positions were recorded, but with hundreds, if not thousands of people moving about.

---

[32] Interview with Tim Korbelak, Harborplace Project Director, Wallace, Roberts and Todd, Philadelphia, Pennsylvania, 25 July 1987.

Whyte (1980) analyzed use in 18 New York city plazas, finding peak use (no moving people) to average less than 90 persons, but as a total for an entire lunchtime hour.[33] The present study's totals represent pedestrian volume at only a single instance. Moreover, Project for Public Spaces (1978) conducted studies of "prominent" open spaces in downtown Seattle. PPS examined plaza activity partly by those seated, standing or leaning, reporting such activity for the course of entire summer days (12 - 5 P.M.). In the PPS study, use averaged 120 and 160 people for each plaza. In the present study, with an hour less of data collection, use averaged 1828 persons for all three days, and 2655 for days two and three.

Consequently, unreliability found in data collected at Harborplace is exemplary of an extreme case, in terms of use. Nevertheless, the implications of such findings should be viewed as worthy of consideration in any site-scale space - these techniques were rigorously tested. Listed below are what was implied from results of these tests.

- During heavy use, or about two-thirds of the duration of data collection in this study, observers were overloaded with data and consequently made erroneous judgements rather than correct recordings of locations of persons.

- The camera has not proven itself as a more reliable, let alone accurate type of observation method, as compared to direct observation.

- With camera "observations," it is as much the retrieval of data that confounds reliability as it is its collection.

- Different measures/tests of reliability occasionally provide contradictory results; each has a weakness and can misrepresent actual data, but misrepresentations can usually be explained if examined logically.

---

[33] Whyte also reported, though, that one could map the location of every sitter in about five minutes. He illustrated a typical siting plan with 88 persons mapped, but did not report any notion of the data's reliability.

- Although reliability is measurable, it is difficult to pinpoint from individual observation to observation with use of summary statistics. Consequently, absolute determination of an exact number of persons that can be accurately mapped is unachievable.

- Reliability does decrease as pedestrian frequency increases but not so uniform a degree that can be used to predict levels or probabilities of reliability. Size of space has not yet been proven as a significant limiting factor in reliability achievement, although when compounded with greatly increasing use numbers it does make mapping more difficult (at the site scale).

- No type of observation method can be determined truly effective (valid and reliable) until another method proven accurate serves as a basis against which to compare it.

Recommendations

Effective Use of Methods

- Choice of methods used in environmental evaluations should be based on reliability or validity given to those techniques previously tested in similar settings with similar research purposes, also on methods' general appropriateness for specific situations.

- Any and every measurement technique should be thoroughly pretested, including data collection, retrieval and introductory analysis, to ensure effectiveness. Particularly, reliability "checks" should be made by simple measures of consistency, although their interpretation should be made legitimately and logically. An "agreement of 70-90% usually

serves as the minimum criterion to be achieved before observation begins, depending on the complexity of the observation scheme."[34]

- Just as use of multiple evaluation methods is necessary to cross-check and validate findings concerning the space, multiple tests of reliability and/or validity are essential to guaranteeing the effective use of the methods themselves; i.e.- to cross-check and compensate for weaknesses from test to test.

- Recording of data should reflect a comonly agreed upon interpretation of activity, not one idiosyncratic.

- When unreliability is found in direct observations of heavily used space the area can be subdivided for mapping, but each area should still be observed by multiple observers so to continue reliability checks.

- When unreliability is discerned in pretesting of direct observations, a skilled observer should examine individual observation approaches to detect biases or any form of random error present in order to remedy it. To do this the "inspector" should discuss with the observers how and why they decided to record information as they did and establish rules of thumb. For example, if all the observers but one agree what represents a stationary person, that observer should adjust to the consensus. Observers should also be encouraged to admit fatigue, boredom or overload. If inconsistencies persist, the method itself may need to be made more clear or unambiguous for those making observations in order to achieve consensus.

- Data should show some reliability across time, even though longitudinal studies are difficult to manage. Nevertheless, single day data collection, a week, month or even year

---

[34] Friedmann, p. 198.

after the initial study can be used to check reliability, although it must be realized that there is the potential for the situation itself to change.

- Direct observation and time-lapse photography can and should be used conjointly, partly with the intent of using camera data as an accurate basis against which to assess the reliability of direct observations, but with precaution taken to ensure the accuracy of camera data, as is listed.

  - Camera Set-up

    - Elevate at an angle where activity can be discerned, but obstructions are not a problem, i.e.- height should be such that fixed (trees, kiosks) and potential objects (people, vehicles) don't obscure anyone that might be directly behind; especially consider distant zones of study where camera "sees" at low angle.

    - Ensure clarity or focus of entire slide by overframing study area and shooting from as close as possible so "edges" of slide do not include data. If compromise between angle wide enough to frame site and level of detail are not suitable, use two cameras; when "getting in" close, maintain unobtrusiveness; be careful with wide angle distortion - generally, wide angle lenses need to be set perpendicular to ground plane, and distortion may not be as apparent through viewfinder as in developed slide or photo.

    - Include clock in picture, with date, to denote observation interval and day, visible in the lower part of the picture frame, close enough to the camera to be clearly legible, but far enough away to be in focus with the rest of the scene.

    - Pretest camera exactly how and where it is to be used, from set-up to retrieval and analysis of data from slide; ensure that all personss are clearly visible in reference to the observation purpose, e.g. - if it is to record activity type, make

sure it can be done; also, determine proper interval for photography, consider a short-enough one so not to miss major activity, but one long-enough so to make changing of film and camera attendence less frequent.

- ▲ Conduct pretesting of camera simultaneously to direct observations to ensure synchronization and examination of the same people. For direct observation mapping requiring longer than a half-minute, it may be necesary to take more than one photograph to concur with mapped data, but this creates extreme difficulty in retrieval and analysis of data.

- Retrieval and Analysis of Data

  - ▲ As evaluation of film data does not save time (it only stores it), be prepared for tedium, boredom and ambiguity.

  - ▲ Construct spatial subdivisioning plan, if one is to be used, in reference to existing lines/elements in the completed design (most reliable data will be collected here); make subdivisions no smaller than needed to address evaluation focus/research questions.

  - ▲ Know what is being looked for in data and develop criteria to recognize; be consistent; re-retrieve data from same photograph to check reliability.

Future Research

- A study similar to this would be useful to assess the reliability of findings, but with a more gradual increase in size of space observed and frequency of pedestrian activity in order to more clearly ascertain where reliability is forfeited.

- Although technical/statistical tests of reliability are illustrated in this study, simpler, less time-consuming techniques should be tested for use by designers with little time or experience with statistics.

- Development and testing of financially feasible methods such as direct observations techniques should take precedence to enable use by more design firms.

- Evaluation methods should be refined for use in the most simple, clear, straightforward and efficient manner possible so to facilitate effective generation and easy interpretation of results.

- Results of environmental evaluations should be presented (or made available) in such a way that can be tested in comparison with other researchers' results.

- Standards of reliability and validity, with accompanying tests or approaches to measuring them need to be developed in order to make it easier for researchers to "check" the reliability and validity of their findings.

- Schools of design should include more emphasis on assessing performance of the built environment in order to better qualify future designers/researchers to reliably and validly conduct such studies.

- Academically completed research should be better integrated into practice by voluntary cooperation with site-designers, including review and suggestions on future use of study approaches and findings.

Conclusion

To conclude, by quoting a source of inspiration for this work, Zeisel (1984) overviews a chapter on research quality:

> Useful research solves already-recognized problems and identifies new ones. Research methods that increase researchers' abilitiy to do this improve the quality of research. If you use methods that allow other people to criticize your research, you can improve your own. If you know when your research findings are applicable and when not, you can act on the world with greater control.[35]

He also states that a "high degree of ... reliability and validity among techniques excludes from scientific observations those that are idiosyncratic, murky, or inconsistent and reflect only one person's unique perspective."[36] This report described how one of these criteria for quality research can be tested for a method used commonly in post occupancy evaluation of exterior space. Hopefully, those reading this report and then conducting an evaluation of any environment will ask themselves, "Have I chosen, used or tested methods in such a way that results can be judged valid and reliable?", and "Have I presented my findings in such a way that they are useful to others attempting to choose and use methods effectively?"

---

[35] Zeisel, p. 86.

[36] Ibid., p. 78.

Selected Bibliography

Amadeo, D., et al. "User Images of Evaluations of a Small Community's Downtown Environment." unpublished paper presented at the *Eleventh Annual Environmental Design Research Association Conference,* Charleston, S.C., 1980.

Anderson, J.R., and D. Butterfield. "Post Occupancy Evaluation and Generalization." *The Challenge of Change: CELA 1981, Abstracts of Papers Presented at the Annual Meeting of the Council of Educators in Landscape Architecture*, p. 1. Department of Landscape Architecture: University of Washington, Seattle, 1981.

Babbie, E. *The Practice of Social Research.* Belmont, Calif.: Wadsworth Publishing Co., 1986.

Becker, F. "A Class Conscious Evaluation: Going Back to the Sacramento Mall." *Landscape Architecture* (October 1973).

Becker, H.S. "Problems of Influence and Proof in Participant Observation." *American Sociological Review*, 23 (1950); pp. 652-660.

Braybrooke, S. "Evaluating Evaluation" *Design and Environment* 5, No. 3 (Fall 1974): pp. 20-25.

Bravat, R.M. *Studying Behavior in Natural Settings.* New York: Holt Rinehart and Winston, 1972.

Brujn, H.S. *The Human Perspective in Sociology: The Methodology of Participant Observation.* Englewood Cliffs, N.J.: Prentice-Hall, 1966.

Cambell, D. "Evaluation of the Built Environment: Lessons From Program Evaluation." *The Behavioral Basis of Design, Book 1: Selected Papers*, pp. unknown. (eds. P. Suedfeld and J, Russell). Stroudsburg, Pa.: Dowden, Hutchinson and Ross, 1976.

Canter, D. "The Purposive Evaluation of Places, A Facet Approach." *Environment and Behavior.* 15, No. 6.

Carmines, C. and R.A. Zeller. *Reliability and Validity Assessment.* Beverly Hills: Sage Publications, 1981.

Clem, P., et al. "A Comparison of Interaction Patterns in an Open Space and a Fixed Plan School." Albuquerque, N.M.: Institute for Environmental Education, 1973. (Mimeographed.)

Collier, J., Jr. *Visual Anthropology: Photography as a Research Method.* New York: Holt, Rinehart and Winston, 1967.

Crank, K.H. "The Assessment of Places." *Advances in Psychological Assesssment* (ed. P. McReynolds). Palo Alto, Calif.: Science and Behavior Books, 1978.

Davis, G. and V. Ayers. "Photographic Recording of Environmental Behavior." *Behavioral Research Methods in Environmental Design.* (ed. W. Michelson). Stroudsburg, Pa.: Dowden, Hutchinson & Ross, 1975.

Davis, T.A. "Evaluating for Environmental Measures." *Proceedings of the Twelth Annual Environmental Design Research Association Conference,* pp. 45-55. College Park, Md., 1982.

Deasy, C.M. "People Watching with a Purpose." *AIA Journal* 54, No. 6 (December 1970): pp. 35-40.

_____. *Design for Human Affairs.* New York: John Wiley and Sons, 1974.

de Jong, D. "Applied Hodology." *Landscape* 17, No. 2 (1967): pp. 10-11.

Eckbo, G. "Evaluating the Evaluation." *Design and Environment.* 2, No. 4: pp. 39-40.

Feldman, E.J. *A Practical Guide to Field Research in the Social Sciences.* Westview, Calif.: n.p., 1981.

Francis, M. "Behavioral Approaches and Issues in Landscape Architectural Education and Practice." *Landscape Journal* 1, No. 2: pp. 92-95.

Friedman, A., C. Zimring, and F. Zube. *Environmental Design Evaluation.* New York: Plenum Press, 1978.

Gilfoil, D. and H. Bowen. "Behavioral Effects of Environmental Design Characteristics in a Public Plaza." unpublished paper presented at the *Eleventh Annual Environmental Design Research Association Conference,* Charleston, S.C., 1980.

Gutman, R., and B. Westergaard. "Building Evaluation, User Satisfaction and Design." Rutgers University, Built Environment Research Paper 17, date unknown.

Guttentag, M. "Models and Methods in Evaluation Research." *Journal of Theory and Social Behavior* 12, No. 4.

Hinkle, D.E., W. Wiersna, and S.G. Jurs. *Applied Statistics for the Behavioral Sciences.* Boston: Houghton Mifflin Company, 1979.

Humphreys, M. "A Post Occupancy Evaluation of Public and Semi-Public Spaces in a Facility for the Aged Using Systematic Observation." Master's thesis, Cornell University, 1982.

Knight, R.C. and P.E. Campbell. "Environmental Evaluation Research: Evaluators' Roles and Inherent Social Commitments." *Environment and Behavior* 12, No. 4.

Lofland, J. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis.* Belmont, Calif.: Wadsworth Publishing Co., 1971.

Lozar, C.C. "Measurement Techniques Toward a Measurement Technology." *Man-Environment Interactions: Evaluations and Applications, Part II.* (ed. D.C. Carson). Stroudsburg, Pa.: Dowden, Huttchinson & Ross, 1974.

Lyle, J. "People Watching in Parks." *Landscape Architecture* (October 1970).

Marans, R.W. and K.F. Spreckelmeyer. *Evaluating Built Environments: A Behavioral Approach.* Survey Research Center: University of Michigan, 1981.

McCall, C.J. "Data Quality Control in Participant Observation." *Issues in Participant Observation* (eds. G. McCall and J.L. Simmons). Reading, Mass.: Addison-Wesley, 1969.

Michelson, W. (ed.) *Behavioral Research Methods in Environmental Design.* Stroudsburg, Pa.: Dowden, Hutchinson & Ross, 1975.

Moore, G. and R.G. Golledge. *Environmental Knowing.* Stroudsburg, Pa.: Dowden, Hutchinson & Ross, 1975.

Moore, G.T. "Architectural Evaluation - The 1982 Progressive Architecture Award Winners." *Environment and Behavior* 14, No. 6 (November 1982): pp. 643-651.

National Park Service, U.S. Department of the Interior. *User Analyis: An Approach to Park Planning and Management.* Washington, D.C.: ASLA, 1982.

Patterson, A.H., and R. Passini. "The Evaluation of Physical Settings: To Measure Attitudes, Behavior, or Both?" *Proceedings of the Fifth Annual Environmental Design Research Association Conference*, pp. 211-219. Milwaukee, Wisconsin: n.p., 1974.

"Post Occupancy Evaluation: The State of the Art." *Research and Design* 1. No. 3 (July 1978).

Preiser, W.F.E. "Analysis of Pedestrain Velocity and Stationary Behavior in a Shopping Mall." Albuquerque, N.M.: Institute for Environmental Education, 1973. (Mimeographed.)

Pushkarev, B. and J.M. Zopin. *Urban Spaces for Pedestrians: An Quantitative Approach.* Cambridge, Mass.: The MIT Press, 1975.

Reizenstein, J.E. "Changing the Definition of Decision-Making as a Way of Increasing the Use of Post Occupancy Evaluation Findings." *Proceedings of the Tenth Annual Environmental Design Research Association Conference*, Buffalo, New York: n.p., 1979.

Rutledge, A.J. *A Visual Approach to Park Design.* New York: Garland STPM Press, 1981.

Sanoff, H. and G. Coates. "Behavioral Mapping: An Ecological Analysis of Activities in a Residential Setting." *International Journal of Environmental Studies* 2 (1971): pp. 227-235.

Schulberg, L. "Behavior Mapping for Design." *Design and Environment* 2, No. 1 (Spring 1971): pp. 34-35.

Stephens, M.A., C.N. Baker, and E.P. Willems. "Self-Observations and Reports of Behavior as a Method for Post Occupancy Evaluation." *Proceedings of the Twelfth Annual Environmental Design Research Association Conference*, pp. 323-330. Ames, Iowa: n.p., 1981.

Ventre, F.T., "Architectural Criticism: Conoisseurship or Evaluation?" College of Architecture and Urban Studies, Virginia Polytechnic Institute and State University: Blacksburg, Virginia, 1985. (Typewritten.)

Videch, A.J. and G. Shapior. "A Comparison of Participant Observation and Survey Data." *Issues in Participant Observation* (eds. G. McCall and J.L. Simmons). Reading, Mass.: Addison-Wesley, 1969.

Webb, E.J. *Unobtrusive Measures.* Chicago: Rand Mcnally, 1966.

Weick, K.E. "Systematic Observational Methods." *The Handbook of Social Psychology,* Vol. II, *Research Methods* (ed. H.M. Proshanky et al.). New York: Holt, Rinehart and Winston, 1970.

Weitzer, A.R. "Research on Environmental Images: The Perception and Use of Urban Parks." Ph.D., DAI, 41, No. 12-B, 4750.

Wener, R.E. "Standardization of Testing in Environmental Evaluations." *Proceedings of the Thirteenth Annual Environmental Design Research Association Conference,* pp. 77-84. College Park, Md.: n.p., 1982.

_____. "Environment-Behavior Research: Success Stories." paper presented at the *Symposium on Evaluation of Occupied Designed Environments*, Georgia Institute of Technology, 1982.

Whyte, W.H. *The Social Life of Small Urban Spaces.* Washington, D.C.: The Conservation Foundation, 1980.

Winkel, G.H. "The Challenge of the Case Study for the Environmental Design Researcher." *Proceedings of the Fourteenth Annual Environmental Design Research Association Conference*, pp. 59-64. Lincoln, Nebraska: n.p., 1983.

Wolf, M. "The Behavior of Pedestrians on 42nd Street, New York City." New York: Graduate Center, City University of New York, 1970. (Mimeographed.)

Yin, R.K. "The Case Study as a Serious Research Strategy." *Knowledge: Creation, Diffusion, Utilization* 3. No. 1 (September 1981): pp. 97-114.

Zeisel, J. *Inquiry by Design: Tools for Environment-Behavior Research.* Cambridge: Cambridge University Press, 1984.

_____. *Sociology and Architectural Design.* New York: Russell Sage Foundation, 1975.

Zimring, C.M. and J.E. Reizenstein. "Post Occupancy Evaluation: An Overview." *Environment and Behavior* 12: pp. 429-451.

_____. "A Primer on Post Occupancy Evaluation." *AIA Journal* (November): pp. unknown.

Zimring, C.M. and F. Wener. "Evaluating Evaluation." *Environment and Behavior* 17: pp. 97-117.

Zube, E.H. *A Multi-Factor Approach to Site Design Evaluation.* Amherst, Mass.: University of Massachusetts Institute for Man and Environment, 1974.

_____. *Environmental Evaluation: Perception and Public Policy.* Monterey: Brooks/Cole Publishing Company, date unknown.

Reviewed Case Studies

1. Love, R.L. "The Fountains of Urban Life." *Urban Life and Culture*, July 1973, pp. 161-209.

2. Reynolds, I., and C. Nicholson. "Housing Site Evaluation." rewritten in *Environmental Design Evaluation*, pp. 166-180. (authors Friedmann, A., C. Zimring, and E. Zube). New York: Plenum Press, 1978. (note: Friedmann et al. cite original information source: "The Estate Outside the Dwelling," 1973.)

3. Brower, S.N. "Recreational Uses of Space: An Inner City Case Study," *Proceedings of the Fifth Annual Environmental Design Research Association Conference*, pp. 53-166. Milwaukee: n.p., 1974.

4. Moore, R.C. "Meanings and Measures of Environmental Quality: Some Findings from Washington Environmental Yard." *Proceedings of the Ninth Annual Environmental Design Research Association Conference*, pp. 287-306. Tuscon: n.p., 1978.

5. Nager, A.R., and W.R. Wentworth. "Urban Park Evaluation." rewritten in *Environmental Design Evaluation*, pp. 155-165. (authors Friedmann, A., C. Zimring, and E. Zube). New York: Plenum Press, 1978. (note: Friedmann et al. cite original information source: "Bryant Park: A Comprehensive Evaluation of Its Image and Use With Implications for Open Space Design," 1976.)

6. Marcus, C.C., "Evaluation: A Tale of Two Spaces, Contrasting lives of a court and plaza in Minneapolis." *AIA Journal*, August 1978, pp. 34-39.

7. Palmer, J.E., and J.H. Crystal. "Evaluating the Accessibility of Designed Environments: National Park Visitor Centers." *Proceedings of the Tenth Annual Environmental Design Research Association Conference*, pp. 307-316. Buffalo: n.p., 1979.

8. Rutledge, A.J. "First National Bank Plaza." rewritten in *Environmental Design Evaluation*, pp. 142-155. (authors Friedmann, A., C. Zimring, and E. Zube). New York: Plenum Press, 1978. (note: Friedmann et al. cite original information source: "First National Bank Plaza, Chicago, Illinois, A Pilot Study in Post Construction Evaluation," 1975.)

9. Kueffer, W.C. "Behavior and Use Patterns in a Tuscon Park." *Proceedings of the 7th Annual Environmental Design Research Association Conference*, pp. 75-80. Vancouver: n.p., 1976.

10. Cohen, H., et al. "Evaluation of a Campus Space." rewritten in *Environmental Design Evaluation*, pp. 132-141. (authors Friedmann, A., C. Zimring, and E. Zube). New York: Plenum Press, 1978. (note: Friedmann et al. cite original information source: "Design Evaluation of a Central Outdoor Space at the University of Massachusetts," 1976.)

11. Share, L.B. "AP Giannini Plaza and Transamerica Park: Effects of Their Physical Characterisitics on Users' Perceptions and Experiences." *Proceedings of the Ninth Annual Environmental Design Research Association Conference*, pp. 127-139. Tuscon: n.p., 1978.

12. Preiser, W.F.E., K.P. Rohane, and M.P. Eshelman. *An Evaluation of Outdoor Space Use at the University of New Mexico.* Albuquerque: Institute for Environmental Education, 1982.

13. Kantrowitz, M. and R. Nordhaus. "The Impact of Post-Occupancy Evaluation Research: A Case Study." *Environment and Behavior* 12, No. 4 (December 1980): pp. 508-519.

14. Burden, A. "Greenacre Park: A Study by Project for Public Spaces, Inc." New York, 1977. (Typewritten.)

15. Miles, R.C., R.S. Cook, and C.B. Roberts. "Plazas for People." New York: Project for Public Spaces, Inc., 1978. (Mimeographed.)

16. Allor, D.J., and R.K. Murphy, Jr. "Group Behavior in a Public Space: Fountain Square, Cincinnati, Ohio." 1979. (Typewritten.)

17. Dozio, M., P. Feddersen, K. Noschis. "Everyday Life on an Insignificant Public Square: Venice." *Ekistics* 298 (January/February 1983): pp. 66-76.

18. Van Valkenburg, M.R. "The Design Implications of Grade School Children's Use of and Attitudes About Two Play Areas in Carle Park, Urbana, Illinois." *Proceedings of the Ninth Annual Environmental Design Research Association Conference,* pp. 307-319. Tuscon: n.p., 1978.

19. Kaplan, R. "Citizen Participation in the Design and Evaluation of a Park." *Environment and Behavior* 12, No. 4 (December 1980): pp. 494-509.

20. Allor, D.J., and S.R. Howe. "The Times of Safety: A Case Study of Fountain Square in Downtown Cincinnati, Ohio." *Proceedings of the Twelth Annual Environmental Design Research Association Conference*, pp. unknown. Ames, Iowa: n.p., 1981.

21. Aguar, C.E. "Longitudinal Research: An Environmental Designer Looks Back." *Proceedings of the Thirteenth Annual Environmental Design Research Association Conference*, pp. 299-309. College Park, Md.: n.p., 1982.

22. Francis, M., and C. Girot. "Mapping Downtown Activity: A Research Approach and Some Design Applications." *Proceedings of the Fourteenth Annual Environmental Design Research Association Conference*, pp. unknown. Lincoln, Nebraska: n.p., 1982.

23. Martin, J., and J. O'Reilly. "Designing Zoos for Children: An Alternative Approach." *Proceedings of the Thirteenth Annual Environmental Design Research Association Conference*, pp. 339-346. College Park, Md.: n.p., 1982.

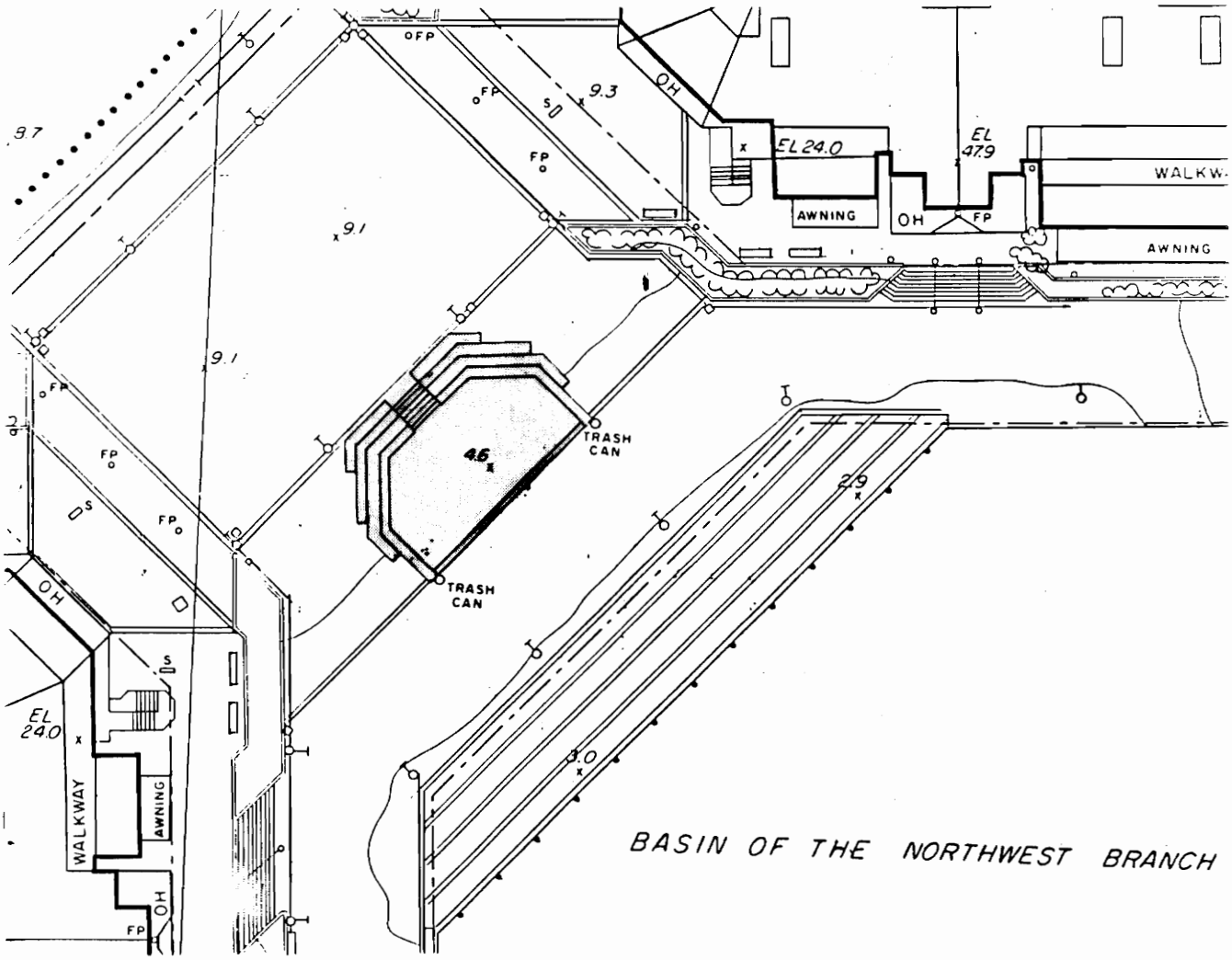
24. Im, S.-B. "Visual Preferences in Enclosed Urban Spaces: An Exploration of a Scientific Approach to Environmental Design." *Environment and Behavior* 16, No. 2 (March 1984): pp. 235-261.

25. Brown, B. "Young Children's Play on Playgrounds: An Observational Study Overview." *Environment and Behavior* (September 1984): pp. 610-625.

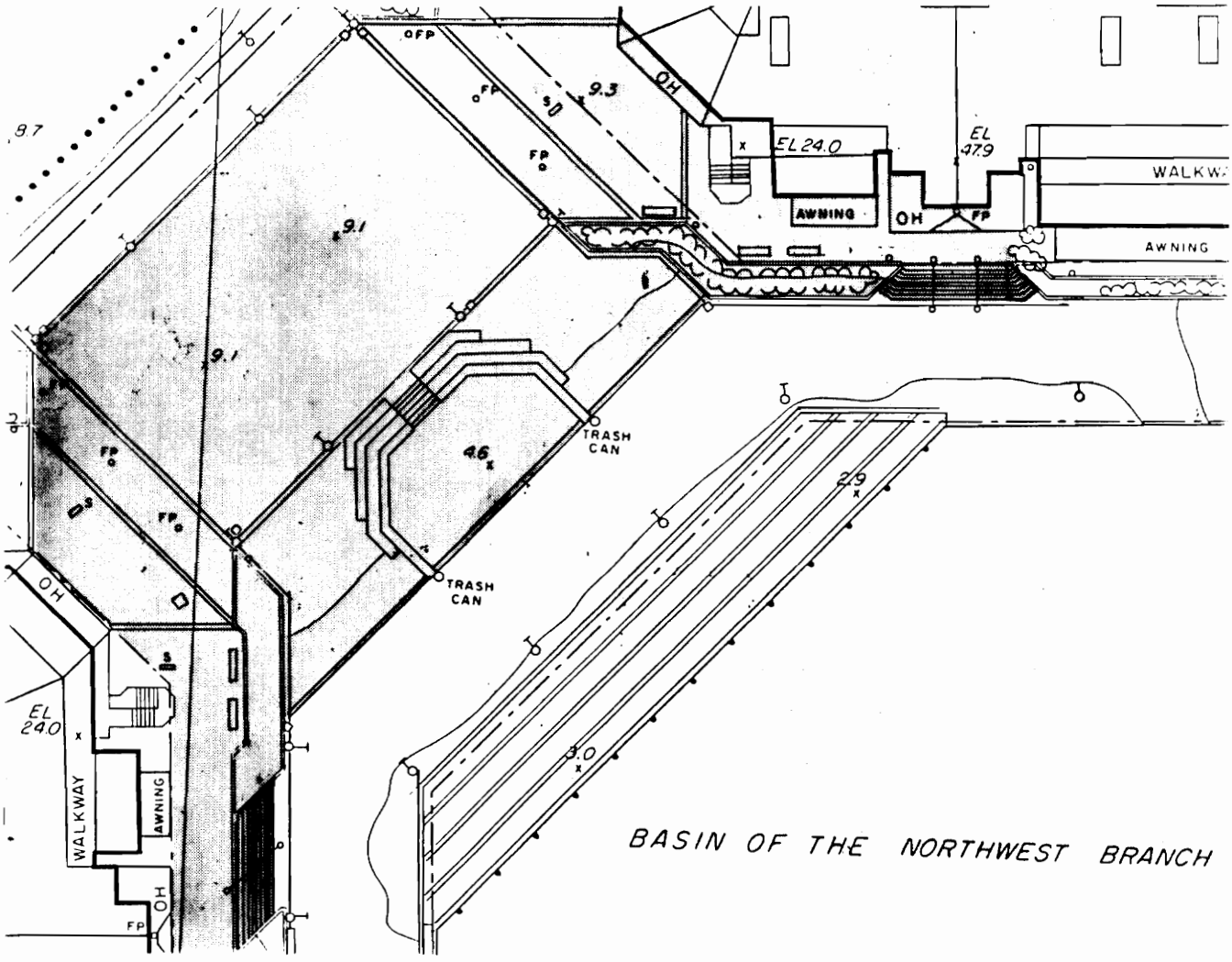
26. Anderson, J. "Scale Models in the POE Process: How They Strengthened the POEs of Two Public Housing Sites." draft submission for *Proceedings of the Seventeenth Annual Environmental Design Research Association Conference*, pp. unknown. Atlanta, n.p., 1986.

27. Chidister, M. "The Effect of Context on the Use of Urban Plazas." *Landscape Journal* 5, No. 2 (Fall 1986): pp. 115-127.

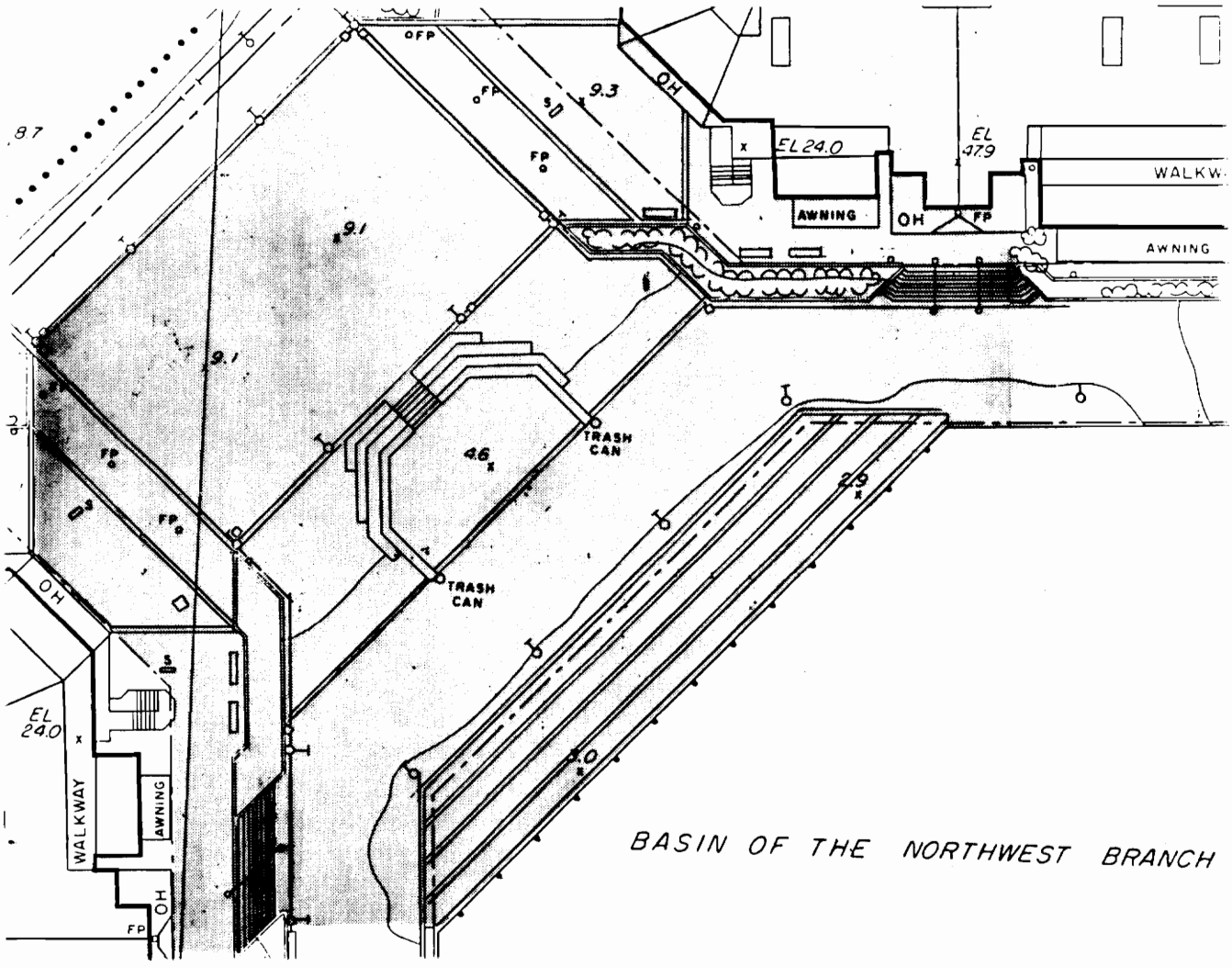
Appendix A. Study Area - Day One



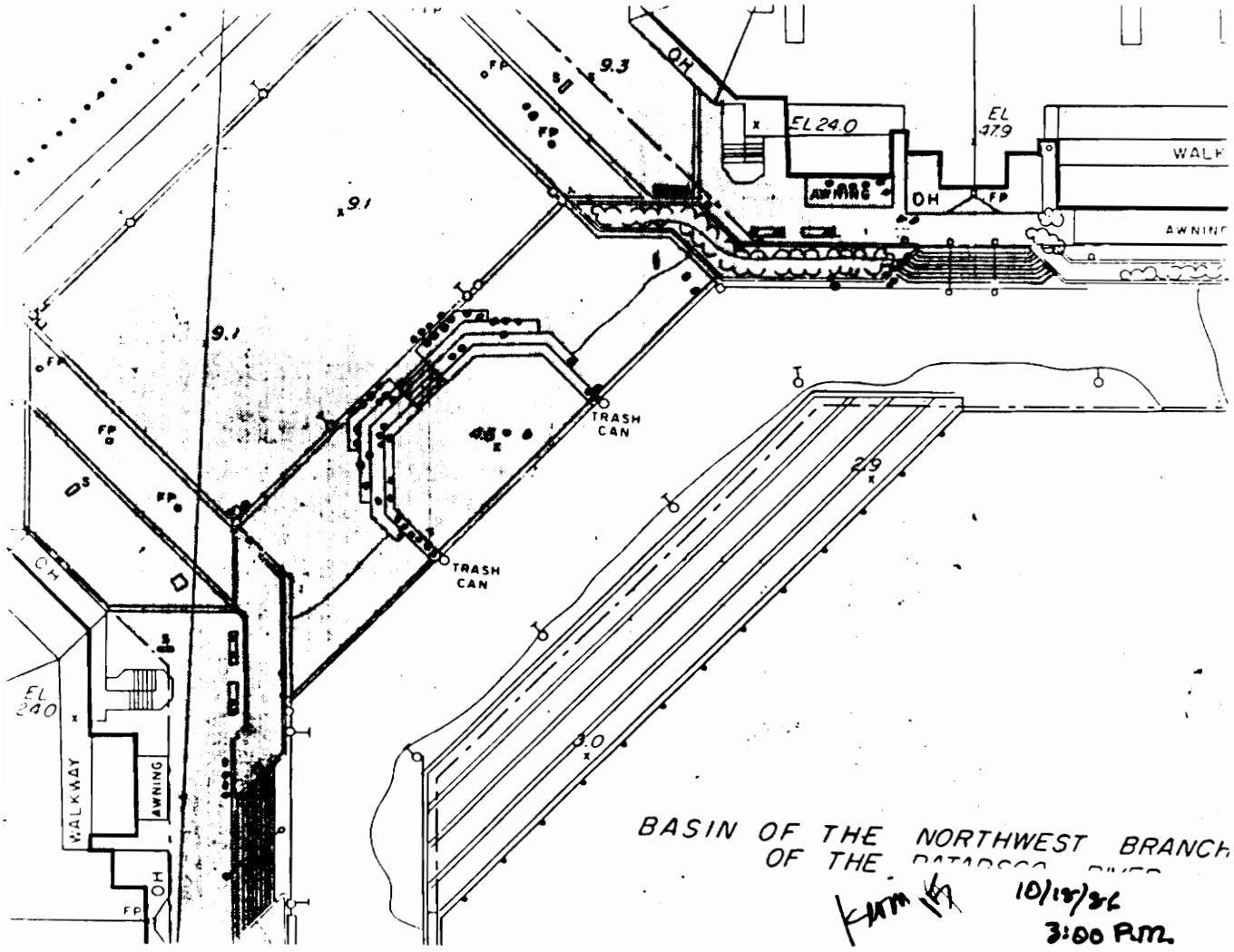
Appendix B. Study Area - Day Two



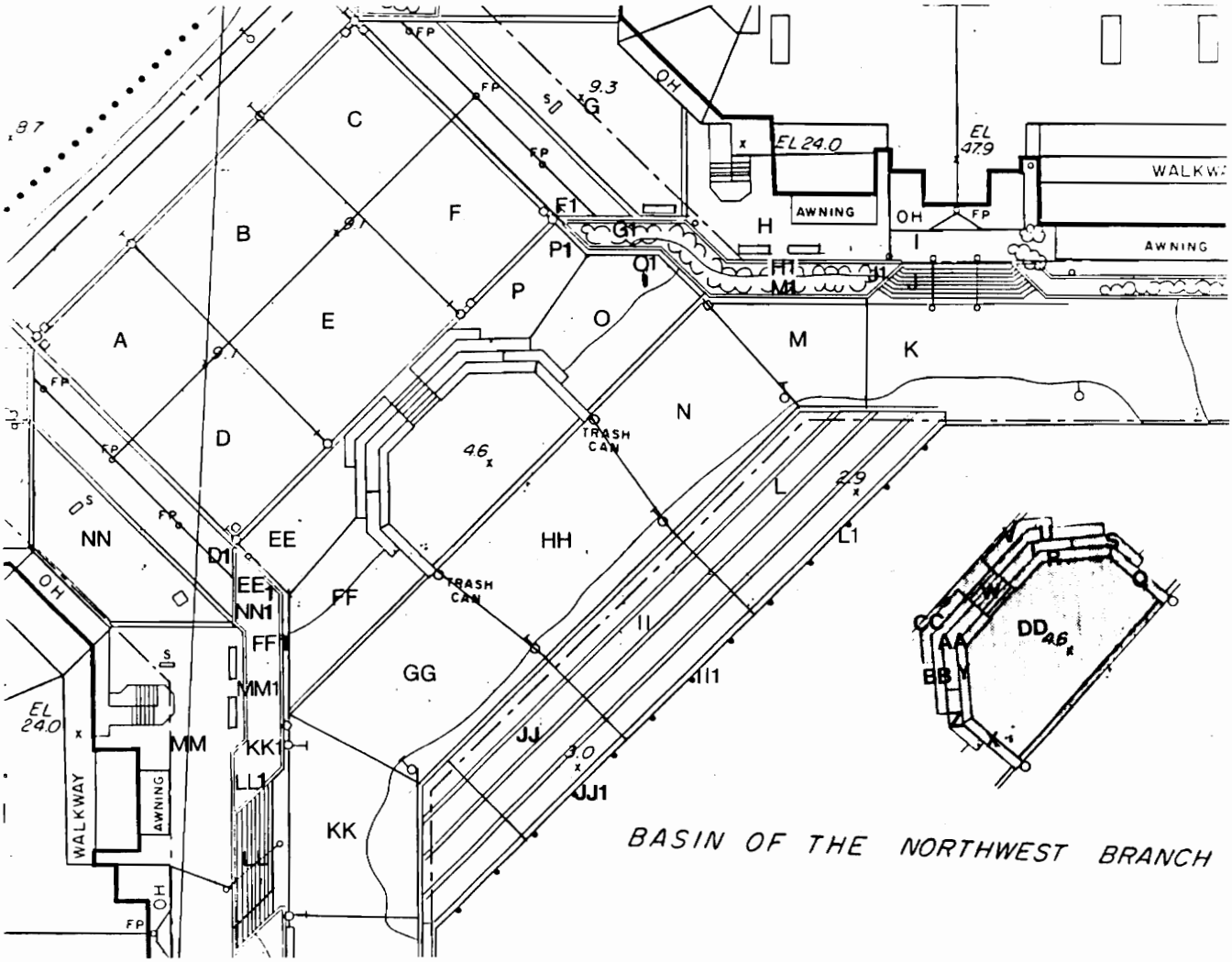
Appendix C. Study Area - Day Three



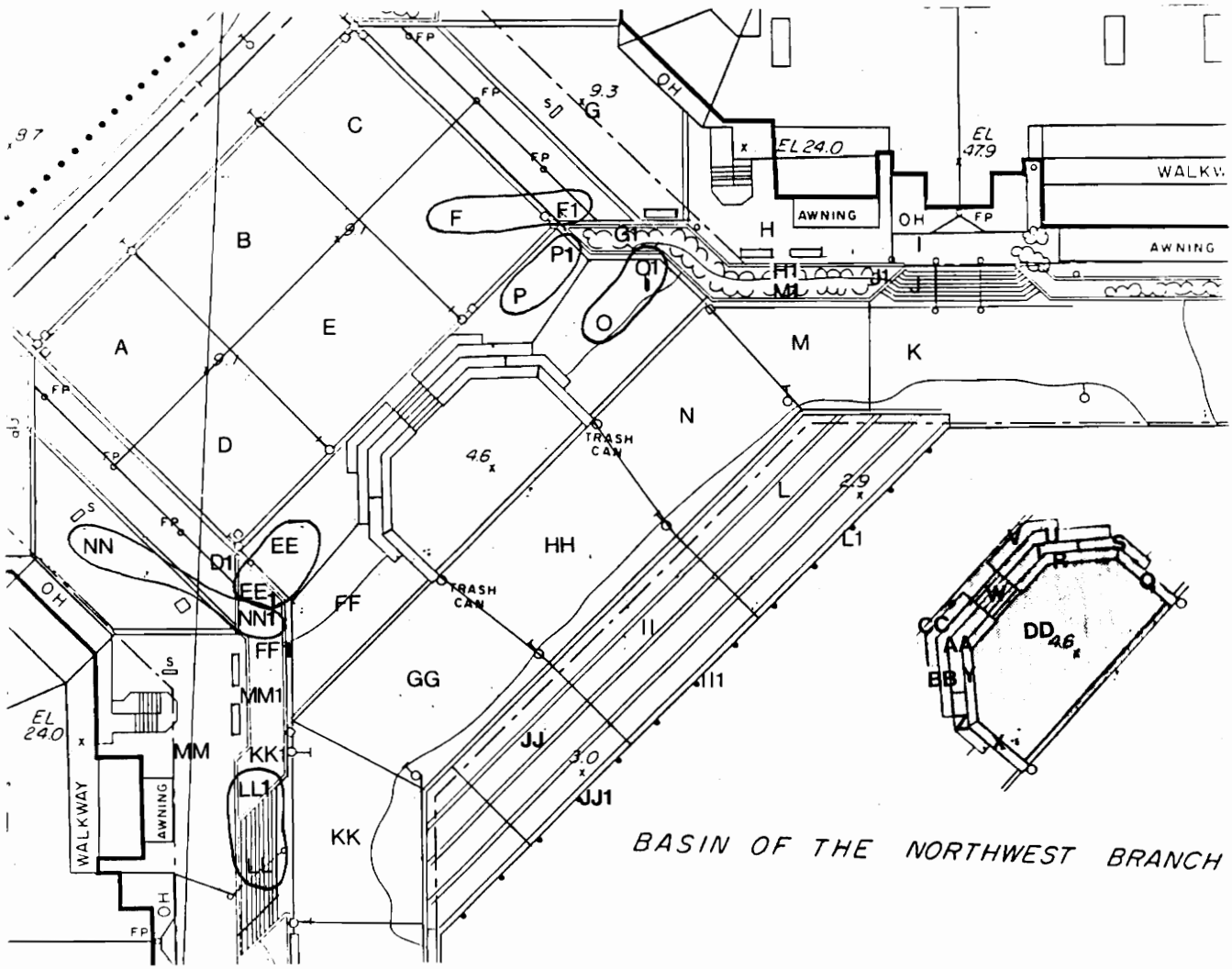
Appendix D. Raw Data Sample



Appendix E. Spatial Subdivisions

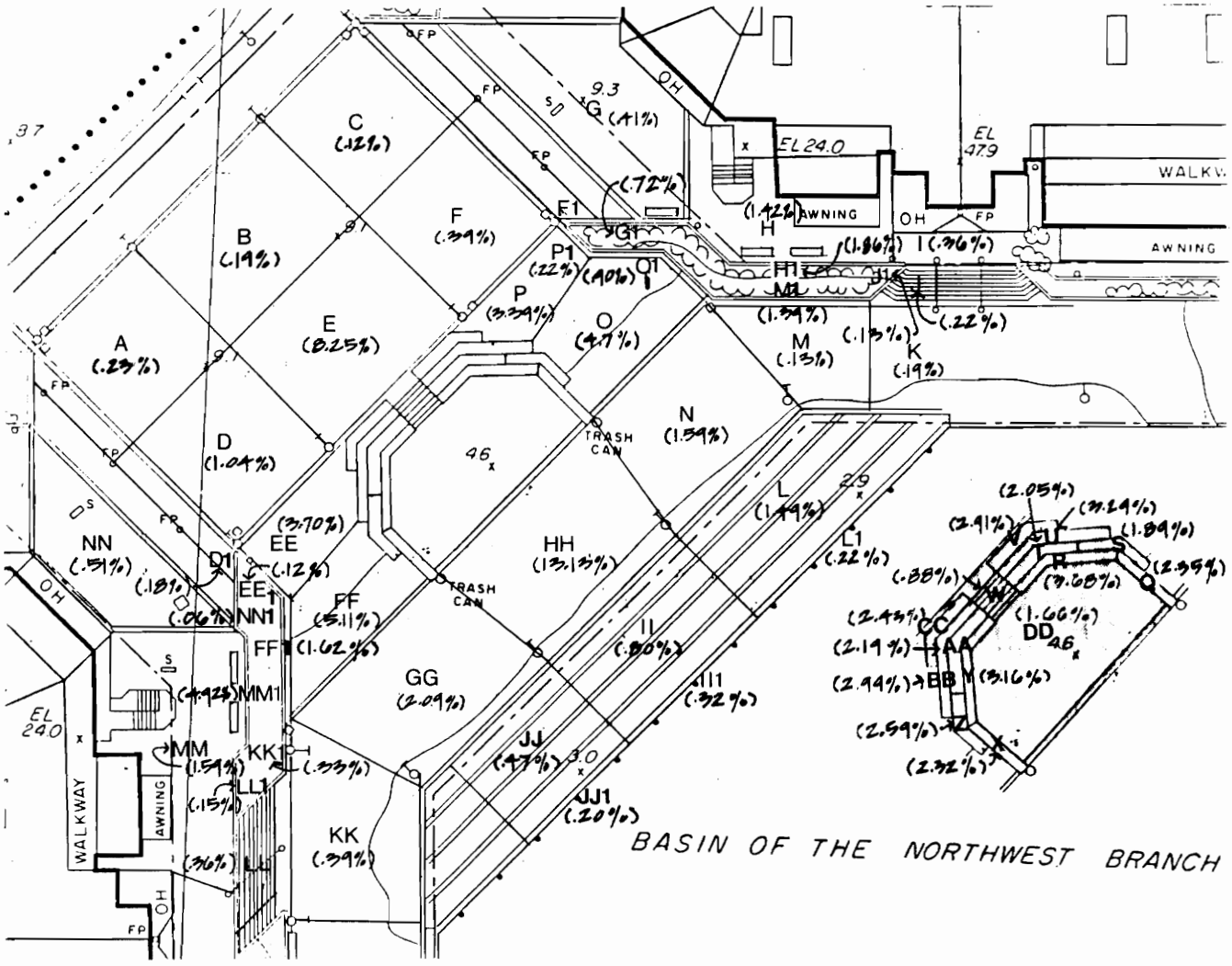


Appendix F. Revised Spatial Subdivisions



Appendix G. User Density

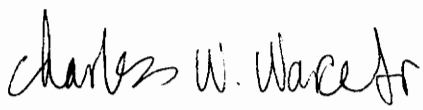
(Represents percentage of total users for three days of data collection)



Vita

Charles W. Ware Jr. completed a Bachelor of Science in Landscape Architecture from Ohio State University in 1985 as an honor graduate. There he was awarded the Faculty Prize in Landscape Architecture. During his junior year at OSU Charles served as an undergradute teaching assistant for a landscape architecture graphic communication course. Between his junior and senior year he interned at the Landscape Architecture Foundation in Washington, D.C., responsible with theoretical development and data base expansion of a landscape architecture research and information clearinghouse (LAFRICH). While completing his senior year thesis Charles worked as a design intern at the Columbus Neighborhood Design Assistance Center. During the same year he was inducted into the Sigma Delta Psi Athletic Honorary for athletic accomplishments.

Before entering graduate school Charles worked as a design intern at John Rahenkamp Consultants, Inc. - Planners, Land Planners, Landscape Architects in Philadelphia, Pa. At VPI & SU in Blacksburg, Va. he received a Master Degree in Landscape Architecture in 1987. Therein he received the Faculty Award for Exceptional Contribution to the Program for his work as a graduate teaching assistant. He also recived the American Society of Landscape Architects Merit Award. For his work on the present document he was given the Stanley W. Abbott Award for Excellence in a Graduate Thesis Project. The thesis research was funded by a VPI & SU Graduate Research and Development Project Award and a Landscape Architecture Foundation Student Research Grant. As part of the top 15% of the graduate school he served as a member of the Gamma Beta Phi Scholastic Honorary and Service Organization and also is a member of the Sigma Delta Psi National Landscape Architecture Honor Society.



Charles W. Ware Jr.