THE UNIVERSITY OF CALGARY

# RESPONSE EFFECTS AND COMPUTERIZED QUESTIONNAIRES: THE ROLE OF THE ENTRY TASK AND RESPONDENT'S PREVIOUS COMPUTER EXPERIENCE

BY

GREGORY R. BRATTON

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF BUSINESS ADMINISTRATION

THE FACULTY OF MANAGEMENT

CALGARY ALBERTA

AUGUST, 1989

ISBN 0-315-54188-1

All Figures, Tables, and Appendices in the thesis "Response Effects and Computerized Questionnaires: The Role of the Entry Task and Respondent's Previous Computer Experience" originate from the author.

The only exception being questions [32] through [36] in Appendix B page 121. These questions originate from the survey "Canadian Tourism Attitude and Motivation Study" performed by Tourism Canada in 1983. This source being described and referenced on page 54 of the thesis.

Signature

Sept. 11/89

Date

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "Response Effects and Computerized Questionnaires: The Role of the Entry Task and Respondent's Previous Computer Experience", submitted by Gregory R. Bratton in partial fulfillment of the requirements for the degree of Masters of Business Administration.

Supervisor, Dr. Peter Newsted
Faculty of Management

Dr. J. R. Brent Ritchie
Faculty of Management

Dr. Larry Katz
Department of Physical Education

Date September 11, 1989

## ABSTRACT

The use of computers in the administration questionnaires is becoming more prevalent in survey research. This technology has been used in business and organization research, with surveys distributed through electronic mail systems or via diskettes. It has also been used in the field, where respondents enter answers directly on the computer; or with the questionnaire operating on un-supervised computers located in public areas. However, few studies have examined the validity and reliability of data collected by this relatively new collection methodology.

This study presents a conceptual framework of variables that create or contribute towards response bias associated with the use of computerized questionnaires. Two of these variables, variation in the entry task, and the respondent's previous computer experience were examined in a field based experiment. The study also evaluated the reliability of a sample collected by a computerized questionnaire used in an unattended methodology. Both components of the study were completed in a field experiment at a tourist attraction.

A between subjects design with three groups was used to compare pen and paper survey responses against those collected from two computerized questionnaire packages; one using a cursor movement answering protocol, and the second using number pad entry. Four response effects were identified for the groups completing computerized questionnaires: an increase in the range of response to scale questions, a variable effect on the mean response to scale questions, a tendency to select only one choice in answering multiple response questions, and longer answers to open ended questions. Response effects were shown to be as much a function of the specific computer entry task, as the computerized interviewing situation in general. This finding suggests that computer collected data should not be compared with data from other methodologies or even different questionnaire software. Respondent's previous computer experience was shown to have only a marginal effect on the size of response effects, and was evident for only one of the two answering protocols.

The sample collected from an unattended version of the questionnaire collected during the same time frame clearly demonstrated that allowing respondents to self-select resulted in a biased sample, which, even after adjusting for a large number of teenage respondents, could not accurately estimate the population frame of visitors to the attraction. The thesis gives some suggestions for alternative uses of questionnaires administered without human intervention.

## ACKNOWLEDGEMENTS

Though I take great pride in claiming to be directly involved in the creation of this document, thanks must go out to those who have helped me in this pursuit. First, a special thanks to my wife, for agreeing to the financial sacrifice that we both shared over these last two years, and more importantly for her patience and understanding when I was about as pleasant as cornered cougar.

I must thank my advisor, Dr. Peter Newsted, who's efforts to give complete and immediate feedback, particularly in the final stages of the thesis, were above and beyond the call of duty. As well, I must thank Dr. Brent Ritchie for encouraging me to undertake a tourism related thesis in the first place, and Dr. Wilfred Zerbe, for his assistance in explaining the appropriate use of multivariate statistics. Through the use of the BITNET electronic mail network, I was also able draw on Dr. John Liefeld's (University of Guelph) wealth of knowledge.

I am an also deeply indebted to my parents for their assistance. My mother played an instrumental role in the field data collection, and my father (himself an academic) gave feedback when it was requested, and more importantly, knew when it was best to keep his opinions to himself.

Finally, I thank Reena Bassen, for her assistance in arranging the keys, scheduling, paperwork and dealing with other hassles that were a necessary part of the attraction's involvement in the study.

## DISCLAIMER

In order to perform this research, it was necessary for the author to sign a licensing agreement with the attraction in question. In order to comply with confidentiality restrictions with respect to the publication of results, it was felt most appropriate to keep the identity of the attraction hidden. The only reporting implication of this decision is that the questionnaire included in Appendix B has been modified in content, but not in form, to hide the identity of the attraction. In the event that,for any reason, a reader of this document is able to ascertain the location of the study and the attraction in question, it must be stated here that "the attraction did not conduct the study or participate in the study or any part thereof and makes no warranty as to the veracity of the results".

## TABLE OF CONTENTS


| | PAGE |
|---|---|
| Title Page | i |
| Approval Sheet | ii |
| Abstract | iii |
| Acknowledgements | v |
| Disclaimer | vi |
| List of Tables | x |
| List of Figures | xii |
| 1. INTRODUCTION | 1 |
| 2. DEFINING BIAS IN SURVEY DATA COLLECTION | 4 |
| 2.1 Sampling Biases | 4 |
| 2.2 Response Effects and Artifacts | 5 |
| 2.3 Survey Theories | 7 |
| 3. RESPONSE EFFECTS AND ARTIFACT IN TRADITIONAL DATA COLLECTION METHODS | 9 |
| 3.1 Collection Method | 10 |
| 3.2 Characteristics of the Question | 10 |
| 3.3 Interviewer Effects | 12 |
| 3.4 Characteristics of the Respondent | 13 |
| 3.5 Methodologies That Reduce Potential Response Effects | 14 |
| 4. RESPONSE EFFECTS AND ARTIFACT IN COMPUTERIZED DATA COLLECTION | 16 |
| 4.1 Field Benefits in Computer Interviewing | 16 |
| 4.2 A Causal Model for Response Effects in Computer Interviewing | 19 |
| 4.2.1. Characteristics of the Respondent | 23 |
| 4.2.2. Interviewer Effects - The Computer as an Interviewer | 26 |
| 4.2.3. Question Content | 30 |

PAGE


4.2.4. Presentation ..... 32
4.2.5. Human Factors ..... 33
4.2.6. Response Task ..... 36
4.2.7. Amount of Feedback ..... 37
4.2.8. Artifact ..... 39
4.3 Summary of Findings ..... 40
4.4 Rationale for This Study ..... 44

5. OBJECTIVES ..... 46

6. HYPOTHESES ..... 48

7. METHODOLOGY ..... 52
7.1 Procedure and Design ..... 52
Part A - Response Effects ..... 52
Part B - Unattended Kiosk ..... 53
7.2 The Questionnaire ..... 54
7.3 Pre-test ..... 56
7.4 Measures ..... 56

8. RESULTS ..... 59
8.1 Part A (Response Effects) ..... 59
8.1.0 Data Analysis and Distribution Assumptions ..... 60
8.1.1 Mean Response to Scale Questions ..... 61
8.1.2 Range of Response to Scale Questions ..... 63
8.1.3 Multiple Response Questions ..... 64
8.1.4 Open-ended Questions ..... 68
8.1.5 Cooperation ..... 69
8.2 Effects of the Entry Task ..... 71
8.2.1 Mean Response to Scale Questions ..... 71
8.2.2 Range of Response to Scale Questions ..... 71
8.2.3 Multiple Response Questions ..... 73

PAGE

8.3 Effects of Previous Experience with Computers ..... 73
8.4 Part B (Unattended Kiosk) ..... 76

9. DISCUSSION ..... 79
9.1 Response Effects For Computerized Completion in General ..... 79
9.1.1 Scale Questions ..... 79
9.1.2 Variance in Answers to Scale Questions ..... 79
9.1.3 Multiple Response Questions ..... 80
9.1.4 Open Ended Questions ..... 80
9.2 Effects of the Entry Task ..... 83
9.3 Effects of Previous Computer Experience ..... 84
9.4 Effects of Other Variables ..... 85
9.4.1 Presentation ..... 85
9.4.2 Content ..... 88
9.5 Estimate of the "Bias Potential" of Variables ..... 89
9.6 Implications ..... 91
9.7 Suggestions for Creating Better Questionnaire Software ..... 92
9.8 Unattended Kiosks ..... 95
9.9 Limitations ..... 98
9.10 Future Studies ..... 99

10. SUMMARY ..... 102

REFERENCES ..... 104

APPENDICES
Appendix A. . Screen Snap Shots of Multiple Response
and Various Scale Type Questions ..... 112
Appendix B Questionnaire
(Content disguised to protect the identity of the attraction) ..... 117

## LIST OF TABLES


| | | PAGE |
|---|---|---|
| Table 1 | Ability of an Unattended Kiosk to Obtain a Representative Sample at Canadian Airports | 26 |
| Table 2 | Reason For Refusal | 59 |
| Table 3 | Comparison of Demographics By Method of Completion | 60 |
| Table 4 | Mean Response to Twenty Scale Questions | 62 |
| Table 5 | Standard Deviation of Responses to Scale Questions | 64 |
| Table 6 | Average Number of End Points Selected | 65 |
| Table 7 | Multiple Response Questions: Number of Choices Selected by Question | 66 |
| Table 8 | Multiple Response Questions: Mean Number of Choices Selected | 67 |
| Table 9 | Average Number of Words and Concepts in Answers to Open-ended Questions | 68 |
| Table 10 | Unanswered Questions: SC Group | 70 |
| Table 11 | Number of Respondents Not Answering Open-ended Questions By Method of Completion | 70 |

PAGE

Table 12 Mean Response to Selected Scale Questions..... 71

Table 13 Correlation of Computer Experience
and Transformed Response Effect Measures ..... 74

Table 14 Correlation Between:Previous Computer Experience
(As a Dichotomous Variable)
and Transformed Response Effect Measures ..... 75

Table 15 Comparison of Distributions of Demographic Variables:
Sample Collected by Unattended Kiosk
Verses an Estimate of the Population ..... 78

Table 16 Frequency of Selecting Only One Choice
for Multiple Response Questions..... 81

Table 17 Distribution of Concepts Presented ..... 82

Table 18 Mean Response to Scale Questions:
Organized by Presentation ..... 87

Table 19 Mean Response to Scale Questions:
Organized by Question Content ..... 88

Table 20 Comparison of Sample Collected by Unattended Kiosk
(person under 18 removed)
Verses an Estimate of the Population ..... 96

## LIST OF FIGURES

PAGE

Figure 1 A Conceptual Framework for Bias in Computerized Data Collection ..... 21

CHAPTER 1

# INTRODUCTION

With the recent proliferation of micro computers and inexpensive software, computers are now finding a place in the survey interview process. Marketing research firms have been using Computer Aided Telephone Interviewing software to aid in the management of large telephone surveys for some years now (Groves et al., 1979). Recent trends, however, move the technology one step closer to the respondent. Data that was traditionally collected by self-completion questionnaires or personal interviews can now be gathered with interviewing programs, where respondents complete a questionnaire interactively (Minno, 1986). The interactive nature and logical ability of these programs make them well suited for field use, particularly to eliminate errors associated with failure to follow complex skipping or branching patterns. As well, the elimination of the data entry tasks makes the collection method attractive, purely from an economic standpoint.

Described as computerized interviewing, computer administered questionnaires, diskette surveys, or interactive questionnaires, this electronic data collection process can take a variety of forms. Electronic mail systems are being used to gather data from persons within an organization (Sproull, 1986). Questionnaire programs can be stored on floppy diskettes and distributed through the mail. Respondents with access to personal computers can complete the questionnaire interactively and return the diskette to the researcher (Higgins et al., 1987). Various marketing companies are using unattended computer terminals to poll on campuses and at shopping malls. Reported benefits are: elimination of interviewer bias, forty to fifty percent savings in data collection costs, elimination of coding and data entry errors, and the ability to track data on a day to day or week to week basis. (Marketing News, 1985a).

However, as with any new data collection method, researchers should first assess the reliability and validity of data collected. Current studies suggest that while interactive

software has a place in the survey data collection process, this technology does introduce interesting artifacts, response effects and sampling biases that would not be found in traditional data collection methods. Researchers have found that it can encourage more guessing to multiple response questions, and generate more positive agreement to scale attitude questions (Liefeld, 1988). It has also been found to elicit more cooperation from respondents and increase response rates (Higgins, 1986; Sproull, 1986; Liefeld, 1986; Allen, 1987; Erdman et al., 1985). It has been suggested that the computer will increase honesty, or reduce the amount of social desirability in the response (Evan and Miller, 1969; Lucas et al., 1977; Allen, 1987; Millstein, 1987; Canoune and Leyhe,1985). While computerized questionnaires can eliminate the requirement for human supervision, studies suggest that unattended terminals in public places may attract an over-representation of younger respondents (Newsted, 1985; Okimoto, 1988). Little work has been performed on the effects of variations in screen design or on-line help facilities, and various input protocols such as light pens and touch screens. In fact, this review was unable to located any studies which incorporated a comparison of the effects of using different questionnaire software.

Identifying the *source* of response effects is valuable in gaining a better understanding of the strengths and the limitations of this data collection tool. This information will, in turn lead to better decisions regarding appropriate use of interactive computer interviewing. In this regard this study has attempted to organize studies undertaken in the various fields of psychology, human factors, and marketing research into a conceptual framework that can be used by future studies in this area.

Presentation will begin with a definition of terminology, arguing that bias in survey data collection is comprised of three distinct components: sampling error; artifact; and response effects elicited by the data collection instrument. The review of the literature has been organized into two sections: the first which reviews response effects and artifact in traditional data collection methodologies, and the second which utilizes this groundwork to treat response effects specific to computerized data collection methodologies. From this discussion two variables are identified for further examination: the characteristics of the entry task, and the respondent's previous computer experience.

The study will also incorporate the use of a relatively new methodology utilizing computerized questionnaires administered from an unattended kiosk, with subjects participating through self-selection. This will provide practitioners will useful comments as to the reliability and validity of data collected through this novel use of the technology.

# CHAPTER 2

# DEFINING BIAS IN SURVEY DATA COLLECTION

The term bias is often used in a generic sense, yet the concept can be divided into three distinct components. Survey results can be biased in one of three ways. First, it is possible to select a sample that is not representative of the population under consideration. The extent to which results are not generalizable to the entire population is referred to as sampling bias. Second, results of an experiment or study can be affected by an uncontrolled or intervening variable, referred to as an artifact. Finally, bias can manifest in measurement errors elicited by the collection instrument. These are referred to as response effects. A brief discussion of sampling bias is followed by a more specific distinction between response effect and artifact, and some common methods for measuring response effects in survey research. This chapter concludes with a summary of two conceptual models for understanding bias in survey research.

## 2.1 Sampling Biases

Survey data collection is based on the principle that a small proportion of a population can be used to make generalizations about the characteristics or opinions of the entire population, provided that a random selection technique is used and each member of the population has equal chance of being selected. It is from these principles that properly conducted national poles of less than one thousand people can predict how the entire country will vote in an election (Nachmias and Nachmias, 1981). The key, however, is in choosing a method that will successfully select a sample that is representative of the population. For example, if questions regarding musical listening preferences were asked of only those persons under twenty-five, the answers given by survey respondents would most certainly differ from those of the general population. The utilization of an unattended terminal, which allows respondents to self-select, creates a great manpower saving in data collection costs. The concern, however, is in selecting an unbiased estimator of the population. Sampling bias is the extent that the

characteristics of the persons attracted to complete a computerized questionnaire are any different from the characteristics of the population frame as a whole (which, in the case of a site exit interview would be considered to be anyone who uses the facility).

## 2.2 Response Effects and Artifact

Artifact and response effects are both essentially unwanted variables that distort results. Artifacts tend to be *unknown* variables which influence the results of an experiment or study. In creating a controlled environment researchers may inadvertently be introducing new variables (or fail to recognize and control for exogenous variables) which would not be present in the "real-world" relationships they are trying to explain. McGuire (1969) portrays the artifact as having a three stage life, the stage of ignorance, the stage of coping, and the exploitation stage.

Response effects, on the other hand, tend to be more associated with measurement errors, rather than unknown variables. A response effect is essentially a measure of the difference between the results obtained by a particular data collection method and the "true" answer. In distinguishing between the two, artifacts that reach the exploitation stage and are understood become response effects, provided that the artifact can be generalized over a variety of situations. For example, a mall intercept survey performed the week after the city's hockey team won the Stanley Cup may have an artifact of overly positive responses to all attitudinal questions. (In fact, the artifact may not be discovered unless follow up surveys were executed and other plausible explanations ruled out.) Once other studies in similar situations are performed (other cities, for example), the effect of losing the final is examined, and some type of explanation is derived, the artifact reaches the exploitation stage. At this point the artifact can be explained in terms of a response effect associated with, say, the general level of jubilation and celebration.

Response effect can take on numerous forms, including:

1) unintentional reporting errors such as telescoping and memory errors (Sudman and Bradburn, 1974);

2) misrepresentation on the part of the respondent in order to maintain positive self presentation (Sudman and Bradburn 1979, p. 163);
3) misinterpretation of the question (Loftus, 1982);
4) response patterns or inaccuracies created by the collection instrument (Liefeld, 1987);
5) non-response, for example to threatening questions (Locander, 1976);
6) interviewer expectations biasing interpretation and recording of responses (Rosenthal and Fode, 1963).

In measuring response effects, results are commonly broken into two components, a true component and an error component. Nachmias and Nachmias (1981) algebraically represent this concept as:

$$X = T + E$$

where: X = actual response
T = true response
E = error or biasing component of actual response

The difficulty in determining the magnitude of a response effect, if indeed one exists, is in estimating the *true* response. A few alternative methodologies are possible, the most accurate being to ask questions about data which is verifiable from other sources. Sudman and Bradburn (1979) performed a large survey of this nature which asked respondents about traffic violation, bankruptcy, drunk driving, and voting behaviors (X). Responses where then compared against actual state and municipal records (T) to determine the error component (E).

In a test-retest procedure measurement instruments are administered to the same group of persons at two different times. The difference in results between the two methods is interpreted as the response effect of administration. The limitations of this method are the influence that completing the first test will have on the results of the second, and the possibility of change in the measured property between the two tests

(Nachmias and Nachmias, 1981). Regardless, this method has been popular in testing computerized administration of psychological tests. (Hurrell and Lombardo, 1984; Russell, Peace and Mellsop, 1986; Resmovic, 1977).

Equivalent forms method uses correlation measures to compare results of two measurement forms. Again, one method is taken as the baseline and differences between methods are interpreted as the response effect. A variation which uses two sample groups has been the favored method used to determine the validity of computerized data collection, where traditional methods of data collection such as pen and paper questionnaires and personal interviews are used as measures of the "true" response (Liefeld, 1988; Higgins, 1987; Kiesler and Sproull, 1986).

Sudman and Bradburn (1974) performed an extensive review of the literature in an attempt to categorize response effects in traditional data collection methods. In order to determine the size of effect of biasing variables they created a scale to measure the magnitude of response effects. This scale quantifies relative effect (RE), or biasing influence of a variable as:

$$RE = \frac{\text{(Actual Response - Validating Response)}}{\text{standard deviation of the population}}$$

The standard deviation of the population is obtained, if possible, from the validating response information, and is necessary to create a common scale. Validating responses were obtained by checking secondary sources such as government records, or by using alternate methods of data collection. For those studies reporting data about attitudes there would be no 'true' or valid answer. In this case alternative verifying methods of data collection were used and the weighted mean of all these responses was used as a basis for comparison.

## 2.3 Survey Theories

While numerous studies have identified specific biases, few works have attempted to create a comprehensive model to categorize the variables at work. Two

conceptual frameworks, however, have been suggested. In order to explain response effects researchers have treated the interview as a social interaction task. Dillman (1978) treats the data collection process as a social exchange with three variables: the cost of responding, the rewards for responding, and trust that the rewards will be delivered. This theory is used to explain response rates and it is suggested that response rates for various data collection methods can be improved by: a) minimizing the cost of responding, b) maximizing the reward for responding and, c) establishing trust that the reward will be delivered.

After completing their extensive analysis of the literature Sudman and Bradburn conceptualized the interview as a " micro-social system in which there are two roles, that of respondent and that of interviewer, joined by the common task of giving and obtaining information." (Sudman and Bradburn 1974, p. 6). The respondent role comprises such variables as the characteristics of the respondent (sex, age, race, yea-sayer) and their motivation (anxiety, hostility, effort). The interviewer role comprises such variables as the characteristics of the interviewer and their role performance (training, expectations). The task, which was found to provide the greatest amount of bias, comprises such variables as: the effect of questions, questionnaire design and the interviewing situation, time and memory factors, and the type of data collected (attitudinal or behavioral questions).

Both models use the context of a social interaction setting, however Dillman approaches the problem with an intent to understand methods of increasing response rates, only one of many potential response effects. Sudman and Bradburn's model is much more global in nature. They would describe cost and rewards of responding as variables that can be affected by characteristics of the task such as the interview situation, and that establishing trust is a function of the roles and actions of the interviewer.

# CHAPTER 3

## RESPONSE EFFECTS AND ARTIFACT IN TRADITIONAL DATA COLLECTION METHODS

Reviewing the literature on response effects in traditional data collection methods is a necessary first step of understanding response effects in computerized data collection methodologies. However, due to the variety of forms response effects can take, and the fact that many artifacts may be situation specific, it is difficult to present a concise synopsis of findings in the area. Most studies are instrument or situation specific and the task of integrating these findings is difficult. Sudman and Bradburn (1979) have been instrumental in attempting to construct a framework for understanding bias. They have also been involved in performing or analyzing many studies dealing with response effects. Yet, even their presentation, using the three component framework of task, interviewer and respondent, fails to adequately deal with the *interaction* effects.

For example does one depict acquiescence as a function of the respondent, or the task? Certainly some respondents will be more likely to agree with statements, but this artifact may only be exhibited in a poorly designed measurement instrument. With respect to question content it would be appealing to be able to state for example, "responses to threatening questions exhibit more bias when compared to responses to non-threatening questions." However, the amount of bias demonstrated will not necessarily be the same depending on presentation of the question (Converse, 1986) or method of administration (Locander, Sudman and Bradburn, 1976). Add to this the possible effects of a respondent demonstrating chronic anxiety (Sudman and Bradburn, 1979 p.64) or response variation due to perceived social class or status of the interviewer (Sudman and Bradburn, 1979, p.109). With this difficulty in mind this chapter will attempt to present a summary of findings in traditional data collection methods, presenting variables that have been shown to bias results: collection method, characteristics of the questions, interviewer effects, and characteristics of the respondent.

## 3.1 Collection Method

Considerable attention has been given to the the response effects that may result from or can be eliminated by specific data collection methods. Traditionally, data collection methods have included: group, face to face, telephone, self-administered, mail or diary collection methods. Randomized response techniques were first introduced by Warner (1965) and have been used on a number of occasions since.[1] However, success in eliminating underreporting to threatening questions has been limited (Locander et al., 1976; Brewer, 1985) and often does not justify the additional costs of administration.

Rosenberg (1969) coined the term 'evaluation apprehension'. This term is used to explain actions that arise because respondents feel they are being judged psychologically and do not wish an unfavorable evaluation. He cites experiments where most subjects prefer to look good rather than cooperate if there is a conflict between demand characteristics and projecting a good appearance. Locander et al. (1976) found that people are more likely to overstate socially desirable traits or actions verbally, and more likely to report undesirable traits or actions if self-completion with pen and paper is the collection method. Bradburn (1983), in a summary of the literature, suggests: that collection method introduces no bias for factual non-threatening questions, that sensitive questions are better answered with anonymous methods, and that motivation is improved with personal contact.

## 3.2 Characteristics of the Question

Specific characteristics of the question, such as order of presentation, have been shown to be important influencing variables. Attitude research suggests that respondents will try to respond in a consistent manner and that the order of questions may increase or decrease cues for consistency. This effect was best shown in a pre-1941 study about

---

1 Respondents are presented with two questions, one of which is threatening. A probability mechanism visible only to the respondent indicates which of the two questions is to be answered. In this manner the *interviewer* is not aware of which question is being answered but by using the known probability of the mechanism, aggregate responses can be calculated.

willingness to allow Americans to join the British or German army. More respondents were willing to allow Americans to join the German army if this question was asked after a similar question about allowing Americans joining the British army than when the two questions were presented in the reverse order (Cantril, 1944). In analyzing effects that question order has on results Bradburn (1983) found that: 1) low saliency questions could be effected by order, 2) that respondents tend to respond in a consistent manner, and as such response to a previous question may affect the result of a question later on in the questionnaire, and 3) fatigue affects response to questions.

Poor wording can contribute to response effects, simply through misinterpretation of the question. Subtle word changes such as replacing "a" with "the" (which implies that the object in question did actually exist) or "occasionally" with "frequently" have been shown to increase positive responses to questions (Loftus, 1982). While practitioners will encourage researchers to keep questions short and simple, studies have shown that increasing the length of the question will increase completeness of responses (Sudman & Bradburn, 1974, p.36) and, in the case of sensitive questions, long open-ended questions can increase the amount of reporting of behavior by two to three times (Sudman & Bradburn, 1979, p.167). Converse (1986) suggests that this phenomenon can be explained by the fact that additional spoken material by interviewer will encourage longer responses and that this aids in recall. Alternatively, it may also simply be a function of allowing the respondent more time to think about their answer.

In analyzing the effect caused by characteristics of the question it is useful to distinguish between effects caused by variation in content, and those caused by variation in question type. Of the two, variation in content appears to have the larger biasing effect. For example, attitudinal questions demonstrated larger response effects than behavioral questions (Sudman and Bradburn, 1974 p.31) across a of variety of task variables. In their review of the literature Sudman and Bradburn (1974) found the threat level of a question to be an important biasing variable, the higher the threat, the higher the response effect. With respect to responses to questions with various levels of threat, Sudman and Bradburn (1979, p. 64) found that persons reporting that questions regarding particular behaviors made them very uneasy, or annoyed them, were less likely

to report ever engaging in that behavior. However, once the behavior had been reported, the respondent's level of acute anxiety had no effect on the frequency of reported behavior. Locander, Sudman and Bradburn (1976) showed that this underreporting response effect to threatening questions is not generally affected by the way in which the question is asked.

Question type does not appear to be as important. In survey research, questions are typically open or closed, with closed presentation including such possibilities as the use of scales or ratios, agree or disagree, or single or multiple choice from a predetermined set of options. Sudman and Bradburn (1979, p.66) found no overall superiority for either open or closed questions; each was more appropriate for different situations.

### 3.3 Interviewer Effects

Research into the area of interviewer bias has concentrated on the response effects that the interviewers may cause in their interaction with the respondent. Rosenthal (1967) found that smiling affected the results of an experiment. In observing undergraduate experimenters, twelve percent smiled at male subjects while seventy percent smiled at female subjects. Sudman and Bradburn (1974) found that younger less experienced interviewers generally caused the largest response effects.

Experimenter expectations can also bias results. Rosenthal and Fode (1963) found that rats being viewed by experimenters expecting brighter behavior showed significantly superior learning compared to rats viewed by experimenters expecting dull behavior. There was, of course, no difference between the two sets of rats. The classic case of interviewer elicited bias involves research on destitute people performed by Rice in 1929 (referenced in Cannell, 1985). In re-reading some of the interviews conducted by assistants he found responses of one interviewer attributed most of the causes of destitution to economic factors, and the responses of another interviewer attributed destitution to the problems of alcohol. Upon talking to the interviewers he found one to be a socialist and the second a prohibitionist. The best defense against interviewer elicited response effects is to ensure that interviewers are properly trained in such things

as body language and prompting, and that they follow instructions carefully. Indeed, recent studies show that interviewer effects are minimal when dealing with trained interviewers, with the exception of the case with questions for which a visible characteristic (such as race) of the interviewer was salient (Sudman and Bradburn, 1974).

## 3.4 Characteristics of the Respondent

One response effect directly attributed to the respondent is error associated with poor recall. Memory errors and saliency are high biasing variables. In a study of hospitalization in a health interview, respondents who hadn't been in the hospital in the last 50 weeks failed to report 40% of hospitalization. Also from the same survey, failure to report a visit to a doctor increased from 15 to 30% from week one to week two suggesting that the less important and the longer the activity-interview interval, the less a person will report. This was augmented by the fact that persons would be less likely to report hospital stays due to the embarrassing nature of the ailment (Cannell, 1981). Sudman and Bradburn (1979) suggest that with low saliency, response effects increase, either as the client is more inclined to guess or will attempt to provide the most 'appropriate' answer .

Researchers have also investigated individual characteristics that may contribute to higher bias in responses. Sudman and Bradburn (1974) suggest that elementary school and college males are the most susceptible to biasing questions, but that this may be more a function of the interviewing situation than the characteristics of the respondent. Their review of the literature (1974) suggests that other respondent characteristics are manifest when there is interaction with other variables. For example response effects are larger for females on threatening questions which potentially evoke socially desirable answers.

Couch and Kenniston (1960) suggest that the acquiescence or agreeing response set is a function of personality, and as such some persons would be more likely to display this response set than others. It has also been suggested that respondent anxiety may have a distorting effect on demonstrated biases. For respondents exhibiting chronic anxiety, low threat questions increased response distortion by about 10 percent. The pattern becomes U-shaped for high threat questions, where persons demonstrating

medium level of chronic anxiety showed the highest level of response distortion (Sudman and Bradburn, 1979).

Crowne and Marlow (1964) have shown that persons with a high need for approval are easier to influence than those with a lower need for approval. However, in an analysis of data collected from a large sample in the city of Chicago, Sudman and Bradburn (1979, p. 85) suggested that variance associated with a ten question Crowne-Marlow scale was a part of the real variance in the data. The authors contend that while response distortion may exist based on the respondents need to portray a particular image, the Crowne-Marlow scale does not effectively measure this tendency.

## 3.5 Methodologies That Reduce Potential Response Effects

Rice (1984) discusses some impacts of data collection in what he terms the 'new media'. He suggests that research on new technology is often after the fact and seldom performed as the new technology is introduced. Often times control groups not using the technology are not included in the research. There can be problems in generalizing findings as the persons using new technology may be atypical of the general population. He suggests the following measurement techniques; 1) multiple measures from several independent sources, 2) objective data sources such as corporate records, 3) unobtrusive measures and 4) a measure of organization climate (Rice,1984). Rosenthal and Rosnow (1969) suggest that post experimental inquiries, non-experiments (where respondents are asked to describe their anticipated actions) and simulation could be used as a means of detecting and measuring potential artifact.

For longitudinal studies Teaborg et al. categorized variables into 1) alpha - change in response, 2) beta - instrument bias such as a change in use of a scale, and 3) gamma - redefining the conceptual domain (invalid findings), categories. They suggest that beta changes or artifact, which in the past have been undetectable, can be identified by including a 'then' measure. This measure is taken at the same time as the post measure and asks the respondent to give their perceptions of how they would currently rate themselves at the time of the *pre* measure. These pre/then/post measurements of means can be examined for correlations and variance of the unidimensional measurement

instrument at both the individual and group level. Then, for example, to determine if gamma change had occurred the correlation between pre/post, pre/then and post/then would be calculated. These would be compared against a control group, where all correlations should be approximately equal.

# CHAPTER 4

# RESPONSE EFFECTS AND ARTIFACT IN COMPUTERIZED DATA COLLECTION

## 4.1 Field Benefits in Computer Interviewing

Computers have been used to aid in data collection in a variety of disciplines. The fields of medicine and psychology have the distinction of the longest history of use. In 1966 Slack et al. used a LINC mini computer and an interactive program to ask patients about allergies. The program made use of branching patterns, asking more detailed questions if a positive response was received for a particular allergy. The computer is also a useful interviewing tool in the psychiatric field, particularly if respondents feel more comfortable answering sensitive questions without others present. (Millstein, 1987; Erdman, Klien and Greist, 1985; Barron, 1987). Researchers have shown that, for the most part, computerized administration of personality measures tests can replicate results of traditional pen a paper administration (Resmovic, 1977; Katz & Dalby, 1981a; Hurrell & Lombardo, 1984; Russell, 1986).

Business applications have surfaced only recently, a result most likely of the recent personal computer (PC) explosion. The diskette survey method was an innovation of an IBM PC user group trying to increase response rates to membership questionnaires. The underlying concept is to use a floppy diskette, rather than paper as the medium for communication. The initial version was simply a print file, respondents were to print a hard-copy which was in turn to be completed and returned. This evolved into an interactive interviewing program, where respondents answer questions "on-line" and simply return the diskette. In a study of 308 respondents, researchers found similar response rates (about 70%) for diskette and traditional mail surveys. Of the respondents given an option of response method, about half utilized the diskette survey (Greenwood, 1987).

Electronic mail systems have been used to increase response rates over traditional mail surveys (Sproull, 1986). Increased response rates are of interest as electronic mail

surveys cannot take advantage of status cues to legitimacy (such as a letter of introduction with university letterhead, or a memo with a CEO signature) that are thought to increase response rates (Dillman, 1978). Electronic mail also can be used in the task of software evaluation. Rushinek et al. (1983) found the following problems common to traditional manual evaluation of systems and software:

1) incomplete forms,
2) illegible forms,
3) biased sample of users,
4) low response rate,
5) long elapsed time between program execution and evaluation,
6) data coding errors,
7) key punch or key operation errors,
8) high error rate in data transcription,
9) user resentment to excessive paperwork.

Rushinek goes on to suggest that all of these evaluation problems could be eliminated with properly executed electronic surveys.

Various marketing companies use unattended terminals to gather survey data. In one instance UPI ASK was able to collect 10,000 responses in a three week period. Traditional data collection techniques would only have gathered 500 responses in that time (Marketing News, 1985b). PC and/or terminals have been used to collect data at trade shows, professional conferences, or events where populations of interest gather. A French marketing group utilizing a terminal at the Rights of Man exhibit found: respondent interest, no psychological barriers, no need for supervision, and few abandoned questionnaires (Franzkowiak, 1985). The Hawaiian Department of Transportation used an unattended terminal with touch screen technology to solicit feedback on the public's satisfaction with the airport's facilities (Okimoto, 1988).

Most studies report that respondents either enjoy or, at worst, are neutral to the computer interview. In a mall intercept interview using personal computers to perform

computerized interviews Liefeld (1987a) experienced almost 100% cooperation. Newsted (1985) compared preference of computer verses personal interview technique. He found that computer users preferred to use the computer for the interview while manual interview respondents were at least neutral towards being interviewed by the computer. In a study of 50 medical patients, 12 preferred physician histories, 18 preferred computer interview, and 20 were neutral. (Slack et al., 1966). In administering the psychological assessment of young children, Katz and Dalby (1981b) found that initially favorable perceptions of computers increased for those children who underwent the computerized form of the assessment. In a review of the psychological and medical uses of computer interviewing, Erdman et al. (1985) found that about 50% of respondents preferred interviewing with a computer over a doctor for: medical histories, suicide risk assessment and structured diagnostic interviews.

The future holds additional possibilities. The pre-employment screening interview, for example, has been automated (Eckhouse, 1986). Prospective employees respond to a questionnaire (administered by a computer using voice synthesis) using the keys on a touch-tone telephone. The company providing this service claims that applicants will be more honest in responding to computerized interviews than they would talking to a human. Hiltz (1979) suggested the creation of a new technique of data collection, "computerized conferencing systems". CCS's could be used to solicit opinions from 'experts' who may already have terminals, for conversational explorations of the attitudes of a small group (taking advantage of the interactive nature of computer interviewing), or to solicit a permanent panel of respondents who are chosen as a statistically representative sample of a more general population. Advantages include being able to perform Delphi techniques, validation of responses, inexpensive follow-up contact, branching questionnaires and on-line help for ambiguous questions.

Researchers should also be looking at computer monitoring for data collection, using unique computer abilities such as clocking response time (Rice, 1984). A marketing research firm utilizing computerized interviewing software uses the completion time as a validity check, citing an instance where a series of twelve minute interviews (shorter than average) were tracked to a dishonest staff member (Bahner, 1987). Also,

longer response times to individual questions could be an indication that the wording is unclear, providing an expedient measure in the pre-testing of questionnaires. Advances in technology will allow researchers to track actual behavioral information, circumventing the traditional survey process entirely. For example, Information Resources International's BehaviorScan works as follows. Consumers are given scanner readable cards with their personal demographic and psychographic profile. As they make purchases, which are checked out by optical scanners, the universal product code and consumer information is automatically captured, enabling the researcher to track purchases directly from the checkout stand at the supermarket (Gorn, 1988).

Computer monitoring is an alternative method of data collection, particularly valuable for organizations collecting internal data about the use of their information systems. Bernard et al. (1981) did an experiment in a computer conferencing system asking users typical questions about what activities they performed on the system, and compared this data to communication behavior monitored by the computer. The comparison showed that people have poor memories, often forgetting and inventing messages. Computer monitoring studies in the past have involved information retrieval systems; measuring transactional data, user log on time, and tracking of commands used, and computer based communications systems; measuring error rates and text content to generate aggregate ratio measures (Rice, 1984). Some of the anticipated issues include managing the large amount of data, privacy, and monitoring longitudinal data.

## 4.2 A Causal Model For Response Effects in Computer Interviewing

As was done for the presentation of bias in traditional data collection methods, an attempt will be made to fit all findings into a framework of potential biasing sources, in this case the framework is for response effects specific to computerized interviewing. The model proposed by Sudman and Bradburn (1974), which includes interviewer, respondent, and common task, serves as the basis for the framework. The task, however, is further subdivided into: question content, presentation, human factors, response task, amount of feedback, and artifact. As well, biases caused by the

involvement of a computer in the interview process are treated as effects created by the "non-human" nature of the interviewer, rather than as a function of the task itself.

The framework also incorporates the systems concept of input-process-output. In the diagram in Figure 1 these are labeled (1) through (3). The input into the model is the question that is being asked. Semantics of the Question (1) is defined as the intent of the question, as understood and phrased by the interviewer. The Survey Process (2) is the exchange between interviewer and respondent. The process is social interaction with the purpose of completing the data collection task. This process then results in an output, the respondent's answers. Results (3) are the combination of true result, sampling bias effects and the sum of all response effects.

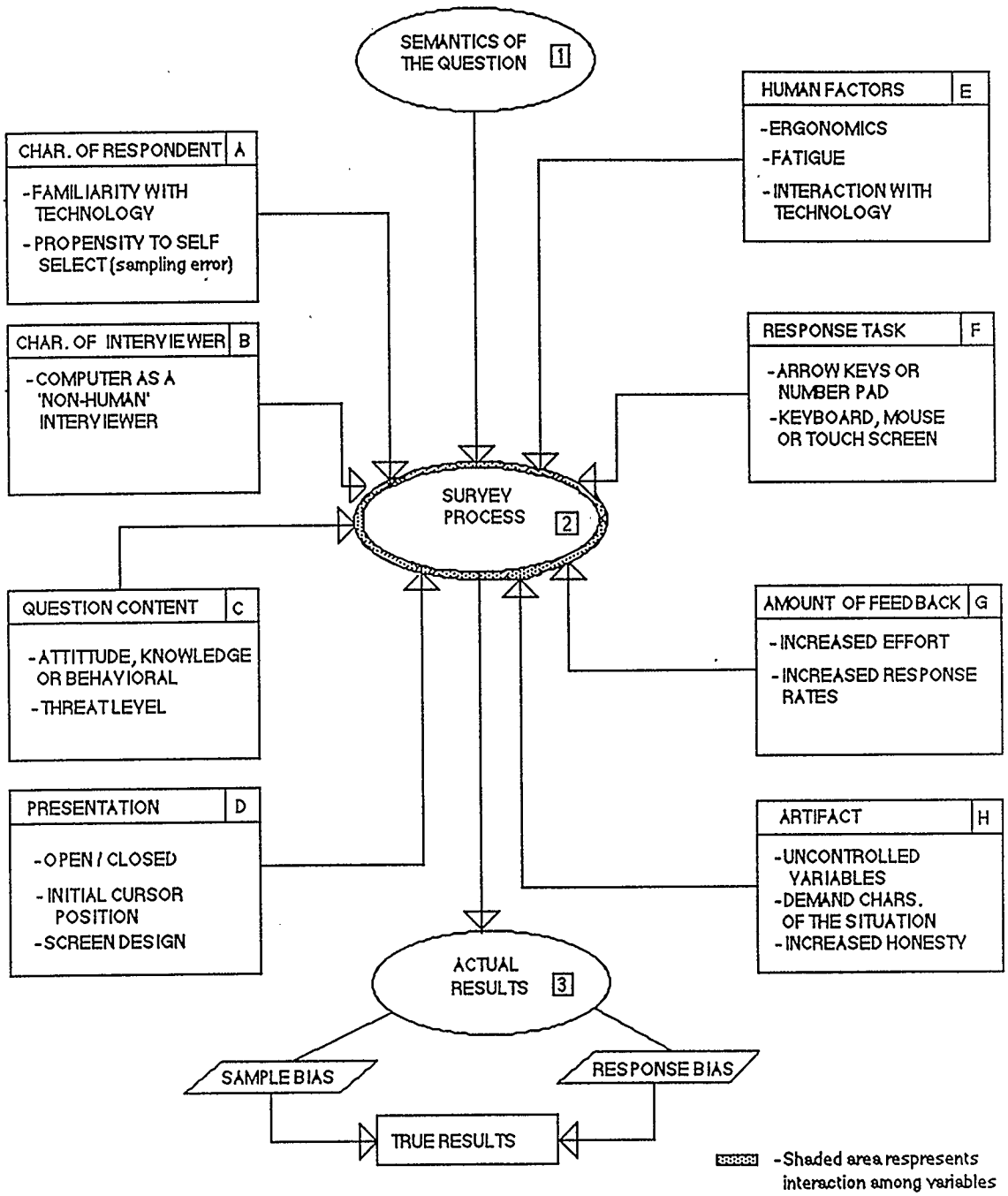
Variables contributing to response effects are labeled (A) through (H). The characteristics of the respondent (A) and the computer as a 'non-human' interviewer (B) are the first two categories. As the task variable creates the greatest source of bias (Sudman and Bradburn, 1974) it is further divided into: question content (C), presentation (D), human factors (E), response task (F), amount of feedback (G), and artifact (H). Each of the biasing variables are briefly defined below and will be discussed in detail later in the paper. Interaction between factors is represented by the shaded area surrounding the Survey Process (2).

<u>(A) Characteristics of the Respondent.</u> A bias produced by the respondent before the interview and of their own accord, not due to other factors. This variable explains individual differences; results consistent across an entire sample would be attributed to another variable rather than the respondent. This includes familiarity with technology, need for approval, and propensity to volunteer.

<u>(B) Characteristics of the Interviewer.</u> Biases produced by the interviewer's involvement in the process. This would include amount of training, expectations, and actions taken to increase response rate and cooperation. The computer can be treated as a 'non-human' interviewer.

**Figure 1.**

A CONCEPTUAL FRAMEWORK FOR
BIAS IN COMPUTERIZED DATA COLLECTION



(C) Question Content. Biases produced by the content of the question. Question content can be grouped regarding either knowledge, attitude, or behavior. Also included in this category are the saliency and threat level of the question.

(D) Presentation. Bias produced by variation in the presentation of a question. This includes variation in length of scales, question wording, order of questions and type of question. Questions can be either closed or open-ended. Closed question types include; pick-1, multiple response (pick-n), ranking and scale response questions. For computerized questionnaires presentation bias also includes initial cursor positioning and screen formats.

(E) Human Factors. Biases produced by variables pertaining to man-machine interaction. This includes physical ergonomic measurements and cognitive thought processes. A 'cognitive' human factor is the extent to which the technology forces the respondent to think or process information in a manner which is different than how they would normally process that information. An example of this would be in completing a ranking. A respondent may usually make a first run through the ranking and make light marks in pencil, then come back and make a final ranking. Using an electronic medium this strategy would generally not be possible; the user would have to adapt his approach to suit the medium.

(F) Response Task. Biases produced by variation in the response task. This can include keyboard entry, verses mouse entry, verses touch screen entry, verses pen and paper completion. For keyboard entry, respondents can user arrows to move the cursor keys, or number keys to indicate placement on a scale.

(G) Amount of Feedback. Bias created by the amount and type of feedback. This variable is concerned specifically with the bias of reduced response rates. The suggestion is that, just as increased personal contact increases response rates, the interactive nature of the computerized questionnaires will have the same effect.

(H) Artifact. Those variables known to bias response but whose nature and relationship to other variables is not completely understood. This includes situational demands and uncontrolled variables.

Interaction among these variables can also be expected. As indicated by the shaded area around the Survey Process (Box 2) there may be more than one factor responsible for an identified bias. As well, interaction effects among factors are not specifically defined in the model. Take, for example, the situation where a white interviewer is questioning a black respondent. Sudman and Bradburn (1974) found that no bias is created unless questions salient to these differences are asked. Therefore, for questions regarding consumer products no bias would be expected and the simple effects of each variable (characteristics of respondent and characteristics of interviewer) would be zero. However, if question content involved racial attitudes, then one would expect a biased response, and accordingly a substantial interaction effect among question content, respondent characteristics and interviewer characteristics. Interaction may also reduce the simple effects of each variable. Threatening questions, for example, have higher response rates if asked in an open-ended format (Sudman and Bradburn, 1979). In this case the interaction between effects of high threat level (question content) and open-ended question (presentation) combine to reduce the total bias. It is a complex task to identify and test for all interaction effects. For the time being it is sufficient to note that they may occur, as the model in Figure 1 is used to organize the review of studies examining response effects in a computerized data collection setting.

### 4.2.1 Characteristics of the Respondent (A)

One factor which might bias results is respondent's familiarity with the technology. It might be expected that some factors such as lack of knowledge or general "computer anxiety" (created by previous negative experiences with computers) may hinder cooperation on the part of the respondent. In fact, most researchers have reported high cooperation rates and favorable responses towards participation in the process

(Liefeld, 1986; Millstein, 1987; Erdman et al., 1985, Kiesler and Sproull, 1986). Higgins et al. (1987), surveying a PC user group, found higher word counts to open-ended questions when comparing diskette to pen and paper interviews. In addition, word counts were not correlated to individuals' rating of their typing abilities. Further research is required to determine if these findings can be generally applied to a population frame less familiar with the technology.

In the education field researchers have looked at the influence that lack of previous computer experience may have on students' scores on computer administered test (Eaves & Smith, 1986; Lee, 1986). While one would not expect comparable levels of anxiety to be created in completing a computerized questionnaire, the findings are worth noting. Eaves and Smith (1986) selected 96 students at random, and placed them in categories relative to their previous experience. The three groups: subjects with no experience, some experience, and extensive experience were then split between a computer and pen and paper examination. Analysis of variance of the six group scores suggested that there was no difference between the two types of test among the three levels of experience. Jo-Ann Lee (1986), on the other hand used a pen and paper test to determine math abilities of one group of students. This score was used to adjust scores in a complete high school math test administered by computer. Of the three groups: experience with computer games only, low computer experience and high computer experience, the games only group, after adjusting for math ability, performed significantly worse on the computerized test. The high experience and low experience scores were very similar, suggesting that even a few instances of "hands-on" experience is sufficient to eliminate any handicap that may exist. These results suggest the possibility that the relationship between previous computer experience and likelihood of demonstrating a bias is a dichotomous rather than linear function.

It does not appear that the computer will serve as a catalyst to increase bias demonstrated to be a function of respondent personality characteristics. Millstein (1987) suggests that individual differences do not interact with mode of administration. In a study asking sensitive questions to a group of adolescent girls by computer and human interviewer, no interaction with need for approval was found.

Characteristics of the respondent will become an important variable if the survey method utilizes a self-selection sampling procedure. Persons of a specific demographic or psychographic profile may be more likely to volunteer, resulting in an unrepresentative sample, which in turn limits the researcher's ability to generalize findings to the entire sampling frame. For example, one would expect that an over-representation of young males, who have a high propensity to use computers and play video games, may be obtained if unattended terminals are used to collect data. Newsted (1985) used a national parks information centre as a setting to compare the data collected from an unattended terminal to periodic one hour samples. He found the median age of the computer respondent group to be younger than the group surveyed in the manual sample. As well the computer group demonstrated a bi-modal peak for education level at grade 8 and university categories. An unattended interactive kiosk at Honolulu Airport was successful in collecting 5000 responses in a one month period. It did, however, collect a unrepresentative sample. When compared to demographics of the Hawaii Visitor Bureau's survey, significantly greater proportions of persons under 19 years of age and smaller proportions of persons over 50 years of age responded to the computerized questionnaire (Okimoto, 1988).

Statistics Canada has unattended computer interviewing kiosks in place at airports in Calgary, Ottawa, Montreal and Toronto. Demographics of respondents are accumulated for each airport. The results of this survey, along with another findings from another study using random sampling techniques (Canadian Facts Airport Audience Study) are shown in Table 1. A comparison of the age distributions of the samples collected reveals that the self-selection computer interviewed group has an under-representation of the 50-65 year old age group and an over-representation of the 13-18 year old age group. Although the unattended kiosks do seem to sample an accurate and consistent proportion of the other two age groups across all airports, including persons over 65 years of age, a more detailed breakdown of the 20-49 group may show more inconsistencies between the two samples.

- TABLE 1 -

Ability of an Unattended Kiosk to Obtain A Representative Sample at Canadian Airports

| *Canada Facts Random Sample* | | *Statistics Canada Unattended Kiosks* | | | | |
|---|---|---|---|---|---|---|
| | | | Tor. | Calg. | Otta. | Mont. |
| Age Group | % | Age Group * | % | % | % | % |
| 15-19 | 3.2 | 13-18 | 14.9 | 9.2 | 10.9 | 15.5 |
| 20-49 | 72.0 | 19-49 | 72.2 | 75.4 | 72.4 | 72.4 |
| 50-64 | 18.3 | 50-64 | 7.2 | 9.3 | 10.7 | 6.2 |
| 65+ | 6.5 | 65+ | 5.6 | 6.1 | 6.1 | 5.9 |

Source: Canadian Facts Airport Audience Study 1988 & (Niman, 1988).

* Note: Canadian Facts interviewed only respondents age 15 or older. In order to compare the two distributions the 6-12 age group of the Statistics Canada respondents was removed. As well, the youngest age group categories do not match precisely.

### 4.2.2 Interviewer Effects - The Computer as an Interviewer (B)

Locander, Sudman and Bradburn (1976) suggested that social contact, or human presence will determine the amount of social desirability in responses. In their framework, Sudman and Bradburn (1974) treated social contact as a function of the task, specifically the data collection method. However, as this study is dealing with only one data collection method, namely computerized questionnaires, it is necessary to modify this framework and treat the computer as an interviewer. In this respect the characteristic of the interviewer, human or otherwise, rather than the task, are used to determine the amount of social contact. In using this framework it is important to understand that interviewing with a computer does not necessarily remove all social contact. In fact the reverse can be true. Okimoto found that a line-up waiting to respond to an unattended kiosk created an audience, which influenced some "apparently unattached men to emphatically press the $75,000+ income button when women were present and watching" (1988, p.359).

The field of psychology has been particularly interested in the validity of psychological tests administered by computer. For the most part the findings suggest that

no adverse effects are present. Greist et al. (1987) found the computer to be a viable alternative for administration of the Diagnostic Interview Schedule. Katz and Dalby (1981a) found high test-retest correlations between pen and paper and computer administration of the Eysenck Personality Inventory. In a study that used 80 students in a repeated measures methodology (computer and booklet administration), Hurrell and Lombardo (1984) found no significant difference between paper and computer administration of Form A of the 16PF test. Russell, Peace and Mellsop (1986) used a four group study to examine the effects that computerized administration of the MMPI had on scores. They found the reliability of the computer then manual administration and manual then computer groups, was equal to that of the computer-computer and manual-manual administration groups. As well, there were no difference in drop out rates. Resmovic (1977) found that method of presentation did not significantly effect mean scores on personality inventories, only that as a result of the counter balanced design, performing the test increased variance in the second test. In a study of 38 respondents undergoing a computer-conducted intake, respondents showed no significant differences in expectations, comfort or rate of return (Barron, 1987).

One study finding differences attributed to method of administration was performed by Canoune and Leyhe (1985). It is important to note that they compared computer verses *human* administration of Gordons 1976 Survey of Interpersonal Values, which include: support, conformity, recognition, independence, benevolence, and leadership. The computer group was higher on the value of recognition (mean of 12 vs 11), and lower on conformity (10.5 vs 12) and benevolence values (17 vs 18). These findings suggest that the human interview group answered in a socially desirable manner. In interpreting these findings, however, keep in mind the fact that differences in the opposite direction were not found for the other three variables and the absolute size of the bias was a difference of about one unit on approximately a twenty point scale.

Lucas et al. (1977) found that the computer was no different than two psychologists in analyzing *indicants* of drinking behavior, however, the computer interview increased reports of present and total *consumption* by about one third. These findings were confirmed by Duffy and Waterton (1984). Other studies in drug and

alcohol use, however, have not supported this hypothesis. Erdman, Klien and Greist (1983) found that pen and paper and computer questionnaires yielded similar results for a group of high school students with the exception of greater reported frequencies of drug and alcohol use in the pen and paper mode. The difficulty with this group is the possibility that the students may be prone to both under-reporting *and* exaggeration. Skinner & Allen (1983) interviewed three groups of "in-house" alcoholic patients using self completion, personal interview and computerized questionnaire methods. They found no difference between methods for an alcoholism screening test and alcohol, drug and tobacco use history. Older persons, however, did prefer the self completion method. They suggests that a potential explanation for the lack of a social desirability bias is that "in-house" patients may be more honest in reporting drinking habits, as a clinical treatment environment eliminates the social pressures for patients to report a perceived norm of drinking levels.

As the amount of social contact has been shown influence the likelihood of social desirability in responses, one would expect personal interviews to elicit the highest amount of bias for questions with the potential for socially desirable answers. In addition, as questionnaires administered by both computer and pen and paper have the same amount of social contact, they would be expected to display the *same amount* of social desirability in answers. Some findings have suggested that this is not the case, that responses to computerized questionnaires display *less* social desirability than comparable pen and paper completion. Researcher have attributed this to the fact that the computer provides a more anonymous setting, as responses enter into a black box of magnetic bits, rather than on a single questionnaire that is returned directly to the interviewer.

Evan and Miller (1969) were the first to claim that the computer would decrease social desirability in responses. In a study using two groups of 30 students they found the computer group to state more positive orientation to religious values. Given the relatively secular environment of a university campus, they interpreted this to indicate less social desirability in responses. In addition, they reported a small tendency to be more honest based on responses to questions regarding Manifest Anxiety (MMPI), Lie Scale (MMPI), and a dichotomized Scole Scale. These findings, however, are not nearly as

conclusive. The proportion of favorable answers to this combined scale was only about five percent higher for the computer group than the pen and paper group. More importantly, differences for individual scales on their own merit (i.e. Lie Scale alone) were not all statistically significant, in fact the *largest* difference between the two groups was for a group of factual and non-judgmental questions designed by the authors to be less likely to elicit socially desirable answers. As the authors utilized a small sample it is entirely possible that difference between groups (about 1.5 of 40 yes/no questions per respondent were answered differently) could have been entirely due to *actual* differences in the personality makeup of the two groups, and not due to method of completion at all. Finally, the authors proposition that differences would be due to increased anonymity of the computer interview was not supported by post interview probing.

Kiesler and Sproull (1986) suggests that the engaging nature of the computer should result in respondents giving more answers, longer answers, and talking more about themselves. As well, the lack of social context will result in less acquiescence, less middle responses, and reporting of more socially undesirable traits. Their findings support this assertion to some degree. While they did not find any difference in tendency to use extreme end points in answering scale questions; in comparison to oral or hard copy, answers to an electronic mail questionnaire tended to be less socially desirable and included more self disclosure. This study, however, used only five questions of the Crowne-Marlow social desirability scale. Liefeld (1986) found the computer group to exhibit *more* social desirability with respect to media habits for the pen and paper completion group, despite respondents stating that the computer interview was less judgmental than other methods.

Certainly findings are not conclusive, and at this point explanations can be considered speculative at best. Kiesler and Sproull's (1986) theory relating to a lack of social context may have some validity. Alternatively, differences between these methods can possibly be explained by a 'big brother' effect, where the nature of the computer (for example, the ability to reject invalid responses) or respondents perceptions of the "power" of computers could encourage respondents to be more honest. Two studies of this nature have been performed in academic settings (Allen, 1987; Sloan, 1988).

However, as the respondents may have suspected that their answers were being verified, these are presented as a function of an artifact. It is also possible that honesty towards computerized questionnaires is a function of cultural perceptions of the computer. For example, Europeans have shown more concern than North America about the possible invasion of privacy associated with large databases. Perhaps Nederhof (1987), in an article which reviewed the same studies as are presented here, is influenced by his perceptions of 'dis-trustful' European attitudes towards computers when he proposes that computer interview responses may include *less* honesty than answers to pen and paper questionnaires.

### 4.2.3 Question Content (C)

When analyzing question by content, questions can be categorized as about knowledge, attitudes, and behaviors. Liefeld (1988) found no task effect (i.e. collection method) on behavioral data and demographic questions. Questions regarding knowledge tend to elicit an interesting response effect. On multiple response (pick-n) type questions Liefeld (1986) found more responses and more incorrect responses than for other collection methods. He labelled this effect "video game syndrome". This phenomenon, where the computer appears to encourage more tries is also supported to some extent by findings of Slacks (1966). In a computer interview regarding allergies the only difference between computer responses and hospital records were false positives on the computer interview responses. As well, Newsted (1985) found better responses to knowledge type questions when comparing computer respondents to personal interviews. Computer respondents, despite seeing fewer exhibits, answered more questions correctly. Unfortunately, as the computer group was allowed to self-select while the pen and paper sample was selected randomly, differences could be due to the fact that respondents who were more knowledgeable were more likely to approach a computerized questionnaire.

Liefeld (1987b) also received more positive responses to attitude and judgmental questions from the computer interviewed group. Millstein (1987) found the only bias created by computer interviewing of adolescent girls was for questions regarding positive

feelings of their experiences, while no bias was demonstrated for questions regarding substance use and symptomatology. As Sudman and Bradburn (1974) found attitudinal questions to be particularly susceptible to bias in traditional data collection, more work is required to determine if computer administration has any additional effects on attitudinal content questions.

With respect to threatening or sensitive questions there is some evidence that the computer is better than humans at eliciting responses in certain situations. Carr (1983) used the computer to take personal psychiatric histories of patients. Overall agreement of the computer history with hospital records was 90%. In fact, the computer interview found additional pertinent information that staff had not previously discerned including: concern over masturbation (30% of respondents), criminal records (26%), reports of blackouts after drinking (23%), and impotency (20%). These facts had not been discovered in four weeks of care, despite patients being interviewed by a registrar, spending several hours with consultants and interviews with a social worker and an orderly. However, the methodology used can not ascertain whether or not these questions were asked by humans with patients choosing not to answer, or asked only by the computer.

On the other hand, Millstein (1987) used the computer and humans to interview two groups of 30 girls on sexual behavior and substance use, gynecologic symptomatology, and affective states. She found no difference among the groups for behavior, substance use and symptomatology reporting. Outside of the clinical setting Higgins et al. (1987) found no bias in responses to sensitive questions when comparing responses to diskette surveys and traditional mail surveys. While the computer has the distinct advantage of performing interactive interviews with no human contact, there is no clear evidence that the computer will not have the same problems that other data collection methods encountered in obtaining truthful answers to sensitive and threatening questions. The benefits exhibited are more in the realm of greater comfort and enjoyment than in improved reporting.

### 4.2.4 Presentation (D)

While bias elicited by poor question wording, order of questions, and use of open verses closed questions can still happen in computerized questionnaires, a more relevant concern is the bias that screen presentation may create. Some formats of questions do not appear to be affected by screen presentation. Liefeld (1988) found no difference for yes/no or pick one questions between mode of interview when comparing self completion, personal and computer interviews. These results, along with findings in the field of psychology regarding the suitability of computerized administration of psychological tests (Greist et al.,1987; Russell et al., 1986; Resmovic, 1977) would suggest that computerized presentation is suitable for questions in a dichotomous or simple multiple choice format.

To date research regarding interval data has addressed presentation in terms of response patterns to scale questions. Groves (1979) suggests that if scales do not use either five or ten points, large differences in response patterns will emerge when comparing telephone interviews to personal completion. Indeed, Liefeld (1988) suggests that results from scale questions cannot be compared with other methods of data collection.

Initial cursor position, for example, may affect response patterns, although Liefeld (1987b) suggests that this is not the case. In a mall intercept study with 800 respondents he found the computer interviewed group to exhibit higher means across a variety of scale questions. This study utilized a scale protocol where the cursor was initially placed at the middle of the scale and for the following two questions was placed at the same point as the respondent's mark on the previous question. In analyzing effects he found no correlation in successive pairs suggesting that the initial placement of the cursor does not affect respondents final placement. Despite the fact that initial placement in the middle of the scale (verses a normal perceptual starting point on the left) could lead to a higher mean, findings regarding a lack of correlation between successive pairs lead him to attribute the increased mean to better "mood" reported by computer respondents. The entire area of presentation is certainly an area that requires detailed attention in future

research. Little is known about how the use of graphics, placement of error messages, and level of on-line help affect users' responses.

### 4.2.5 Human Factors (E)

Much research has been performed on the man-machine interface component. The question is two-fold, the suitability of video media presentation for completion of the task, and the ability of the user to interact with the technology. With regard to the first component, two tasks which have been researched are proof-reading for errors and reading text for comprehension. In reviewing the research Switchenko (1984) states that there is no correlation between proof-reading errors, reading scores and method of presentations, and that proof-reading from a Cathode Ray Terminal (CRT) appears to take longer than proofreading from paper.

The results on reading rates are mixed. Wright and Lickorish (1984) investigated the requirements of referees in working with article submissions in an electronic form. Volunteer referees were given two submissions, one in electronic form and the other paper, with half of the referees receiving each article in each form. Referees were deemed to require legibility, movement through text, annotation and communication. Wright and Lickorish found that as a group the referees were taking 20% longer to deal with the screen as compared to the paper version. There was also a non-significant tendency for the paper article to be given a poorer rating.

In a study which compared reading (for a two hour period) from a book to reading text from a television screen Muter et al. (1982) found that participants read 28.5% slower from the screen. The study did not control such variables as mean words per page (120 per screen verses 400 per page), number of characters per line, and did not control for such variables as individual differences in reading abilities which can exist between subjects in different groups. Using personal computers in a mall intercept Liefeld (1986) found some older respondents could not read the screen due to poor eyesight and as a result experienced lower completion rates for the computer interview than for self completion and personal interviews.

Cushman (1986) compared reading comprehension for various forms of presentation, microfiche (negative and positive image), CRT (negative and positive image), and printed form. Respondents were periodically tested for fatigue as well as being monitored for speed and tested at the end of the reading for comprehension. For the microfiche, black letters on white matte board was no more fatiguing than paper; for CRT, light characters on a dark background was no more fatiguing than paper. However, for both mediums the opposite image was found to be more fatiguing. Reading performance (rate x comprehension) was similar for all presentation methods. This research would suggest that when creating screen formats in extended computer interviews that light characters on dark backgrounds should be used to reduce fatigue.

Askwal (1985) had sixteen psychology students search and integrate fictional paragraphs from paper and CRT formats. The students were measured for reading speed, asked to make logical inferences from the reading as well as rate the difficulty of the task. She found no difference between reading times or difficulty ratings of the task for CRT and paper presentation mediums. There was, however, a difference in search time and strategies. Search times for both mediums increased the same amount as search demands were increased. In the paper medium subjects searched almost twice as much information as with the CRT, yet in the computerized situation search times were almost twice as long. As the group was comprised of generally novice computer users, this difference may be explained by unfamiliarity with the technology which prevented the subjects from using their normal search strategy. For example if users forget or are unsure how to use the 'back-page' key they would not be able to use a 'zigzag' type of search and would have to adapt their search strategy accordingly.

Switchenko (1984) also supports findings that there is no inherent disadvantage to reading text from a screen rather than in a printed form. Participants were given two readings of the same material for both a difficult and easy article. She found reading time for the difficult article was longer than for the easy article on both days, yet found no difference in reading times between the two presentation modes.

One interesting finding is a tendency to longer completion times for computer interview data. In comparing completion times for electronic mail and personal interview

processes computer interviewed respondents spent an average of 74 minutes completing the survey verses one hour for personal interview (Sproull,1986). Liefeld (1986) and Lucas et al. (1977) also found computer interviews to take longer than self completion and personal interviews. The research suggests that the longer completion time may be mostly attributed to learning how to interact with the new technology (and increased requirement for instructions on how to respond) rather than an ergonomic factor associated with using the CRT medium. Liefeld (1986) also attributes the extended time to less pressure to complete the interview and more thought being put into answers. It is possible then, that longer completion times are explained by respondent motivation due to collection method, rather than as a human factor limitation.

Stemming from a review of studies in the area of human factors, Hansen and Hass (1988) suggest that four primary and three secondary factors can be used to explain differences in performance with associated mediums. They suggest that page size, legibility, responsiveness (of the system), and tangibility (the extent to which the state of the system appears to be modifiable via physical apparatus) are influential in determining user performance. These primary factors combine to produce three secondary factors: sense of directiveness (degree that changes on the screen are a direct result of users actions); sense of engagement (feeling that the system is holding an interesting conversation with the user); sense of text (grasp of structural and semantic arrangement of the text). This framework was used to explain lower spatial recall scores for personal computer verses pen and paper completion tasks, and longer mean time to locate text on computer verses pen and paper documents (Hansen and Hass, 1988).

From this collection of research it appears that if presentation variables are controlled there is no inherent disadvantage to reading data in an electronic form (allowing that respondents bring their reading glasses to the interview). The differences tend to appear in those studies which do not rigidly control variables to absolutely replicate the pen and paper task. From a strictly controlled environment it can be argued that the use of a CRT screen does not affect the information presentation component of the communication process.

There does appear to be a human factor bias in processing the data. Limiting factors appear in task completion when the subject attempts to use the technology in a cognitive method unlike that with which they would approach the task using pen and paper. Both the referees and subjects searching data were required to interact with the computer, while in other studies the interaction would be only a simple next page keystroke. At this stage a tentative methodology rule of thumb to reduce human factors response effects can be formulated. In order to reduce response effects due to human factor variables researchers should limit the knowledge of technology required and replicate the pen and paper cognitive process as much as possible. Perhaps the next generation, brought up with computers in the classroom will develop a different cognitive process and not display the same response effects.

#### 4.2.6 Response Task (F)

In addition to 'cognitive' human factors the characteristics of the response task can also contribute towards biasing responses. In some cases the response task may not affect responses. In a study of a PC users group, Higgins et al. (1986) found that the requirement for keyboard entry did not limit response content for open-ended questions. On the other hand, one response effect which may be a function of the entry task is the range of responses in answers to scale questions. In comparing computer interviewed respondents against self completion and personal interview Liefeld (1987b) found that the standard deviation of responses for computer interview respondents was lower than the standard deviation of responses for self completion and personal interview respondents on 20 of 27 questions. In this study the cursor was placed in the centre of a ten point scale and respondents used arrow keys to move the cursor.

This is in contrast to the findings of O'Brien and Dugdale (1978). In comparing computer verses personal interviews of a sample of women in Great Britain they found a wider range of responses to scaled questions. Allen (1987) found that 6 out of 12 items had a significantly greater variation of responses when comparing computer interviews to self completion of machine readable questionnaires. The effect of the computer interview, which utilized software requiring entry using number keys, was to increase the

standard deviation by approximately 25% on a ten point scale. Sproull (1986) found an increase in the number of extreme responses (electronic verses paper completion); out of 48 hypothetical questions 35 had more extreme responses for electronic completion. Yet in a further study with Kiesler (1986) no differences in extremity of response were found. These differences may be explained by the variation in hardware and software protocol used, in particular whether cursor movement or number selection is required.

Findings for response effects have an impact on findings in other categories as variations in response patterns are often used as measures for other response effects such as acquiescence and social desirability. Yet the research has provided no clear indication of the importance of the protocol of the entry task (which can vary depending on the software used) as source of response effects. If the entry task does contribute to bias, then this variable must be taken into consideration when attempting to measure the larger phenomenon, "completion by computer". As well, additional research is required regarding the potential effects of light pen, touch screen, and mouse response mechanisms.

### 4.2.7 Amount of Feedback (G)

This variable is concerned with response effects in the form of non-response. The basis for amount of feedback as a biasing variable is derived from Sudman and Bradburn's (1979) proposition that cooperation and response rate will be increased with subject-interviewer interaction. This assertion is broadly supported in the difference in average response rates among personal, telephone and mail survey techniques (Sproull, 1986). The Total Design Method (Dillman, 1978) for example, can be employed to consistently obtain mail response rates of 75%. While Dillman attributes participation to attention to administrative details rather than the nature of the collection medium it can be argued that such actions as follow up phone calls, timing of mailings and other administrative details are increasing "interaction" in the interview. In fact, Sproull (1986) suggests that researchers may find that they are able to increase response rates with no substantial increase in costs by using electronic surveys in conjunction with other data collection methods. Regardless of the proposed *source* of variance in response rates, all

of the "feasibility" studies regarding computerized data collection have looked at its ability to increase response rates.

For electronic mail or diskette mailing techniques, their interactive nature and ease of immediate response may contribute towards higher response rates, when compared against traditional mail surveys. Ninety percent of participants in an electronic mail survey said that they would participate in another survey (Sproull, 1986). Higgins et al. (1987) measured cooperation levels by asking respondents if they had returned the last questionnaire mailed to them. Of those participants completing the diskette questionnaire, 39% of respondents had not replied to previous questionnaire while only 24% of the control group (paper respondent) had not replied. Sproull (1986) found a response rate of 73% for electronic mail respondents compared to 87 % for personal interviews. These figures compare to an internal mail questionnaire (circulated in the same company at the same time period) response rate of 55%. Overall these studies show an increased response rate for diskette and electronic verses traditional mail surveys.

The research is not overwhelmingly convincing. In other studies Kiesler and Sproull (1986) found only a 67% response rate for electronic surveys verses 75% for paper survey. Higgins et al. (1987) found virtually equal response rates when comparing a traditional mail survey compared to a diskette mailing. This figure may have been biased by the fact that 60% of non-respondents did not have access to 100% IBM compatible personal computers and as such were unable to use the program.

In comparing results from a mail survey and a diskette survey Higgins, et al. (1987) found that mean diskette reply time was lower, and that diskette respondents gave more words and more points in answering open-ended questions. These factors were combined to indicate higher respondent rapport. This group, PC users, was by no means a random sample and it is not known if the findings can be generalized to other populations.

The research tends to show that using electronic mail and diskettes to conduct surveys may increase response rates over traditional mail surveys. While the term feedback traditionally involved human contact, it is suggested that the interactive nature of the computerized questionnaire (including abilities to branch based on responses and to

retrieve data from previous answers) constitutes increased feedback in the same manner that a personal interview will have more feedback than a pen and paper interview. One cannot, however, emphatically say that increased response rates are due solely to the interactive nature of computerized questionnaires. It is also quite possible that higher response rate may only be due to a novelty effect of receiving a survey electronically or on diskette. To determine the correct causal relationship it would be necessary to perform a study which varied the "interactiveness" of the questionnaire and, for example, compare response rates in organizations that are new to electronic mail surveys against those of an organization that has become used to communicating through the electronic medium.

### 4.2.8 Artifact (H)

As was stated earlier, artifacts tend to be unknown variables which influence the results of an experiment or study. However, as more is understood about an artifact, and if theoretical links can be made to the collection method, an artifact can make the transition to response effect and as such would fall under the most appropriate preceding category. The relationship between respondent apprehension and increased honesty is one such example. Respondent apprehension may be stimulated by the use of computers to complete the interview, which in turn, may lead to increased honesty in answers. However, in testing for this effect it would be quite possible to manifest a situation where respondents could suspect that the researcher has data to verify answers and that this artifact, not computer completion, would be the cause of increased honesty. Two studies in campus settings have shown minimal support for this idea. In a preliminary study Sloan (1988) found a non-significant tendency to be more honest in reporting grade point averages for those students using computer interviews compared to self completion. Allen (1987) found respondents completing a machine readable questionnaire over-reported their GPA's, while those completing the same questionnaire on-line did not. Unfortunately, the on-line group were also more likely to incorrectly state the semester of their first term.

In using computerized questionnaires to evaluate electronic mail systems, researchers must also be wary of changes in user expectations. Taspcott (1982, p. 207)

monitored the use of an electronic mail system in an organization over time. He found that as measures of information obtained from the system increased over a nine month period, so did measures of information required to perform normal expected duties. Licker et al. (1986) formulated a similar model suggesting that adoption of an electronic mail system was a cyclical relationship between use, value, and preference. Both studies suggest that measurement over time would be required to discover this artifact.

## 4.3 Summary of Findings

While such effects as wording and order of presentation no doubt can occur in computerized questionnaires, this study is only concerned with those response effects specific to computerized questionnaires. In articles dealing with computerized data collection three main ideas tend to prevail. The first is the idea that the computer will increase honesty, or reduce the amount of social desirability in the response. (Evan and Miller, 1969; Lucas et al., 1977; Allen, 1987; Millstein, 1987; Canoune and Leyhe,1985; Sproull, 1986). The suggestion is that the instrument may change a respondent's perceived need to maintain positive self presentation. This can be measured by:

1) variation in honesty in responses.
2) variation in amount of social desirability in responses.

Certainly the difference between face to face and self completion interviews is well documented, with most theory suggesting that the amount of social contact will determine expected bias. In this respect computer completion should be no different than that of pen and paper self completion. As well, criticisms have been made of the Crowne-Marlow scale (Sudman and Bradburn, 1979) and no reliable alternatives have been suggested (Nederhof, 1985).

The second theme is the measurement of user attitudes towards the technique and the effect that this has on completion rates (Higgins, 1986; Sproull, 1986; Liefeld, 1986; Allen, 1987; Erdman et al., 1985). While not a response effect in the strictest sense, the

collection instrument can also bring about differences in respondent effort. Respondent effort can be operationalized as:

1) percentage of completed interviews and percentage of answered questions,
2) a function of completion time, where increased duration could be interpreted as an indication of increased effort.

It has already been shown that electronic surveys are viewed positively by respondents and can help to increase response rates. There has not, however, been any work to determine if increased completion times are indeed a result of increased effort or simply a function of user's inability to adapt to new software.

The third major area is the extent that the computer will effect responses to various type of questions. Liefeld (1988) has done the most detailed work in this area, identifying some response effects specific to question type. Response effects created by the instrument may manifest as follows:

1) more agreement to scale questions,
2) variation in response patterns for scale questions, measured by the standard deviation of responses,
3) variation in response patterns to multiple response questions,
4) variation in content (in absolute terms) to open ended questions,
5) variation in rank order for rating questions,
6) entry errors on the part of the respondent.

Finally, a conceptual framework for explaining the causes of response effects in computerized data collection is not complete. Most studies only go as far as suggesting that response effects are a result of computer interviewing in general. For example, Liefeld (1986) suggests that increased mood may be the cause of response effects in his studies. Kiesler and Sproull (1986) suggest that selection of more extreme responses is a result of reduced social context. The research in this area is at a stage that it is no longer

sufficient to identify the possible response effects that may be elicited, but necessary to determine *sources*. Understanding the source of a response effect will give a better framework from which to identify other potential biasing variables. Below is a summary of findings to date and questions not yet addressed, presented within the context of the conceptual framework discussed in this paper.

CHARACTERISTICS OF THE RESPONDENT. For computerized questionnaires, the most salient characteristic of the respondent is the user's previous computer experience. No studies have attempted to determine the effect that the user's computer experience will have on the likelihood of exhibiting response effects. The relationship between "hands-on" experience and likelihood of exhibiting a response effect may be linear, or a dichotomous step function. Perhaps one or two previous "hands-on" experiences are sufficient to eliminate the possibility of response effects associated with fear of computers or the user's inability to complete the questionnaire. It may also be necessary to create a more complete measure of previous computer experience than the number of hours of "hands-on" computer experience. Knowledge of computer concepts, and attitudes towards computers may also affect the likelihood of demonstrating response effects.

In addition, a potential for *sampling* bias exists if respondents are allowed to self-select. Initial studies have shown, not surprisingly, that a larger than expected proportion of teenagers will be attracted. Sex, formal education, and previous experience with computers may also affect whether a respondent chooses to participate in a survey conducted by an unattended terminal. No work has been done on the psychographic profile of the user that may be attracted.

INTERVIEWER EFFECTS. In the field of psychology the computer has been found to be an unbiased administrator of psychological tests. Respondents are quite congenial towards the technology, and there is evidence that this increase in mood is reflected in respondents' answers. There is some indication that the computer will serve to increase honesty in respondents, particularly if compared to a human interviewer. Kiesler and Sproull (1986) suggest that less social desirability is evidenced by more extreme

responses to scale questions, however, these findings were not confirmed by Liefeld (1987b). There is little grounded theory (in terms of linking effect and theorized cause through empirical findings) behind other findings in this area. Often the size of these effects are not taken in proper perspective and results may be a function of the interview setting rather than use of the technology. In the case of unattended terminals research is also required as to the possible effects of "group" completion of questionnaires, and lack of privacy in completing the questionnaire.

QUESTION CONTENT. This area appears to pose a high potential for creating response effects. While simple behavioral questions are not affected, such response effects as "video game syndrome" for multiple response knowledge questions, and higher responses to attitudinal questions have been demonstrated. Additional work is required to confirm these findings in other settings. With respect to sensitive and threatening content, there is no clear evidence that the computer will not have the same problems that other data collect methods encounter when attempting to obtain truthful responses.

PRESENTATION. Only one work addresses this area (Liefeld, 1987b). The suggestion is that for scale questions initial cursor position will not effect answers to scale questions. Little is known about how the use of graphics, placement of error messages, and level of on-line help effect user responses. Future studies in this area would require the luxury of access to two different software packages utilizing different protocols. As prices come down, and more packages become available this type of study may be common in the future.

HUMAN FACTORS. No inherent ergonomic disadvantage in using video technology for the communication process has been demonstrated. There may be response effects that are derived from interacting with the technology, particularly if the user must adopt new cognitive processes. Computer questionnaires generally take longer to complete; the degree that this is a function of previous experience has not been identified.

RESPONSE TASK. The findings here are mixed. Liefeld (1987b) found that an entry protocol requiring cursor movement to answer scale questions reduces the range of response to scale questions. Other studies, using number pad entry, have found more extreme responses in answers. These findings are of particular interest as researchers use response patterns to scale questions to measure other effects such as acquiescence, extremity of opinion, and social desirability. For example, Kiesler and Sproull (1986) suggests that a wider range of responses for computer interview respondents was an indication of less social conformity in answers. Yet no studies to date have attempted to determine if results are a function of the specific entry protocol, rather than "computer interviewing" in general. The effects of light pen, touch screen and mouse response mechanisms are unknown at this time.

AMOUNT OF FEEDBACK. The interactive nature of computerized questionnaires appears to increase response rates and effort in responding to questions. It is not known, however, the role that the novelty of the medium has in increasing cooperation.

ARTIFACT. Claims have been made that, even when controlling for social contact, the computer will create more honesty in responses. There is, however, no strong rationale which may explain *why* this alleged effect occurs. A possible explanation is the respondents' perception of the abilities of the computer. Additional studies need to be performed within the context of this explanation.

## 4.4 Rationale for This Study

Many of the previous questions can only be answered in a laboratory setting, utilizing two questionnaires on different software and strict controls of respondents' mood, computer experience and with access to verifying data. This approach however, would not answer the questions that the practitioner asks. "Can I use the computer in the field, and if so what response effects should I be prepared to safeguard against? Can respondents understand and complete a questionnaire without human instruction? Will

answers be any different than those I would get from traditional pen and paper methods? Will the computer "illiterate" be able to complete the questionnaire without bias?"

While it is not within the scope of this study to look at all potential biasing variables, an examination of the influence that variation of the *entry task* (Figure 1, Box F), and the varying levels of *previous computer experience* (Figure 1, Box A), have on response effects will provide a good start for the development of a theoretical framework of response effects in computerized data collection.

The importance of the entry task as a biasing variable is an important first step in laying the foundation for future studies. Other response effects such as acquiescence, social conformity, and engagement in the process are often measured by examining response patterns to scale questions. Yet the research has provided no clear indication that response effects are not simply a function of the protocol of the entry task (which can vary depending on the software used), rather than the phenomenon being measured, "completion by computer". As well, few studies have been performed with inexperienced computer users, where the potential for bias is much greater, in fact the very nature of the relationship between response effects and computer experience is not known. Finally, the study will limit itself to only the following response effects that are thought to be elicited by the data collection instrument and examine:

a) agreement to scale questions ( means),
b) response patterns to scale questions (standard deviation),
c) response frequencies for each choice on multiple response questions,
d) length and content of open-ended questions.

This information, in combination with an evaluation of the ability of an unattended terminal to obtain a sample representative of the local population frame, will give the practitioner useful information regarding the suitability of this new data collection technique for his or her research needs, as well as providing more information for grounded theory in this area.

# CHAPTER 5

## OBJECTIVES

While computerized interviewing can be used in a multitude of disciplines, this study will focus on its applicability in the area of tourism research. The study should shed light on the potential use for attraction/event exit surveys and as a cost effective instrument to obtain general traveller data. From a theoretical perspective the objectives of the study are:

(I) To examine response biases found in other studies relating to computerized questionnaires. This will include variation in mean response to scale questions, variation in response patterns to scale questions, response frequencies for each choice on multiple response questions (video game syndrome), and length and content of open ended questions.

(II) To determine if variation in the entry task causes response effects in answers. Specifically the study will examine three entry tasks: pen and paper completion, computerized questionnaire utilizing cursor movement for selection of answers, and computerized questionnaire utilizing number pad entry. This three group study will also help to determine if response effects are the result of a "non-human" yet interactive computerized interview, or more specifically a function of the specific computer entry task.

(III) To determine if a respondent's previous computer experience has any effect on the likelihood of demonstrating response effects. Most studies to date have been on the computer "literate" and none have examined the very real effect that lack of previous experience may have on the way respondents answer questions.

As well, the study will investigate the validity of data collected by a methodology which allows respondents to *self-select at an unattended kiosk* operating a computerized questionnaire. The study will examine practical aspects regarding the use of this methodology in a tourist attraction field setting, specifically:

(IV) To determine if a sample collected by a computerized questionnaire on an unattended terminal is an unbiased estimator of the attraction population. This has important ramifications for the researcher. If the sample is not representative, some type of human involvement will be required to implement suitable sampling procedures.

# CHAPTER 6

## HYPOTHESES

Stemming from the objectives presented in chapter 5, the study has four main predictions.

(1) The "non-human" yet interactive nature of the computerized interview will cause response effects in answers.

Previous studies have shown that computerized collection has increased the mean response to scale questions (Liefeld, 1987b), increased the number of choices for multiple response questions (Liefeld, 1988), increased the number of words in open ended questions (Higgins et al, 1987), and increased cooperation (Sproull, 1986; Higgins, 1987). As well, some studies have shown computerized collection to increase the variance in responses to scale questions (Sproull, 1986; Allen, 1987). It is necessary to confirm that these response effects, as compared against pen and paper completion, exist in an alternative setting.

(1.1) Completing a computerized questionnaire will effect the mean response to scale questions.

(1.2) Completing a computerized questionnaire will effect the variance in answers to scale questions.

(1.3) As compared with pen and paper completion, computerized interviews will result in more choices selected for multiple response questions.

(1.4) As compared with pen and paper completion, answers to open ended questions will contain more words and concepts.

(1.5) Computerized interviews will have fewer unanswered questions and fewer unfinished questionnaires than the pen and paper completion group.

(2) Response effects are a function of the entry task.

Some contradictory findings may be attributable to the differences in software utilized by various researchers. The flexibility of computerized questionnaires allows for a number of possible different answering protocols. Two of the most common: are to use a number pad, with choices identified by specific numbers; or a cursor movement "shoot and pick" protocol, where answers are selected by using a cursor to highlight the most appropriate selection. In contrast to Sproull (1986) and Allen (1987), Liefeld (1988) found that computerized data collection using a cursor movement protocol *reduced* the variance of responses to scale questions. As well, the software that Liefeld used in obtaining an increased tendency to select choices to multiple response questions utilized this cursor movement with the ability to toggle selections on and off. This novel, or game-like protocol may encourage more responses than a relatively simple number pad entry protocol.

(2.1) For computerized questionnaires, an entry task requiring cursor movement will result in different mean responses to scale questions than an entry task requiring number pad entry.

(2.2a) As compared to pen and paper completion, computerized interviews using a cursor movement protocol to select answers will result in fewer extreme answers to scale questions.

(2.2b) As compared with pen and paper completion, computerized interviews using number pad entry to select answers will result in more extreme answers to scale questions.

(2.3) For computerized questionnaires, an entry task requiring cursor movement to select choices will result in more answers to multiple response questions than will an entry task that uses number pad entry.

(3) The magnitude of response bias will be affected by previous computer experience.

If response effects are a function of the respondent being unable to operate the software, persons with more previous computer experience would be less likely to display response effects. As well, persons with *less* experience may have poor typing skills, thereby limiting the length of answers to open ended questions.

(3.1) Respondents with more experience will be less likely to demonstrate an overall increase in mean response to scale questions.

(3.2a) Respondents with more experience will be less likely to demonstrate narrow range of responses to scale questions (cursor movement entry).

(3.2b) Respondents with more experience will be less likely to demonstrate extreme responses to scale questions (number pad entry).

(3.3) Respondents with more computer experience will give fewer answers to multiple response questions.

(3.4) Respondents with more computer experience will give more words and concepts to open ended questions.

(3.5) Respondents with more computer experience will have shorter completion times.

(4) An unattended kiosk will attract a non-representative sample.

Newsted (1985) and Niman (1988) have shown unattended kiosks to attract younger persons with more education. Associated with this would be an expectation of higher household incomes. Persons with previous computer experience may be more likely to approach an unattended terminal. A visit to the local video game arcade would also suggest the likelihood of attracting more males. Self selection is also a function of time and interest. Distant visitors may have more time or inclination to complete a questionnaire, as would persons with higher interest in and involvement with the attraction.

As compared with a random sample of the attraction population:

(4.1) An unattended kiosk (self-selection) will attract a sample biased with respect to age distribution, specifically a larger than expected proportion of young (13-25) respondents, and a smaller than expected proportion of old (50+) respondents.

(4.2) An unattended kiosk (self-selection) will attract a larger than expected proportion of persons with higher levels of formal education.

(4.3) An unattended kiosk (self-selection) will attract a larger than expected proportion of persons with higher household incomes.

(4.4) An unattended kiosk (self-selection) will attract a larger than expected proportion of computer users.

(4.5) An unattended kiosk (self-selection) will attract a larger than expected proportion of male respondents.

(4.6) An unattended kiosk (self-selection) will attract a sample biased with respect to geographic distribution. (Visitors whose normal residence is a large distance from the attraction may be more likely to self-select).

(4.7) An unattended kiosk (self-selection) will attract a larger than expected proportion of annual pass holders.

# CHAPTER 7

# METHODOLOGY

## 7.1 Procedure and Design

In order to test all predictions the study had two distinct components, both of which were conducted during the same three week period. An exhibit at a local attraction was selected as the data collection site, where respondents were asked to complete a typical on-site visitor questionnaire.

Part A. Part A was designed to examine the response effects that may be elicited by completing a computerized questionnaire. This was done by a between subjects design with three randomly selected groups: one self-completing the questionnaire by pen and paper (SC), and two completing interactive computerized questionnaires on personal computers with standard keyboards; one requiring a cursor movement protocol to answer questions (CM), and the other requiring number pad entry to answer questions (NP).

A stratified sampling strategy was used. Examination of previous years' daily visitation rates showed that during the average week approximately 40% of visits occurred on Mondays to Thursdays, 30% of visits occurred on Fridays and Saturdays, and 30% of visits occurred on Sundays. Visitor traffic during the pre-test was relatively constant during the entire day with a marginal slow down during noon, and the first and last hour of opening much slower than the rest of the day. Previous experience indicated that persons with young children were more likely to visit in the morning. In order to reflect this estimated fluctuation in volume, ten sampling periods were scheduled over three weeks: one period for each of Monday to Thursday (40%), two Saturdays and one Friday (30%), and three Sundays (30%).[2] Five of the samples were performed in the morning, starting between 9:30 and 10:30, and five performed in the afternoon, starting between 12:45 and 1:30. It took anywhere from three to five hours

2 Due to inclement weather (a spring blizzard) one Sunday was replaced with a Saturday sample.

to complete the daily quota of 60 respondents, depending on traffic volume.

During each sampling period, every 3rd person entering the exhibit was selected.[3] Once selected, respondents were randomly assigned to a method of completion, in total 200 persons were interviewed by a computerized questionnaire utilizing cursor movement (CM), 200 persons by a computerized questionnaire utilizing number pad entry (NP), and 200 persons completed the questionnaire using pen and paper. If a person refused, the respondents sex, and reason for refusal (if given) were recorded. An estimate of the respondents age was also made and recorded at the time of refusal.

Part B. Part B was designed to determine if an unbiased population estimate could be collected by allowing respondents to self-select at an unattended kiosk (a personal computer on top of a wooden stand, located in a visible location with a large "Visitor Survey" sign). The sample collected in Part A was used as an estimate of the true sampling frame of persons visiting the exhibit, hence the requirement for conducting both components of the study over the same time period.

Respondents in Part B sampled by self-selection, choosing to participate by approaching and completing a questionnaire located at an unattended kiosk (UK) utilizing only one completion method, the same software as NP above. The questionnaire was a shorter split version (to reduce total completion time) of the questionnaire described above, with a different half running on alternating days. Each version included all demographic and computer experience related questions, half of the attraction related questions, and none of the travel related questions. It also included one question as to whether the questionnaire was answered alone, in a group, or with other persons watching. To facilitate ease of use it included a short pictorial-type tutorial indicating location of keys (which were also color coded and labeled), on-line help, and the ability to back up to previous questions or restart from the beginning. The software

3 On two very slow days this was adjusted to every second person. Note as well that the interview site could process at most only four computerized questionnaires at one time, and as such was occasionally not flexible enough to adjust to sudden increases in visitor traffic. In this case the next third person from a point in time (namely the freeing up of one computer) to enter the exhibit was selected.

also incorporated other features necessary to run in an unattended environment, including an automatic logging of start and finish times for each questionnaire.[4]

The unattended computerized questionnaire was operational the entire three week period, with the exception of those times that the sampling in Part A was taking place. The only exception to this was one three hour period (2:00 to 5:00) due to a software 'lock-up', and a two hour period (12:00 to 2:00) while a malfunctioning keyboard was replaced.

## 7.2 The Questionnaire

The final form of the questionnaire included 11 demographic questions, 4 multiple response questions, 3 open-ended questions and 20 Likert-type scale questions. All of the multiple response and open-ended questions were provided by the attraction management, as were 12 of the scale questions. Three scale questions were required to measure attitudes towards computers, and the questionnaire also included 5 questions about vacation travel to obtain a broader range of content without excessive variation on the apparent theme of the questionnaire. These questions were selected from the Canadian Tourism Attitude and Motivation Study (Tourism Canada, 1983), based on reasonably normal response distributions and means that were either neutral, slightly positive or slightly negative. It was felt that these distributions would be sensitive to response effects in either direction.

The questionnaire was created on three mediums: pen and paper; with the NP group utilizing the MAZE interviewing software package (Fiske, 1987); and the CM group utilizing the CAPPA interviewing software package (Green, 1985).[5] All three mediums used the same questions. However, the computerized questionnaires also included a practice section, which was used in conjunction with verbal instructions to demonstrate how questions were to be answered. Both computerized questionnaire

4 These features included: an internal trap of <Ctrl C>, a shell operating environment that allowed password-only escape to DOS (to deal with potential "hackers"), and a routine that started a new questionnaire in the event of no keyboard activity for a period of 90 seconds (to deal with the problem of persons walking away in the middle of a questionnaire).

5 This is the same software used by Liefeld (1986, 1987, 1988).

groups used personal computers with standard keyboards to answer the questions. Function keys, and other non-essential keys were covered with tape to reduce confusion.

The entry task of the two computer groups differed in the following way. The NP group used the numbers at the top of the keyboard to make their selection, and pressed the Enter key to proceed to the next question. The CM group used a cursor movement protocol that required (1) using one of four arrow keys to MOVE the cursor to their choice, (2) pressing the Enter key (color coded red) to PICK their choice, and (3) pressing the F10 key (color coded blue) to proceed to the next question. This would, however, vary slightly for different types of questions. (In addition to the written descriptions below, Appendix A has screen snap shots of example multiple response and scale type questions.)

For multiple response questions the NP entry protocol required the respondent to enter all appropriate numbers (each possible selection was preceded by a number) into an eight character field. The CM program used an entry protocol that required the respondent to move the cursor in front of the desired selection and press the red Enter key to "pick" (at which point the background of the selected answer turned red). This process was repeated until all selections were picked, at which time the respondent pressed the blue F10 key to proceed to the next question.

For scale questions the NP program labelled each of the five points: strongly agree, agree, neutral, disagree, strongly disagree. Respondents simply entered the number that best reflected their feelings. The CM software used a horizontal line with five possible cursor positions and labels "strongly agree" and "strongly disagree" at either end point. Verbal instructions were given on how to select "agree", "disagree" and "neutral" responses.

For open-ended questions, the entry task for both computer groups was almost identical, with the exception that the CM group had a larger entry field (10 lines) than both the NP (4 lines) and the SC paper version (4 lines). Both software programs required the respondent to use the keyboard to type in their response and press the Enter key (NP) or the blue F10 key (CM) to proceed to the next question.

## 7.3 Pre-test

Pre-testing involved design and testing of the two computerized questionnaires on university students, and a preliminary on-site test for the paper questionnaire (ten respondents) to determine if length was appropriate and to confirm that questions were understood and interpreted properly. Once these tasks were completed, a two day pre-test at the study site was done for 60 respondents, 20 questionnaires completed by pen and paper and 20 performed on each software package. Post-questionnaire interviewing necessitated minor modifications to the computer questionnaires, including the replacement of a self guided tutorial (which took too long to complete) with four example questions and verbal instructions. Pre-test respondents indicated that instructions for answering questions were clear, on-line help was adequate, and that no questions were ambiguous in nature.

The unattended terminal was left at the study site and observed for a two day period. This pre-test suggested that improved visibility was required (a large "VISITOR SURVEY" sign subsequently created by the attractions graphic department was adequate), and that keys should have been identified with colors rather than names. The "F1" key, for example, could not be located by novice users. These modifications were made for the actual study.

## 7.4 Measures

The following measures were used to quantify response effects:

(1) Mean Response to Scale Questions. It is possible that the computerized completion has a novelty-enjoyment effect that increases the number and extent of positive responses to scale questions. To determine if this is the case, the mean response to 17 questions can be compared across methods of completion.[6]

---

6 The three scale questions not included were about the respondents attitudes towards computers. As over one quarter of the respondents had no previous computer experience, it was felt that the actual first experience of using a computer may directly affect their answers to this question. For example, a respondent who has never used a computer before may not be able to make a judgement about their apprehension, whereas the same respondent, if selected for completion in one of the two computer groups, could base this answer on their only experience with computers, the actual completion of the questionnaire. As such, other variables could be responsible for differences between the two computer and self completion groups.

(2) Range of Response to Scale Questions. If a five point scale is used, a wider range of responses can be approximated by counting the number of times each respondent selected an extreme point (either 'strongly agree' or 'strongly disagree'). This "tendency to use extremes" measure was calculated by allocating respondents 1 point for every response answered with an end point (either positive or negative) for each of fourteen scale questions.[7] While each person would be expected to have a different score, sample sizes are large enough that each group should have a similar distribution of scores, and a comparison can be made to discern if there are differences across method of completion.

(3) Number of Choices for Multiple Response Questions. To determine if a video game syndrome, as reported in Liefeld (1986) exists, the total number of choices to all multiple response questions in the questionnaire were added together to create a measure of "tendency to select".

(4) Length of Responses to Open-ended Questions. Total number of words was simply counted for each response to each of the three questions. Counting concepts was a little more involved. A concept was defined as a "distinct thought or idea". A preliminary examination of 100 respondents created the concept categories. In addition, due to difficulties that other researchers have had in consistent identification of concepts (Perschel, 1972), the idea of a concept modifier was used to differentiate between simple and more complete answers. For example, "I brought visitors" was coded as one concept - BRING VISITORS. The answer "To bring friends from Toronto" was coded as two concepts - BRING VISITORS and the concept modifier TORONTO. As the actual title of the concept modifiers were of no substantial value to either the attraction or the research design, all modifiers were counted in the category OTHER. Using this measure, the number of concepts counted would not be a function of the subjective categories created in the initial analysis, and longer, more elaborate answers would have higher concept counts.

An undergraduate student was given instructions for coding and a sub-sample of

7 Three questions about the attraction used a five point scale with different semantic labels: Poor, Fair, Good, Very Good, and Excellent. The question of extremes would not apply to this type of labeling

90 questions to code. This same subset was coded by a second reviewer in a similar fashion to determine consistency. A correlation coefficient (for the number of concepts coded for each question, rather than consistency in labeling the concept) between the two coders of .9 was obtained. Any differences (for the first ninety respondents) between the two coders were discussed, and the student then proceeded to code all 1800 responses.

# CHAPTER 8

# RESULTS

## 8.1 Part A (Response Effects)

A total of 87 of the 687 persons selected refused to participate in the study for an overall response rate of 87.3 percent. Reasons for refusal are presented in Table 2.

- TABLE 2 -

Reason For Refusal

| | Number | Percentage |
|---|---|---|
| No Reason Given | 39 | 45% |
| First Visit & First Exhibit Viewed | 17 | 20% |
| Can't Leave Children Alone | 14 | 16% |
| No English | 8 | 9% |
| Incomplete Questionnaire | 3 * | 3% |
| Employee at the Attraction | 2 | 2% |
| No time | 2 | 2% |
| Refused to Use Computer | 2 | 2% |
| No Reading Glasses | 1 | 1% |
| TOTAL | 87 | 100% |

* Paper questionnaires completed in less than 3 minutes were examined in the field, and if incomplete were discarded and replaced by another respondent (randomly selected).

Because persons were randomly assigned to different methods of completion, all three treatment groups should demonstrate the same respondent profiles. A Chi-square analysis of the three groups (CM, SC, NP) on the following demographic variables: age, sex, income, education, place of residence, number of dependent children, amount of hands-on computer experience, and annual membership at the attraction showed no significant differences in distribution among the three groups (see Table 3). In addition, only two persons initially agreed, but subsequently refused upon being assigned to one of the computer completion groups. This analysis confirms that proper controls were taken in randomly assigning respondents to completion method.

- TABLE 3 -

Comparison of Demographics By Method of Completion

| Variable | $Chi^2$ (df) | | Probability |
|---|---|---|---|
| Sex | .2 | (2) | .90 |
| Age | 10.9 | (12) | .53 |
| Income | 6.6 | (12) | .94 |
| Education | 6.4 | (10) | .78 |
| Residence | 15.3 | (14) | .36 |
| # of Children | 8.8 | (8) | .36 |
| Annual Pass | 2.1 | (2) | .35 |
| Computer Experience | 8.2 | (8) | .41 |

## 8.1.0 Data Analysis and Distribution Assumptions

Prior to data analysis all variables were examined for missing values and the fit between their distribution and the assumptions for multivariate analysis. Two parametric tests, ANOVA, and MANOVA were used; the data screening for each is discussed in turn.

Analysis of Variance. This test is used repeatedly throughout the results section. While the univariate F test of mean differences is frequently said to be robust to violations of normality and homogeneity of variance with large and equal sample sizes, Tabachnick and Fidell (1989) recommend transforming variables to fit these assumptions. Providing there are no outliers, Harris (1975) suggests that formal homogeneity of variance checks are not necessary if the ratio of the smallest to largest sample is no greater than 4:1 and the ratio of the smallest to largest variance is no greater than 20:1. These conditions are satisfied for all analysis in this study. To satisfy normality, any significantly skewed variables (at $p<.001$) were transformed logarithmically or by taking the square root. This transformation resulted in no variable having more than 3 univariate outliers, a number which would be expected given the large sample size. To ease interpretation, presentation in tables utilizes the means of the original variable and the F statistic of the transformed variables.

Multivariate Analysis of Variance. This analysis was performed treating the 17 scale questions as unique variables. Four questions had a "No Opinion" response as pre-testing showed that many people had not seen these exhibits. This resulted in a large number of missing values for three of these variables (100,192,289). As a regression analysis showed that no combination of other variables could accurately predict this variable, missing values for these variables were simply assigned the mean of each variable. Missing values for the other questions, predominantly from the SC group, were dropped, resulting in 544 of the original 600 cases. Nine variables were significantly skewed (at $p<.001$), and were transformed logarithmically, by taking the square root, or by taking the square root of the reflected variable. Three variables were still skewed after these transformations, however, Mardia (1971) shows that providing the smallest cell has at least 20 degrees of freedom, MANOVA is robust to modest violations of normality if the violation is created by skewedness rather than outliers. A check for univariate outliers revealed that none of the variables had more than the expected number with a sample this large. Four multivariate outliers (within their own group) with a Mahalnobis distance greater than $p<.001$ were removed, leaving 540 cases. Formal homogeneity of variance-covariance matrices checks are not necessary due to the acceptable ratio of variances and sample sizes (Harris,1975). Linearity can be assured from the transformation of skewed variables, and multicollinearity is not a problem as correlations among dependent variables are quite low.

### 8.1.1 Mean Response to Scale Questions

A first step of the analysis was a MANOVA. Treating each of 17 questions as dependent variables and collection method as the independent variable, MANOVA showed a significant collection method was found to have a significant effect on the dependent variables (Wilks' approximate $F=1.96$, $p=.001$). [8] Next, a series of univariate ANOVAs for each of the questions showed significant differences: two travel questions, one about reducing hours, one about increasing type X exhibits, and one about attitudes towards computers. Mean responses to each question are shown in

8 The three computer related questions were not included in this analysis (see footnote 6 for rationale).

- TABLE 4 -

Mean Response (+) to Twenty Scale Questions

| Question | CM | SC | NP | F | (df) | p.= |
|---|---|---|---|---|---|---|
| ATTRACTION | | | | | | |
| Rate exhibit A * | 1.86 | *1.98* | **1.81** | 2.45 | (2,579) | NS |
| Rate exhibit B * | **1.57** | *1.63* | 1.58 | .34 | (2,499) | NS |
| Rate exhibit C * | **1.55** | *1.76* | 1.67 | 2.18 | (2,407) | NS |
| Quality of Info? * | *1.48* | 1.45 | **1.34** | 1.22 | (2,310) | NS |
| More of exhibit X | **2.23** | *2.46* | 2.37 | 4.17 | (2,588) | .0015 |
| More of exhibit Y | *2.41* | 2.37 | **2.36** | .16 | (2,587) | NS |
| Label displays? * | *2.04* | 1.98 | **1.92** | .52 | (2,586) | NS |
| City aware of B? | 2.69 | *2.69* | **2.55** | 1.33 | (2,577) | NS |
| Up Gate Price? | *3.41* | 3.35 | **3.24** | 1.06 | (2,578) | NS |
| Up Annual Pass? | *3.17* | **2.99** | 3.02 | 1.35 | (2,569) | NS |
| Pay to Park? * | 3.47 | **3.39** | *3.52* | .76 | (2,577) | NS |
| Reduce Hours? * | *3.84* | **3.56** | 3.66 | 4.64 | (2,572) | .01 |
| TRAVEL | | | | | | |
| Travel Same Place | *3.22* | 3.14 | **3.07** | .81 | (2,581) | NS |
| Money Well Spent * | *1.91* | 1.88 | **1.83** | .31 | (2,583) | NS |
| Pay More for Extras | *2.92* | 2.67 | **2.59** | 4.87 | (2,583) | .008 |
| Don't Travel Much * | *2.57* | 2.50 | **2.38** | 1.39 | (2,585) | NS |
| Use Travel Agent | *3.75* | 3.32 | **3.25** | 10.9 | (2,583) | .0001 |
| ATTITUDES TOWARDS COMPUTERS | | | | | | |
| Understand | *2.99* | **2.88** | 2.89 | .46 | (2,576) | NS |
| Feel Apprehensive | *3.66* | **3.48** | 3.64 | 1.4 | (2,573) | NS |
| Read About | *3.46* | **3.03** | 3.2 | 6.56 | (2,573) | .0005 |

(+) As the computer questionnaires coded a Strongly Agree as = 1 and Strongly Disagree as = 5, a *lower* mean represents stronger agreement with the statement.
* These variables were transformed to meet the parametric assumption of normality by either taking the square root, or log. For easier interpretation, the means of the original variable are presented along with the F statistic and probabilities for the transformed variable.
**BOLD** - most positive of three groups *Italics* - most negative of the three groups

Table 4.

To perform specific comparisons, this MANOVA was repeated, comparing each of the two computer groups against the SC group on an individual basis. This resulted in a significant effect (Wilks' F=2.19,p=.005) for the SC-CM pair, *but not* for the SC-NP pair (Wilks' F=.96,p=.503). While this result does suggest that there are differences between computer and paper questionnaire completion, the prediction that "use of a computer" will cause a response effect for mean responses to scale questions can only be supported if *both* computer groups are different than the SC group. As this is not the case the prediction 1.1 can not be supported.

### 8.1.2 Range of Response to Scale Questions

As can be seen in Table 5 the CM group has the largest standard deviation in 15 of 17 questions and the Paper group has the smallest standard deviation in 11 of 17 questions. A Cochran C homogeneity of variance test was used to determine if these variances were likely to have been obtained from the same population. This test, using the untransformed variables, showed significantly different distributions for 11 of 17 scale questions ($p<.01$).

The "tendency to use extremes" variable was calculated by summing the number of end points selected over the fourteen scale questions using the 'strongly agree' to 'strongly disagree' labelling. This variable was positively skewed with values ranging from 0 (no end points chosen in any scale questions) to 13 (end points selected in all but one question). Means and standard deviations of the number of end points selected are presented in Table 6.

Analysis of variance of the number of end points by collection method show significant differences among the groups. A subsequent Duncan test ($p<.05$) showed all three means to be significantly different from each other. In particular, the CM group used about 50% more end points than either of the other groups. As both computer groups generated significantly more extreme responses than the SC group, it would appear the computer collection has some effect on the variance in responses to scale questions. However, as there is a significant difference between the two computer

- TABLE 5 -

Standard Deviation of Responses to Scale Questions

| Question | CM | SC | NP | Homogeneity of Variance Cochrans C | p.= |
|---|---|---|---|---|---|
| ATTRACTION | | | | | |
| Rate exhibit A | *.75* | .78 | **.81** | .358 | NS |
| Rate exhibit B | **.75** | .74 | *.71* | .351 | NS |
| Rate exhibit C | *.80* | **.94** | .84 | .340 | NS |
| Quality of Info? | **.72** | .66 | *.65* | .375 | NS |
| Up Gate Price? | **1.24** | *1.13* | 1.19 | .362 | NS |
| Up Annual Pass? | **1.24** | *1.03* | 1.08 | .411 | .011 |
| Pay to Park? | **1.53** | *1.24* | 1.25 | .428 | .002 |
| Reduce Hours? | **1.34** | 1.14 | *1.07* | .423 | .003 |
| More of Exhibit X | **.93** | *.69* | .82 | .429 | .001 |
| More of Exhibit Y | **.97** | *.74* | .78 | .450 | .000 |
| Label displays? | **1.09** | *.76* | .87 | .473 | .000 |
| City aware of B? | **1.07** | .90 | *.89* | .419 | .004 |
| TRAVEL | | | | | |
| Travel Same Place | **1.26** | *.97* | 1.13 | .416 | .005 |
| Money Well Spent | **1.00** | *.72* | .75 | .477 | .000 |
| Pay More for Extras | **1.19** | *1.00* | 1.05 | .402 | .023 |
| Don't Travel Much | **1.19** | *.94* | 1.00 | .429 | .001 |
| Use Travel Agent | **1.19** | *1.08* | 1.13 | .367 | NS |

**BOLD** - largest standard deviation. - *Italics* smallest standard deviation.

completion groups, the prediction that range of responses will be affected simply by"using a computer" cannot be supported without some reservation.

### 8.1.3 Multiple Response Questions

Multiple response questions allow respondents to select as many choices as are applicable. Two (possible) biases for this type of questions are possible. First, respondents may be more likely to select a *specific* response as a result of the collection

- TABLE 6 -

Average Number of End Points Selected

| Group | CM | SC | NP |
|---|---|---|---|
| Mean | 5.32 | 3.05 | 3.76 |
| Standard Deviation | 2.76 | 1.95 | 2.25 |

$F(2,597) = 48.7$, $p=.000$ Results are for the skewed variable, normalization does not change the significance level.

method. In order to determine if this was true, the actual frequency of positive responses (i.e. the number of times a choice was selected) was compared across method of completion. In answering the four multiple response questions, respondents couldpotentially select a total of 23 choices (respondents answered either question 2a or 2b, see Table 7). Comparisons were made across method of completion, treating each one of these choices as a dichotomous variables. Seven variables (choices) had significant Chi-square values but there was not any clear cut trend for more (or less) selections of answers for any method. Interestingly, differences were more likely to occur in the first possible choice of each question (5 of 7 times), though again, there is no definite trend where one collection method was consistently higher or lower than the others.

Second, collection method could also effect the number of choices selected over the entire questionnaire. This "tendency to select" variable had scores ranging from 1 to 22 (where 23 represents selecting all choices on all multiple response questions). Analysis of variance of this variable by collection method showed no significant differences among methods of collection (see Table 8).

However, as Liefeld (1986) demonstrated a "video game syndrome" for multiple response questions asking about knowledge of outlets in a shopping malls, a post-hoc distinction was made based on the content of questions. The "tendency to select"

- TABLE 7 -

Multiple Response Questions:
Number of Choices Selected by Question

| QUESTION | CM | SC | NP | $Chi^2$ (2 df) | prob |
|---|---|---|---|---|---|
| 1. Reason for visiting | | | | | |
| Reason # 1 | 77 | 105 | 97 | 8.28 | .015 |
| Reason # 2 | 31 | 24 | 31 | 1.30 | NS |
| Reason # 3 | 138 | 151 | 142 | 2.25 | NS |
| Reason # 4 | 91 | 115 | 91 | 7.84 | .020 |
| Reason # 5 | 109 | 145 | 109 | 18.35 | .0001 |
| Reason # 6 | 16 | 16 | 17 | .04 | NS |
| Reason # 7 | 49 | 65 | 61 | 3.32 | NS |
| 2a. Reasons for Annual Pass Purchase | | | | | |
| Reason # 1 | 54 | 55 | 38 | 6.54 | .0378 |
| Reason # 2 | 8 | 11 | 8 | .70 | NS |
| Reason # 3 | 34 | 26 | 27 | 2.06 | NS |
| Reason # 4 | 27 | 27 | 21 | .08 | NS |
| Reason # 5 | 43 | 45 | 39 | 1.42 | NS |
| Reason # 6 | 9 | 10 | 9 | .27 | NS |
| 2b. Reason for Not Purchasing an Annual Pass | | | | | |
| Reason # 1 | 34 | 30 | 20 | 6.76 | .035 |
| Reason # 2 | 13 | 14 | 10 | 1.52 | NS |
| Reason # 3 | 9 | 7 | 13 | 1.23 | NS |
| Reason # 4 | 41 | 53 | 51 | 3.61 | NS |
| Reason # 5 | 23 | 31 | 29 | 2.18 | NS |
| Reason # 6 | 36 | 40 | 43 | .71 | NS |
| 3. Familiarity With Attraction Programs | | | | | |
| Program # 1 | 74 | 31 | 59 | 23.98 | .0000 |
| Program # 2 | 52 | 37 | 56 | 5.47 | NS |
| Program # 3 | 19 | 21 | 29 | 2.75 | NS |
| Program # 4 | 36 | 25 | 31 | 2.33 | NS |
| Program # 5 | 10 | 16 | 22 | 4.89 | NS |
| Program # 6 | 62 | 60 | 65 | .29 | NS |
| Program # 7 | 54 | 61 | 70 | 3.01 | NS |
| 4. Familiarity With Public Service Efforts | | | | | |
| Service # 1 | 44 | 41 | 66 | 9.77 | .0076 |
| Service # 2 | 43 | 31 | 29 | 3.99 | NS |
| Service # 3 | 44 | 28 | 44 | 5.35 | NS |

variable was calculated for the questions asking about reasons for subjects' actions by totalling the number of answers selected for questions 1 and 2a/2b in Table 7. A second variable for questions asking about subjects' knowledge of programs and public service efforts was created in the same manner, by totalling questions 3 and 4. As can be seen from Table 7 for "reasons" questions only, the SC group selected about 15% more reasons than the average of both computer groups (Duncan, $p<.05$). However, the same analysis for the "knowledge" tendency to select variable demonstrated that the *two computer groups* on average reported significantly higher awareness levels, with about 29% more choices than the SC group (Duncan, $p<.05$). These findings suggest not a simple effect, but an interaction between collection method and question content (see Table 8). The prediction that "use of a computer" will increase the number of choices selected for multiple response questions can only be supported for knowledge type questions.

- TABLE 8 -

Multiple Response Questions:
Mean Number * of Choices Selected

| Questions | CM | SC | NP | F(2,597) | Prob. |
|---|---|---|---|---|---|
| All (1,2a/2b,3,4) | $6.4^{a}$ | $6.61^{a}$ | $6.62^{a}$ | 3 | NS |
| Reasons (1,2a/2b) | $4.21^{a}$ | $4.85^{b}$ | $4.26^{a}$ | 6.5 | .0016 |
| Knowledge (3,4) | $2.19^{a}$ | $1.76^{b}$ | $2.36^{a}$ | 12.3 | .000 |

$^{ab}$ Means with different letters are significantly different (Duncan, $p<.05$).
* F statistic is reported for transformed variables.

### 8.1.4 Open-ended Questions

A comparison of response length (for both number of words and concepts presented) across method of completion was performed for each of the 3 open-ended questions. Totals for all three questions were then added together to calculate the overall number of words and concepts given by each respondent. The results of this analysis are presented in Table 9.

- TABLE 9 -

Average Number of Words * and Concepts *
in Answers to Open-ended Questions

| | CM | SC | NP | F(2,597) | Prob. |
|---|---|---|---|---|---|
| # of Words (log10) | | | | | |
| Question 1 | 6.55[a] | 7.88[a] | **4.92**[b] | 10.97 | .0000 |
| 2 | 5.89[a] | **2.98**[b] | 5.04[a] | 21.93 | .0000 |
| 3 | 6.32[a] | **3.81**[b] | 5.72[a] | 11.50 | .0000 |
| Total | 18.76[a] | **14.67**[b] | 15.68[b] | 5.92 | .0028 |
| # of Concepts (log10) | | | | | |
| Question 1 | 1.63[a] | 2.02[b] | **1.47**[a] | 9.29 | .0001 |
| 2 | 1.43[a] | **.92**[b] | 1.30[a] | 13.24 | .0000 |
| 3 | .96[a] | **.82**[a] | .99[a] | 2.04 | NS |
| Total | 4.02[a] | **3.75**[a] | 3.76[a] | 1.91 | NS |

* It was necessary to take the log of each variable to normalize distributions. For ease of interpretation, the untransformed means are presented, however, the F statistic represents ANOVA of the normalized variables.
[abc] Means with different letters are significantly different (Duncan, $p<.05$).
**BOLD** - lowest of the three means

Subsequent Duncan tests ($p<.05$) showed that while the SC group presented more words and concepts than the computer groups in the first question, both NP and CM groups presented significantly more words in answering questions two and three, and more concepts in answering question two. Overall, the CM group presented more words than the SC, however, this was not accompanied by a significantly greater

number of concepts. The results clearly show that the act of keyboard entry to open-ended questions in and of itself, is not a limiting factor. The prediction that "use of computer" will increase the length of responses to open-ended questions can be supported.

### 8.1.5 Cooperation

Only two persons refused to complete the questionnaire upon being informed that they were assigned to one of the two computer groups. Completion rate was similar for all three groups. Three persons in the NP group did not complete the questionnaire, all for technical reasons (one due to a keyboard error and two due to software 'freeze-ups'). Four persons did not complete the CM questionnaire, and three person in the SC group returned questionnaire in such a short time that all questions could not have been properly read and answered. The high completion rates are probably a function of the short length of the questionnaire; the average completion time was 8.2 minutes for the SC group, 9.6 minutes for the NP group, and 10.9 minutes for the CM group.

Both computer groups required respondents to answer each question before continuing to the next question. Therefore, the only unanswered questions in these groups were a result of the six partially completed questionnaires, leaving less than 1% of all questions unanswered. The SC group, on the other hand, could mark the no opinion or neutral box but often chose not to answer questions. For this group an average of 5.6% of questions per questionnaire were not answered. See Table 10 for a breakdown of unanswered questions.

The SC group was also less likely than both of the computer groups to answer open-ended question at all (Duncan, $p<.05$), answering on average only 1.6 of 3 open-ended questions, while both computer groups answered an average of 2.0 open-ended questions. Table 11 shows the distribution of number of respondents who did not provide any answer to the open-ended questions, by question. As can be seen, respondents in the SC group were less likely to provide any answer to both questions 2 and 3. All of these findings support the prediction that cooperation can be increased by computerized data collection.

- TABLE 10 -

Unanswered Questions: SC Group

| Question # | Missing | Question # | Missing |
|---|---|---|---|
| # children | 3 | pay to park | 17 |
| income | 5 | reduce hours | 22 |
| exhibit A | 8 | knowledge of D | 10 |
| exhibit B | 16 | vacation 1 | 12 |
| exhibit C | 32 * | vacation 2 | 10 |
| increase X | 10 | vacation 3 | 10 |
| increase Y | 11 | vacation 4 | 8 |
| label displays | 12 | vacation 5 | 10 |
| city aware of B | 21 | Computers | |
| quality of info. | 11 | -hands on | 10 |
| annual pass | 4 | -software | 10 |
| increase price | 16 | -understand | 16 |
| increase pass price | 25 | -apprehensive | 20 |
| | | -read about | 20 |
| | | Total Unanswered Questions | 349 |

* Exhibit was not open at time of survey.

- TABLE 11 -

Number of Respondents Not Answering Open-ended Questions
By Method of Completion

| | CM | SC | NP |
|---|---|---|---|
| Question 1 | 15 | 10 | 16 |
| Question 2 | 55 | 117 | 57 |
| Question 3 | 123 | 158 | 122 |
| TOTAL | 193 | 235 | 195 |

Chi-square = 4.28, (4 df), p=.0064

### 8.2.1 Effects of Entry Task on Mean Response to Scale Questions

As can be seen in Table 12, of the four scale questions with significantly different means, this difference was at least partly attributable to differences between the two computer groups in three questions (Duncan, $p<.05$). Performing the same MANOVA analysis that tested prediction 1.1, but including only the two computer groups, showed a significant effect (Wilks $F=2.49, p=.001$). These findings support the prediction that specific differences in the entry task can effect the answers to scale questions.

- TABLE 12 -

Mean Response to Selected Scale Questions

| Question | CM | SC | NP |
|---|---|---|---|
| Increase Exhibit X | $2.23^{a}$ | $2.46^{b}$ | $2.37^{ab}$ |
| Reduce Hours? * | $3.84^{a}$ | $3.56^{b}$ | $3.66^{b}$ |
| Pay More for Extras | $2.92^{a}$ | $2.67^{ab}$ | $2.59^{b}$ |
| Use Travel Agent | $3.75^{a}$ | $3.32^{b}$ | $.25^{b}$ |

* Significance tests were performed on the normalized distribution of this variable, means presented are before transformation.
$^{ab}$ Means with different letters are significantly different (Duncan, $p<.05$).

### 8.2.2 Effects of Entry Task on Range of Response to Scale Questions

As can be seen in Table 6 (Section 1.2), the prediction that cursor movement protocol would decrease the range of response to scale questions is clearly not supported, as the CM group ($m=5.32$) used end points on the scale far more often (Duncan, $p<.05$) than both the NP group ($m=3.76$)and the SC group ($m=3.05$). The prediction that the number pad protocol would increase the range of responses, however, is supported, as the NP ($m=3.76$) group used slightly more end points than the SC ($m=3.05$) group (Duncan, $p<.05$).

The apparent reason for this unexpected finding becomes clear if one looks at the homogeneity of variance tests presented in Table 5. Notice that only six questions have equal variance across all groups. Furthermore, presentation for four of these six used an up-down scale which labelled *all* points on the five point scale.[9] It is important to note that for all other questions, the CM software only permitted labelling of the two ends of the scale ("strongly agree" and "strongly disagree" verses "agree" and "disagree") while both other methods had labels on all scale points. While respondents were given complete verbal instructions regarding how to select "neutral", "agree", and "disagree", it is possible that these instructions were forgotten.

Continuing on this line of logic, if respondents where unaware of the "agree" and "disagree" points then they would make use of more "strongly agree" and "strongly disagree" selections. An examination of the tendency to use end points in testing prediction 1.2 showed this to be the case. As well, if CM respondents were indeed using more "strongly agree/disagree" choices over "agree/disagree" then the number of "agree" and "disagree" choices for the CM group should also be lower. A crosstabulation of the number of times respondents used either the "agree" and "disagree" options against method of completion showed the CM group had a significantly larger number of respondents (47 respondents verses SC=12 & NP=10) who selected "agree" or "disagree" less than 3 times in answering 13 scale questions (Chi-square=80.13,p=.000).

These results clearly show that the cursor entry task, in combination with a presentation difference (the omission of labels for the three middle points of the scale) created a response effect of increasing the likelihood of using end points of the scale. In effect, many of the CM group were answering questions using a three, verses five point scale.

---

9 This was done as these questions required: No Opinion, Poor, Fair, Good, Very Good, and Excellent labels. This variation was not possible using the standard CAPPA scale question format.

### 8.2.3 Effects of Entry Task on Multiple Response Questions

The comparison of overall "tendency to select" multiple response choices performed for prediction 1.3 showed no significant differences among any of the groups. As well, the two computer groups showed no significant differences in the number (NP=4.22, CM=4.25) of selections for either the personal reason for visiting (Duncan p>.05), or knowledge type (NP=2.36, CM=2.19, Duncan p>.05) multiple response questions. The prediction that the software-specific entry protocol will effect the number of choices for multiple response questions is not supported.

## 8.3 Effects of Previous Experience with Computers

At least one quarter of the respondents sampled had no previous hands on experience with computers. To determine if lack of experience increased the likelihood of demonstrating a response bias, a scale to measure respondents' computer experience was created from six questions about computer experience. This scale included:

1) Amount of previous hands on use (0-4).
2) Variety of software used (0-5).
3) Number of courses taken (recoded 0-4).
4) A self-assessment of the respondents understanding of how computers work (1-5).
5) A self-assessment of apprehension about working with computers (1-5).
6) A self-assessment of the respondents effort to read about computers (1-5).

A Cronbach's alpha of .68 was obtained from the pre-test data. For the actual study the alpha increased to .833. In examining the data it was evident that two questions regarding attitudes towards computers should be dropped to form the final scale. The question about reading had a significantly different mean across the three methods of completion ($F(2,573)=6.56, p=.0015$). As well, the question about apprehension had a low correlation to other items. This modification resulted in a four question scale (questions 1 to 4) with a Cronbach's alpha of .846.

This measure was then correlated to measures of each of the response effects

demonstrated, for each of the two computer groups. In order to meet parametric assumptions of normal distributions, the following response effect measure variables were logarithmically transformed: tendency to use extremes, tendency to select, total words presented , total concepts presented, and completion time.

As the correlation of each scale question against previous computer experience showed some positive and some negative relationships, the original strategy of summing the 17 scale questions into one overall "mean-score" variable was felt to be inappropriate. Instead, answers to each of the individual questions were correlated to previous computer experience for both groups. For the CM group only two questions had significant correlations: increasing admission prices (r= -.19, p=.003) and increasing the price of annual passes (r= -.20, p=.003). The NP group had three significant correlations: increasing the price of annual passes (r= -.13, p=.032), vacationing at a regular destination (r= .20, p=.003), and using travel agents to select destinations (r= -.12, p=.042).

- TABLE 13 -

Correlation of Computer Experience and Transformed Response Effect Measures

| | Cursor Movement | | Number Pad | |
|---|---|---|---|---|
| | Pearson R | prob. | Pearson R | prob. |
| 3.2 * Tendency to Use Extremes (log10) | -.128 | .037 | -.099 | NS |
| 3.3 Total Multiple Response Choices selected (log10) | .203 | .002 | .083 | NS |
| 3.4 Total Words Presented (log10) | .111 | NS | .018 | NS |
| 3.4 Total Concepts Presented (log10) | .122 | .045 | .046 | NS |
| 3.5 Completion Time (log10) | -.240 | .000 | -.149 | .019 |

* Numbering identifies the specific prediction being tested.

Table 13 includes the correlations between previous computer experience and measures of response effects for the CM and NP groups. As can be seen the only effect that previous computer experience has on response effects for both computer groups was to reduce the completion time. There were, however, a number of significant relationships between experience and measures of response effects for the group using the cursor movement answering protocol. For this group, more experienced subjects were more likely to provide more answers to multiple response questions, use fewer end points in answering scale questions, and provide marginally more concepts in answering open-ended questions.

It is also possible that the likelihood of demonstrating a response effect is not related to computer experience in a linear fashion, but is a step function relationship, where once a respondent reaches a certain minimum level of experience, response effects are eliminated or substantially reduced. To test for this possibility the computer

- TABLE 14 -

Correlation Between:
Previous Computer Experience (As a Dichotomous Variable)
and Transformed Response Effect Measures

| | Cursor Movement | | Number Pad | |
|---|---|---|---|---|
| | Pearson R | prob. | Pearson R | prob |
| 3.2 * Tendency to Use Extremes (log10) | -.062 | NS | -.089 | NS |
| 3.3 Total Multiple Response Choices selected (log10) | .167 | .010 | .058 | NS |
| 3.4 Total Words Presented (log10) | .086 | NS | -.031 | NS |
| 3.4 Total Concepts Presented (log10) | .081 | NS | -.028 | NS |
| 3.5 Completion Time (log10) | -.268 | .000 | -.145 | .022 |

* Numbering identifies the specific prediction being tested.

experience was recoded into a dichotomous variable. A score of three or less was arbitrarily chosen as the cutoff point as this approximated the percentage of persons reporting no previous hands on usage. Table 14 presents the correlations between the dichotomously coded computer experience and the transformed measures of response effects. This manipulation of the previous computer experience variable actually weakens the relationship, due in part to the loss of power associated with dichotomously coding a continuous variable. While it is clear that people with little or no previous experience will take longer to complete the questionnaire, the prediction that the amount of previous computer experience will effect the likelihood of demonstrating a response effect can only be supported for the cursor movement protocol group.

## 8.4 Part B (Unattended Kiosk)

The unattended kiosk (UK), collected 589 questionnaires over the three week period. Eighty-one of these were not finished leaving a total of 507 useable questionnaires. Responses were then examined for validity, using the response obtained from the pen and paper responses collected from the SC group in Part A as guide. Any responses claiming more than 75 visits to the attraction in the past year (the attraction had a small proportion of quite frequent visitors), or more than 30 courses about computers were eliminated. This left a total of 489 valid responses.

To determine the reliability of the sample collected by the unattended kiosk, it was first necessary to make an estimate of the actual exhibit population, using the sample of all 600 respondents collected in Part A. As the actual visitor attendance fluctuated on a daily basis, it was necessary to weight each day based on the number of visitors that actually passed through the exhibit. This was done by using the daily attraction attendance figures to weight the questionnaires on a day by day basis. Each day's sample weight (wd) was calculated as follows:

$$wd = \frac{\text{daily attendance}}{xv * (\text{sample total} / 3)}$$

where:
xv = % of weekly visits expected on each day based on previous years' attendance.

The sample from the unattended kiosk was weighted as well. While the number of questionnaires collected each day fluctuated, it did not vary in direct proportion to the number of visitors. As well, on those days in which sampling in Part A was taking place, the unattended kiosk was only operational for half of the day. To correct for this, half of the attendance for each of 10 sampling days (Part A) was removed from the daily and total attendance figure. The kiosk group weight for each day's sample (wdk) was:

$$\text{wdk} = \frac{\text{daily attendance / total attendance}}{\text{\# responses / total responses}}$$

where: daily attendance and total attendance were reduced to compensate for 10 periods of in-operation.

A comparison of differences in the distribution of demographic variables showed significant differences on all variables (see Table 15). As hypothesized, the unattended kiosk attracted more younger persons, particularly persons under the age of 18. It also attracted a higher than expected proportion of male respondents. As predicted, the group that self-selected had more hands on experience, and were more likely to be out of town visitors. Finally, persons with higher involvement in the attraction, as measured by the ownership of a seasons pass, were also more likely to use the unattended kiosk than those persons who did not own a membership. Two sampling prediction were not supported as the differences were opposite of the predicted direction. One was with respect to income, where the unattended kiosk attracted more persons with reported household income of under $10,000. The second was a larger than expected proportion of persons with only some high school education. Obviously the ability of an unattended kiosk to estimate the characteristics of a population is severely affected by the self-selection process.

- TABLE 15 -

Comparison of Distributions of Demographic Variables:
Sample Collected by Unattended Kiosk
Verses an Estimate of the Population

**SEX**

| | Pop. | Kiosk |
|---|---|---|
| Male | 41.7% | 48.5% |
| Female | 58.3% | 51.5% |

$Chi^2$=5.3,(2 df), p=.02

**ANNUAL PASS**

| | Pop. | Kiosk |
|---|---|---|
| Yes | 26.1% | 36.8% |
| No | 73.9% | 63.2% |

$Chi^2$=14.9,(2 df),p=.0001

**PLACE OF RESIDENCE**

| | Pop. | Kiosk |
|---|---|---|
| Calgary | 63.1% | 64.4% |
| S. Alta. | 8.5%% | 8.3% |
| N. Alta. | 11.5% | 4.1% |
| B.C. | 5.7% | 5.9% |
| Sask. | 2.7% | 3.8% |
| Canada | 4.4% | 3.6% |
| USA | 2.5% | 5.6% |
| Overseas | 1.5% | 4.3% |

$Chi^2$=37.1,(14 df),p=.000

**INCOME ($000)**

| | Pop. | Kiosk |
|---|---|---|
| < 10 | 6.6% | 15% |
| 10-20 | 10% | 9.5% |
| 20-30 | 13% | 11.4% |
| 30-40 | 15.4% | 11.2% |
| 40-50 | 14.9% | 10.6% |
| 50-60 | 10% | 8.8% |
| > 60 | 17.4% | 17.9% |
| No Answer | 12.7% | 15.6% |

$Chi^2$=31.6,(14 df),p=.000

**COMPUTER EXPERIENCE**

| | Pop. | Kiosk |
|---|---|---|
| None | 26.2% | 13.7% |
| Once/Twice | 11.3% | 13.3% |
| Occasional | 23.2% | 21.9% |
| Regular | 17.4% | 23.7% |
| Extensive | 21.8% | 27.3% |

$Chi^2$=31.2,(8 df),.000

**EDUCATION**

| | Pop. | Kiosk |
|---|---|---|
| Some High | 11.2% | 23.8% |
| High School | 23.6% | 17.6% |
| Some College | 19.1% | 19.2% |
| Trade | 19.1% | 12.9% |
| University. | 17.9% | 16.9% |
| Post Grad. | 9.2% | 9.7% |

$Chi^2$=41.2,(10 df),p=.000

**AGE**

| | Pop. | Kiosk |
|---|---|---|
| < 18 | 1.7% | 20.3% |
| 18-25 | 19.1% | 21.3% |
| 26-35 | 45.8% | 39.8% |
| 36-45 | 17.9% | 12.3% |
| 46-55 | 4.8% | 3.5% |
| 56-65 | 7.4% | 2.0% |
| > 65 | 3.3% | .8% |

$Chi^2$=159.9,(12df),p=.000

# CHAPTER 9

# DISCUSSION

## 9.1 Response Effects for Computerized Completion in General

As the study examined different response effects associated with scale, multiple response and open-ended questions the conclusion is not as straight forward as "yes computer completion causes a response effect" or "no it does not". Numerous factors appear to be at play; some types of questions appear to be effected by computerized collection while others are not, in some cases the specific entry task plays a role while in other cases no differences are seen, and computer experience is relevant for one type of software but not the other. However, some general conclusions for each type of question can be made.

### 9.1.1 Scale Questions

While there were differences for individual scale questions, this study found no consistent novelty or other unidirectional effect where the simple action of using the computer to complete the questionnaire increases or decreases the mean response to scale questions. As well, if differences did occur between groups, they were just as likely to be between the two computer groups, as between each computer group and the pen and paper completion group. While researchers have found respondents to report greater liking for computer verses pen and paper interviews (Newsted, 1985), this study can not support Liefeld's (1986) suggestion that a "novelty effect" will in any way increase agreement to scale questions.

### 9.1.2 Variance In Answers to Scale Questions

Using a number pad entry protocol increases the variance in responses to scale questions over and above traditional pen and paper completion. This conclusion is added to a growing list of similar findings from marketing based studies in a variety of consumer settings (O'Brien & Dugdale,1978; Allen,1987). This increased variance may be due to

the fact that respondents are given no visual clues as to the end points of a scale. *However*, as the CM group is substantially wider than the NP group, it appears that absolute size of the variance of responses is also a function of the specific presentation and entry task protocol, and not purely a function of increased task absorption or engagement, as suggested by Kiesler and Sproull (1986).

### 9.1.3 Multiple Response Questions

No *overall* "video game syndrome" (Liefeld, 1986) was demonstrated for either of the two computer collection methods. Instead, any tendency to select multiple response choices appears to be a function of the content of the multiple response question. Division of multiple response questions based on content did show a "video game syndrome" effect for knowledge/awareness questions, but not for questions soliciting reasons for respondents' actions, where the pen and paper completion group actually selected the greatest number of choices.

Given that this study was using the same software in a similar situation, the inability to replicate Liefeld's quite substantial "video game syndrome" response effect was unsettling. However, further analysis revealed a plausible explanation. In analyzing the distributions of the number of choices selected for each question (rather than the overall mean number of choices selected) it was clearly seen that both computer groups were more likely to pick only one choice for each question (see Table 16). Despite verbal and visual instructions at the beginning and clear written instructions on each screen to "Select As Many As Appropriate" some subjects were still answering pick-n type questions as pick-1 type questions. Possible explanations for this finding are discussed as they relate to the effect of the respondent's previous computer experience (section 9.3).

### 9.1.4 Open-ended Questions

The results of this study suggest the requirement of keyboard entry to answer open-ended questions is not a concern, in fact, the computerized interviewing situation increases the likelihood of answering open-ended questions (particularly later in the questionnaire) as well as the length of these answers. Higgins et al. (1987) came to

similar conclusions for a sample of PC user group, however, this group would have been quite familiar with the technology. The fact that almost 30% of respondents in this study had no previous hands on computer experience, and the lack of a significant correlation between previous experience and length of answers, shows that the requirement for keyboard entry does not hinder even novice users from answering open-ended questions.

- TABLE 16 -

Frequency of Selecting Only One Choice for Multiple Response Questions

| Question | CM | SC | NP |
|---|---|---|---|
| REASONS | | | |
| For Visit | 82 | **22** | 75 |
| Annual Pass | 12 | **6** | 8 |
| No Annual Pass | 120 | **94** | 135 |
| KNOWLEDGE | | | |
| Program | 114 | **50** | 94 |
| Service Efforts | 171 | **165** | 168 |
| TOTAL | 499[a] | **337**[b] | 480[a] |

**BOLD** - least number of *single* choices selected
[ab] Means with different letters are significantly different (Duncan, $p<.05$).

It also appears that though persons are typing more, this is not accompanied by an increase in presentation of more distinct concepts. Counting concepts was originally incorporated into the design in case persons with slow typing speed used fewer words to describe a concept than would be done on paper. But as can be seen by comparing the word/concept ratio (CM = 4.67, SC = 3.91, NP = 4.17), this was not the case, with both computer groups using more words to discuss the same concept. A review of the coded answers showed that often long answers would repeat the same idea with different words.

It is also worthwhile to examine content, what people were talking *about*, rather than just the length of answers. Kiesler & Sproull (1986), for example, demonstrated that computerized administration of open-ended questions resulted in more self centered

answers, than did pen and paper completion. The cooperative nature of this study involved coding the answers to open-ended questions for the benefit of the attraction management. A comparison of the distributions of concept categories across the three methods of completion would provide, while by no means comprehensive, a method to investigate the possibility that collection method does effect the content of answers to open ended questions. However, as is demonstrated by the non-significant Chi-square values in Table 17, the overall distribution of ideas, based on a marketing oriented categorization, is essentially the same for both computer and paper and pen completion methods. These results would suggest that while computerized data collection may effect the *length* of answers to open-ended questions, it will not effect the *content* of answers, or at least content as far as a practitioner is concerned.

- TABLE 17 -

Distribution of Concepts Presented

| | CM | SC | NP |
|---|---|---|---|
| REASON FOR VISITING | | | |
| Friends/Family | 101 | 112 | 96 |
| Tourist/Holidays | 20 | 29 | 21 |
| Weather/Outdoors | 47 | 53 | 44 |
| Activities | 68 | 90 | 60 |
| Exhibits/Attributes | 38 | 58 | 42 |
| TOTAL | 274 | 342 | 263 |

Chi-square=3.15 (8 df),p=NS

| | | | |
|---|---|---|---|
| ADDED TO ATTRACTION? | | | |
| No change | 13 | 6 | 9 |
| More Exhibits | 82 | 58 | 94 |
| Events/Activities | 28 | 22 | 26 |
| Improve Grounds | 50 | 31 | 41 |
| TOTAL | 173 | 117 | 170 |

Chi-square=3.31 (6 df),p=NS

## 9.2 Effect of the Entry Task

For scale questions, the differences between the two computer groups were often as much or more than the differences between each computer group and the pen and paper completion group. There is strong evidence that the entry task will have a significant effect on the range of responses to scale questions, as the cursor movement protocol group selected almost 40% more end points in answering the questionnaire than did the number pad protocol group.

While differences were found between the cursor movement protocol and SC group, they were opposite to the predicted direction. In this study the cursor movement group variance was larger than both the self completion and the number pad group variances, which is inconsistent with Liefeld's (1988) findings. Liefeld used a ten-point scale which required five contacts to move the cursor to the end point, or a repeated movement that resulted if the cursor was held down, whereas this study only required two contacts to get to the end, and less likelihood that new users would notice the repeat movement resulting from holding down the cursor. These differences may have contributed to discrepant findings. How much this result is an effect of differences in presentation, can not be determined, as it appears that the inability to label all scale points on the cursor movement software may have resulted in some of the respondents treating the 5 point scale as if it only had 3 points.[10]

While entry task does appear to be an influencing variable for scale questions, it appears the specific answering protocol has no effect on the number of multiple response choices selected. This is a bit surprising, particularly as the cursor movement entry task used more visual graphics (cursor movement, more colors and background highlighting of selected answers) than the relatively simple number pad entry task (number pad entry into an eight character field). Despite these differences, there was no overall difference between either computer group and the pen and paper completion method. Even in examining the effect for both of the "reasons for action" multiple response questions and the program knowledge questions *both* computer groups selected almost the identical

10 If in fact, the reason for the difference was missing labels, this response effect occurred despite verbal instructions about unlabeled points at the onset of the questionnaire.

number of responses. Perhaps differences would have arisen if fewer respondents in both computer groups had answered the questions as a pick-*1* rather than pick-*N* question.

## 9.3 Effect of Previous Computer Experience

Previous computer experience appears to have little substantial effect on the likelihood of eliciting a response effect. While Lee (1986) showed students with no previous experience to be handicapped by computerized completion of math tests, even dichotomous coding of the computer experience variable did not strengthen the relationship between previous experience and the number of response effects demonstrated. Of the 17 scale questions asked of each group, at most three questions had significant correlations with computer experience, and these were not all in the same direction. Two of the correlations in the CM group: increasing admission and increasing annual pass prices, may be due to the fact that these were the first two of four questions presented together, and that some less experienced users had difficulty in understanding the mechanics of this type of presentation. Two plausible explanations are come to mind, either respondents unknowingly left the cursor at 'strongly disagree', or that frustration in answering was projected into a more negative responses to alternatives.

There are differences though, for each of the two types of software. An increase in previous computer experience was seen to reduce the number of end points selected, increase the number of multiple response questions selected and marginally increase the number of concepts presented in answering open-ended questions, *but only for the cursor movement answering protocol group*. For this group, those persons with less experienced were *more* likely to pick more end points and select less multiple response choices, two response effects that may have been a result of respondents' failure to read or comprehend instructions. Is seems then, that due to the nature of the task, persons in this group with more previous experience were better able to read and comprehend instructions.

The same relationship, however is not demonstrated for the number pad answering protocol. The lack of a correlation between previous experience and the "tendency to select" for the number pad protocol would preclude the explanation that greater previous

experience facilitates the comprehension of instructions, which in turn reduces response effects. Instead, it would suggest that the response effect is inherent in the nature of the entry task; a user that sees one entry field may assume, regardless of instructions presented on the screen, that only one choice is required.

Finally, the reader should also keep in mind that while the relationships found in the cursor movement group are significant, the proportion of variance explained by previous experience ($r^2$) is quite small, with the highest correlation (computer experience with overall tendency to select, utilizing a cursor movement protocol) explaining only 4% of the variance in the number of multiple response choices selected. As well, it was demonstrated quite clearly that completion time is affected by experience and as such, cannot be used as a surrogate measure for cooperation and effort.

## 9.4 Effect of Other Variables

While not in the objectives of this study, two other variables identified in the conceptual framework presented early in this paper (section 4.2) merit some discussion. Post-hoc analysis showed that both presentation (Figure 1, Box D), and question content (Figure 1, Box C) to have a high potential for eliciting bias.

### 9.4.1 Presentation

Entry task and presentation are quite closely related. Yet inconsistencies in the findings (where entry variations created differences in answers to scale questions but not for multiple response questions), suggests examining the potential biasing effect of presentation in isolation. With respect to scale questions, the nature of the questions and inflexibility of the software did not allow for complete control over variations in presentation. As such, the questions could be arbitrarily subdivided, where the potential for variation in presentation to influence results was judged as being relatively equal for all questions within each category. In addition to the written descriptions below, Appendix A has example screen snap shots for each category of question.

Category 1 (9 questions) - This group was judged as the standard presentation group.

Category 2 (4 questions) - Presentation is quite similar for all three methods of completion. Due to the lack of flexibility in the CM software (with respect to labeling scale points) questions in Group 2 required an up-down cursor movement, all points of the scale were labelled, and once selected, answers were back-lighted in red. This made the answering procedure more similar to the NP group protocol, and presentation almost identical for all three groups.

Category 3 (4 questions) - For this group all four questions related to proposed alternatives for dealing with a reduction in operating revenues, with the potential that agreement with one option could effect responses to other options. Both the CM and SC group presented the question so that the respondent could see all four questions at once. The NP software presented the questions one at a time. For all other questions, both computer interviewed groups (NP and CM) presented the questions one at a time.

The means originally presented in Table 4 are reorganized by this arbitrary distinction and presented in Table 18. Note in particular, the lack of significant difference in Category 2, where presentation of all three groups is judged as most similar.[11] Also, in Category 3 the NP group was presented four related questions one at a time, yet this appeared to have no influence on mean responses, as the only significant difference in this group is attributable to the CM group being different than both SC and NP.

To verify this post-hoc prediction, a MANOVA treating each question as an independent variable was performed for each of Categories 1, 2 and 3. This analysis showed a collection method effect for questions in Category 1 (Wilks' F= 2.48, p=.001), an effect approaching significance for questions in Category 3 (Wilks' F=1.88, p=.06), and no collection method effect for questions in Category 2(Wilks' F=1.47, p=.16). This analysis suggests that entry task effects can be offset by reducing the differences in

11 A review of the homogeneity of variance test for each question (Table 5) also shows that none of the questions in group 2 have significantly different variances.

- TABLE 18 -

Mean Response to Scale Questions:
Organization by Presentation

| Question | CM | SC | NP | ANOVA F | ANOVA p.= |
|---|---|---|---|---|---|
| Category 1 | | | | | |
| More of exhibit X | **2.23** | *2.36* | 2.37 | 10.27 | .0059 |
| More of exhibit Y | *2.41* | 2.37 | **2.36** | .13 | NS |
| Label displays? | *2.04* | 1.98 | **1.92** | 3.32 | NS |
| City aware of B? | 2.69 | *2.69* | **2.55** | 3.61 | NS |
| Travel Same Place | *3.22* | 3.14 | **3.07** | 1.24 | NS |
| Money Well Spent | *1.91* | 1.88 | **1.83** | 2.11 | NS |
| Pay More for Extras | *2.92* | 2.67 | **2.59** | 8.43 | .015 |
| Don't Travel Much | *2.57* | 2.50 | **2.38** | 3.02 | NS |
| Use Travel Agent | *3.75* | 3.32 | **3.25** | 21.97 | .0000 |
| Category 2 | | | | | |
| Rate exhibit A | 1.86 | *1.98* | **1.81** | 5.95 | NS |
| Rate exhibit B | **1.57** | *1.63* | 1.58 | .94 | NS |
| Rate exhibit C | **1.55** | *1.76* | 1.67 | 4.44 | NS |
| Quality of Info? | *1.48* | 1.45 | **1.34** | 3.53 | NS |
| Category 3 | | | | | |
| Up Gate Price? | *3.41* | 3.35 | **3.24** | 1.78 | NS |
| Up Annual Pass? | *3.17* | **2.99** | 3.02 | 2.20 | NS |
| Pay to Park? | 3.47 | **3.39** | *3.52* | 2.73 | NS |
| Reduce Hours? | *3.84* | **3.56** | 3.66 | 14.23 | .0008 |

**BOLD** - most positive of three groups

presentation, a conclusion that is supported by findings in human factors research. Switchenko (1984), Cushman (1985), and Askwal (1986) all found that with strict controls for presentation in place, reading comprehension is not effected by the use of CRT verses the traditional ink on paper medium. Finally, presenting related questions one at a time, verses being able to see all questions at once, does not seem to increase the sizeof the response effect over and above existing effects caused by variation in the entry task.

### 9.4.2 Content

There are two findings that suggest question content can also be an important biasing variable. First, the 'video game syndrome' was demonstrated to some extent for the knowledge type questions, but not for those questions asking about reasons for visiting and purchasing an annual pass. Both computer groups gave fewer reasons than did the pen and paper group. Second, differences are seen if questions are divided into two groups, (a) questions about the attraction, and (b) less salient and more subjective questions about vacation travel. The mean response data originally presented in Table 4 is reorganized by this arbitrary distinction based on question content and presented in Table 19. This organization suggests a distinct pattern for the less salient travel questions, but not for questions about the attraction.

- TABLE 19 -

Mean Response to Scale Questions:
Organization by Question Content

| Question | CM | SC | NP | ANOVA F | ANOVA p.= |
|---|---|---|---|---|---|
| ATTRACTION | | | | | |
| More of exhibit X | **2.23** | *2.36* | 2.37 | 10.27 | .0059 |
| More of exhibit Y | *2.41* | 2.37 | **2.36** | .13 | NS |
| Label displays? | *2.04* | 1.98 | **1.92** | 3.32 | NS |
| City aware of B? | 2.69 | *2.69* | **2.55** | 3.61 | NS |
| Rate exhibit A | 1.86 | *1.98* | **1.81** | 5.95 | NS |
| Rate exhibit B | **1.57** | *1.63* | 1.58 | .94 | NS |
| Rate exhibit C | **1.55** | *1.76* | 1.67 | 4.44 | NS |
| Quality of Info? | *1.48* | 1.45 | **1.34** | 3.53 | NS |
| Up Gate Price? | *3.41* | 3.35 | **3.24** | 1.78 | NS |
| Up Annual Pass? | *3.17* | **2.99** | 3.02 | 2.20 | NS |
| Pay to Park? | 3.47 | **3.39** | *3.52* | 2.73 | NS |
| Reduce Hours? | *3.84* | **3.56** | 3.66 | 14.23 | .0008 |
| TRAVEL | | | | | |
| Travel Same Place | *3.22* | 3.14 | **3.07** | 1.24 | NS |
| Money Well Spent | *1.91* | 1.88 | **1.83** | 2.11 | NS |
| Pay More for Extras | *2.92* | 2.67 | **2.59** | 8.43 | .015 |
| Don't Travel Much | *2.57* | 2.50 | **2.38** | 3.02 | NS |
| Use Travel Agent | *3.75* | 3.32 | **3.25** | 21.97 | .0000 |

**BOLD** - most positive of three groups

While there can be no disagreement as to which presentation category questions should be assigned to, any theoretical distinction between questions based on their content should be supported by an empirical distinction in the actual data. However, as the attraction questions were provided by attraction management, and travel related questions were chosen based on the distribution of responses rather than in the interest of combining questions to form a scale measure, a factor analysis reveals the obvious; the 17 questions factor into 7 different factors, with no interpretable groupings. Without this empirical support, it is not reasonable to combine questions into two content based measures of "travel" and "attraction". Needless to say, future research should incorporate content measures pre-tested for empirical support of theoretical distinctions.

## 9.5 Estimate of "Bias Potential" of Variables

Returning to the conceptual framework presented in Figure 1, it is now possible to make some comment on the strength and nature of the effect that each variable has on response effects.

[A] Characteristics of the Respondent: Respondents with less previous experience are more likely to demonstrate biases, but this variable appears to be relevant only in the case of "poor design", where design is a function of presentation and entry task variables. Note that while biases are statistically significant - the percent of variation in responses explained by previous experience is quite small (at most 5%).

[B] Computer as an Interviewer: Other studies have attributed biases to the general activity of "use of a computer". This study has demonstrated the need for more precise definitions, in the same physical setting, the use of different software resulted in differences as substantial as those found between computer and pen and paper completion results. As well, results suggest that detailed and thorough instruction at the beginning of the questionnaire are required to ensure all respondents complete the questionnaire in the proper manner. On the positive side, the novelty of computer use did not translate into

higher means, or any other form of 'video game syndrome', and cooperation was increased.

[C] Content: As in other studies, question content appears to play an important role, particularly through interaction with other variables such as presentation. While answers to factual questions are not subject to computer administered response effects, other content categories that are particularly prone to response effect in the computerized collection are not known at this time.

[D] Presentation : Presentation and the entry task are closely related. In this study, the inability of different entry tasks to elicit response effects for both scale type *and* multiple response questions would suggest that response effects are more prone to variation in presentation rather than the entry task side of the association. As well, an entry task bias should be relatively consistent over the entire questionnaire given a consistent answering protocol, while presentation can change with each type of question and even each screen in the questionnaire. Therefore, presentation may be a greater source of variation than the entry task. This area warrants additional examination, particularly from the field of computer science.

[E] Human Factors: While the results of this study do not explicitly lend themselves to making comment on this variable, some conjectures can be made. Foreign or novel answering processes increase the likelihood of response effects; as the process becomes more similar to traditional questionnaire answering tasks, the likelihood of a response effect is reduced. The action of locating and moving a cursor appears to be more difficult to learn than using a number pad protocol, particularly for the novice users. As the man-machine interaction in computerized questionnaires is not nearly as complicated as in other software or applications, 'cognitive' human factor variables do not appear to be of great concern.

[F] Response Task: Variation in the entry task affected the mean response to scale questions, however, for each of four questions significantly affected, variation in the collection method accounted for only 1.5% to 3.5% of observed variance. From a human factors perspective, once the respondent learns the mechanics of the entry task, the bias potential of this variable should be removed. The key factor for the entry task, is to ensure that respondents understand the mechanics of the answering protocol at the beginning of the interview. As well, a Select-Next, verses Point-Select-Next answering protocol (discussed in detail in section 9.7), appears quicker and less prone to response effects.

[G] Amount of Feedback: The interactive yet structured nature of a computer interview insured cooperation and the elimination of unanswered questions. The tremendous advantages of this effect are reaped at the data coding and analysis stage, particularly in the elimination of missing answers.

[H] Artifact: The very nature of artifacts (unknown variables or influences) makes it difficult to comment on their importance. This study did show how susceptible this type of research is to artifact, exemplified by the failure to label middle points of a scale being a possible explanation for quite substantial difference in use of scale end points.

## 9.6 Implications

Researchers should be aware of the implications of wider range of responses to scale questions that are found in answers to computer administered questionnaires. Extreme agreement or disagreement is usually interpreted to mean strong opinions. Any studies that are counting the number of extreme responses in an absolute, verses relative context will have to adjust the computer collected data due to this tendency.[12] This comment applies to market researchers making sales potential estimates based on Likert scale

12 This particular study for example, found that over the entire questionnaire, the number pad entry group selected about 24% more end points than the pen and paper group, and that the cursor movement group, in turn, selected about 40% more end points than the number pad group.

questions about purchase intentions. Estimates may have to be re-calibrated, otherwise the tendency for computerized questionnaires to elicit extreme responses may result in an overestimation of potential users of the product or service. As well, longitudinal or ongoing research, which switches from traditional pen and paper to more efficient computerized questionnaires should also be aware of the implications of this move.

Inconsistencies among various studies (Liefeld,1988; Sproull,1986; Kiesler & Sproull,1986; Allen,1987), and the failure to predict direction in this study, suggest that variance in responses to scale questions may be a function of mechanical differences in presentation and answering protocols. If the large variances for the cursor movement group in this study are, as proposed, a function of incomplete labeling of all scale points, then it appears that computer questionnaires are much more sensitive to scale presentation and answering protocol than are traditional pen and paper scaling tasks. As such, Menezes and Elbert (1979) findings that different scaling methods do not effect results in traditional self-completion questionnaires are likely not applicable to computerized administration.

Given this *mechanical* effect that various entry tasks have on range of response to scale questions, researchers comparing collection methods effects should be warned against using variance as a surrogate measure for other constructs. As well, the correlation between completion time and previous experience should caution researchers against using completion time as a surrogate measure of effort. Prime examples of these pitfalls are Kiesler and Sproull's (1986) suggestion that increased variance in answers to scale questions is an indication of increased absorption in the task, or Liefeld's (1986) utilization of increased completion times as support for his proposition of a "novelty effect".

## 9.7 Suggestions for Creating Better Questionnaire Software

Given that both computer groups tended to answer the pick-n questions more like pick-1 questions it appears that respondents are less likely to read or understand instructions in the computerized format. As well, the increased variance in answers to scale questions in the cursor movement groups may also be a function of subjects

forgetting verbal instructions. To eliminate response effects resulting from the entry task, it is essential that subjects have a complete understanding of the mechanics and concepts required in the answering protocol. This requires field administrators with the ability to discern the difference between a respondent who is "going through the motions" and one who understands the instructions that are being given. No matter how simple the software, there is no substitute for example questions at the beginning of the questionnaire. These instructions must cover all of the possible actions a respondent can take. For multiple response questions this would include: using help, selecting one choice, selecting more than one choice, un-selecting a choice, and even going back to a previous question to make changes.

User instructions aside, some comments about the design of both software packages can be made. The CM questionnaire presented much more extraneous information on the screen including: a question, a scale labelled at two ends, a command bar that was seldom used, written information that was not necessary to answer questions, a flashing cursor in the middle of the scale, and a three step answering process - (1) point, (2) select, and (3) go to next question. Once the respondent selected their answer it was *possible* to move the cursor to another point on the scale and as such the respondent would be presented with: (1) a mark for their selection, (2) new cursor position if they wished to change their answer, and (3) a small number above the scale which indicated their actual answer. It would be possible to think that their answer was in the "new" cursor position, move to the next question and the computer would (unknown to the user) record the first selection. The point of this drawn out explanation is that a two step selection process: (1) select, then (2) go to next question, may be easier for novice users to understand. Completion times, the greater tendency for biases to exist for the cursor movement protocol software, and field experience in giving respondents assistance would support the contention that a cursor movement protocol is more difficult for users to understand and as such more likely to lead to response effects than a number pad entry protocol.

Research into interactive software design is still very much at an early stage. Using Hill's principles for human-computer interface design (1987, p.115) some generic comments can be made about sound questionnaire software design.

- Make the software easy to use. The selecting protocol should be simple to understand and clearly identify which choice has been selected. Use a neutral colored tape to cover up keys that are not required.

- While the software has to be flexible enough to allow for various question types, an overall consistency is required for answering protocols (i.e. all questions require the ENTER key to proceed), and in the placement of questions, response areas, error and help messages. As well, the the presentation and entry protocol should be consistent with the type of answer required. For multiple response questions, a screen design and protocol that visibly presents more than one entry field *implicitly instructs* the respondent that *more than one answer* can be chosen.

- Use familiar terms and concepts. Do not be afraid to color code or use tape to rename keys such as - Help, Go Back, Next.

- Provide on-line help. Ideally, this help should be in depth and at two levels, one for the question answering protocol and one to elaborate on the actual question being asked. Short help instructions should be automatically called in the event of out of range answers.

- Make the software fail-safe. Young children are excellent test labs. In the IBM PC environment, software should trap interrupts, otherwise a small plastic shim under the CTRL key can stop this key from being depressed.

- One of Hill's recommendation is to allow for the forestalling of prompts, so that that users familiar with the system need not wait for complete screen re-draws to execute their next action. This type of interface, however, *is not at all appropriate* for computerized

questionnaire software, as it leads to novice users getting further and further into trouble with each incorrect press of a key. A NEXT key is required to ensure orderly progress through the questionnaire. Associated with the requirement for a NEXT key is the requirement for a one stroke GO BACK key to be used to verify answers to previous questions.

- In the unattended environment, eliminate *some* unreasonableness checks. The answers to these questions can be later used as a basis to reject deceitful respondents from the sample. For example, by leaving the number of visits per year without a range check, it was possible to eliminate a number of response that were clearly completed by a teenager playing on the machine (for example answers of 9999, 7777, and 888 visits per year). If such reasonableness checks are incorporated in the design, these answers would be rejected at time of entry, the respondent would be forced to re-enter a reasonable response to continue, and the researcher would be left with no mechanism to eliminate these fictitious questionnaires from the sample.

## 9.8 Unattended Kiosks

It was demonstrated quite clearly that an unattended kiosk suffers from the sampling bias associated with allowing respondents to self-select. This is not surprising. The results of Newsted (1985) and data collected from unattended kiosks at Canadian airports (Niman, 1988) suggested that a significantly larger than expected proportion of teenage males would be attracted to the unattended kiosk. It was not clear, however, if a representative sample could be estimated by simply correcting for over-representation of this group. To investigate this possibility, all persons under the age of 18 were removed from the unattended sample, and this adjusted sample (still weighted) was compared against the weighted population estimate described in section 8.4.

As can be seen in Table 20, the unattended kiosk's ability to estimate the population is improved by this adjustment, but is still likely to attract a sample that is significantly younger, has more experience with computers, more likely to own an annual pass to the attraction, and more likely to be from out of town. As well, persons with high

- TABLE 20 -

Comparison of Sample Collected by Unattended Kiosk
(person under 18 removed)
Against an Estimate of the Population

**SEX**

| | Pop. | Kiosk |
|---|---|---|
| Male | 41.7% | 45.5% |
| Female | 58.3% | 54.5% |

$Chi^2$=1.3,(2 df),p=NS

**ANNUAL PASS**

| | Pop. | Kiosk |
|---|---|---|
| Yes | 26.1% | 36.6% |
| No | 73.9% | 63.4% |

$Chi^2$=11.97,(df 1),p=.0005

**PLACE OF RESIDENCE**

| | Pop. | Kiosk |
|---|---|---|
| Calgary | 63.1% | 66.8% |
| S. Alta. | 8.5% | 7.7% |
| N. Alta. | 11.5% | 4.2% |
| B.C. | 5.7% | 7.9% |
| Sask. | 2.7% | 2.6% |
| Canada | 4.4% | 2.6% |
| USA | 2.5% | 4.7% |
| Overseas | 1.5% | 3.4% |

$Chi^2$=26.21,(df 7),p=.0005

**INCOME ($000)**

| | Pop. | Kiosk |
|---|---|---|
| < 10 | 6.6% | 8.9% |
| 10-20 | 10% | 12.1% |
| 20-30 | 13% | 13.6% |
| 30-40 | 15.4% | 14.4% |
| 40-50 | 14.9% | 11.3% |
| 50-60 | 10% | 11% |
| > 60 | 17.4% | 17.8% |
| No Answer | 12.7% | 10.8% |

$Chi^2$=6.33,(df 7),p=NS

**COMPUTER EXPERIENCE**

| | Pop. | Kiosk |
|---|---|---|
| None | 26.2% | 15% |
| Once/Twice | 11.3% | 14.5% |
| Occasional | 23.2% | 20% |
| Regular | 17.4% | 21.3% |
| Extensive | 21.8% | 28.9% |

$Chi^2$=23.3,(df 4),p=.0001

**EDUCATION**

| | Pop. | Kiosk |
|---|---|---|
| Some High | 11.2% | 8.4% |
| H. School | 23.6% | 19.2% |
| Some Col. | 19.1% | 26.8% |
| Trade | 19.1% | 16.6% |
| Univ. | 17.9% | 20.2% |
| Post Grad. | 9.2% | 8.7% |

$Chi^2$=12.25,(df 5),p=.032

**AGE**

| | Pop. | Kiosk |
|---|---|---|
| 18-25 | 19.3 | 29.7% |
| 26-35 | 46.6% | 48.6% |
| 36-45 | 18.1% | 14.4% |
| 46-55 | 4.9% | 3.9% |
| 56-65 | 7.5% | 2.4% |
| > 65 | 3.4% | 1.0% |

$Chi^2$=30.25,(df 5),p=.0000

school education or less are not as likely to complete the questionnaire. It is possible to speculate that the source of this bias is both a combination of the self-selection process (more annual pass holders and more out of town visitors), and the requirement of computer completion (younger, better educated respondents with more hands on experience).

There is an additional concern. In comparing the adjusted distributions in Table 20 against the original distribution in Table 15, it was evident that persons under 18 were entering a large number of fictitious responses. For example, of the 148 persons under the age of eighteen removed from the sample, 22 reported having post-graduate degrees, 19 alleged to be university graduates, 10 had managed to obtain trade school diplomas, and 7 claimed to have had some college courses. The extent that this type of fictitious or deceptive answering applies to other age categories can not be ascertained, nor can teenagers who lied about their age be eliminated from the sample.

The final problem of the unattended kiosk methodology, is with respect to the sampling unit. Most persons who visited the attraction did so in a group, and as this group tends to stay together throughout the visit about 27% of respondents reporting completing the questionnaire in a group or were "coached" by a friend. However, in a group completion circumstance only one person will be able to actually enter answers. If the group is a family, it is likely that parents will let their children perform this task, given high probability that children have used computers at school and will want to be involved. Therefore, the increased proportion of persons under 18 answering the questionnaire may not just be a function of unattended teenagers being attracted to the kiosk, but also of parents allowing their children to operate the keyboard and answer questions. In light of this finding, any unattended kiosk methodology will necessarily have to collect demographic information about other members in the respondent's group.

There is obviously an economic and pragmatic based temptation to use the computerized questionnaire software in an unattended methodology due to the considerable savings in field administration and data entry costs. Although, it appears that this savings would be offset by the necessity of creating a suitable weighting mechanism to adjust for the unrepresentative nature of obtained samples. This limitation, however,

does not entirely rule out data collection through unattended kiosks at attractions. If the demographic distribution of the attraction population is known, questionnaires can be randomly selected from a population base collected by the unattended terminal.[13] Other possibilities include the use of unattended kiosks in pre-testing or exploratory type research.

While sampling biases limit the generalizability of absolute measures, this hindrance should not stop an unattended kiosk from detecting a change in levels over time. Used in this way, unattended kiosks could be used as a "quality barometer", to monitor relative change of evaluation of exhibits, concessions, and attraction staff performance on an ongoing basis. The immediacy of feedback from computerized data collection methods makes it suitable as an on-line (granted expensive, but certainly sophisticated) suggestion box. Attractions that are quite heavily into computerized technology in existing exhibits may find the inclusion of visitor questionnaires a logical extension of existing resources and in line with the image and goals of the attraction.

## 9.9 Limitations

A few comments about the generalizability of these findings are required. With many potentially influential variables, the ability to control for unwanted artifacts was difficult. For example, testing for entry task effects required the use of two types of software, which in turn used different screen designs. While attempts were made to keep presentation equal for both software packages, such things as use of colors, background instructions, location of question area, and layout of response area would necessarily be somewhat different. As well the very nature of a cursor response task (five alternative entry fields) requires different presentation than a number pad entry response task (which may have five selection markers but only one entry field). There is therefore a practical difficulty in separating presentation from the entry task. As such, all conclusions about effects of the entry task will have some element of presentation influence and conversely, presentation effects will have an element of entry-task influence. The conceptual distinction between the two variables is still quite necessary, however, as presentation can

13 This procedure was not appropriate in this study, as the unattended questionnaire did not include all of the questions asked in the questionnaire administered in part A.

vary from question to question, while one response task protocol or set of protocols will be used for all questions in the questionnaire.

Though unlikely, it is possible that some of the response effects demonstrated in this study were a function of insufficient verbal and visual instructions. During administration, respondents appeared to understand the instructions. As well, the field staff made explicit efforts to be precise, complete and consistent, particularly in demonstrating how to select middle points for the cursor movement protocol. Each respondent was also shown how to select more than one response on the multiple response questions. The field experience gave no indication that either one of the two response effects (increased use of extremes in cursor movement group, and computer completion tendency to select only one choice to pick n questions) were occurring, nor did they show up in analysis of pre-test data. Nevertheless it is recognized that results could have been partially a function of respondents not knowing that they could select unlabeled scale points, or more than one response to pick-n questions. The extent that these same biases would have been demonstrated if respondents were given two or three examples of each type of question at the onset of the questionnaire is not known. The more important issue though, is to recognize that these response effects did occur, and could quite easily be replicated by unknowing researchers in similar situations.

## 9.10 Future Studies

More complete and precise techniques for identifying and measuring response effects associated with computerized data collection are required. A simple comparison of means, variance and frequency of answers across collection methods may identify some response effects but totally miss others. For example, in this study examining the absolute number of choices to multiple response questions showed only a marginal effect. It was only a post-hoc analysis of the distribution of responses to each question that discovered the computer groups' tendency to select only one answer to pick-n type questions.

For scale questions, collection method appears to affect independent questions in different directions, making the creation of an overall measure quite difficult. Just as

psychologists have validated batteries of tests (eg. Katz & Dalby,1981a; Hurrell & Lombardo, 1984; Russell et al.,1986, Greist et al.,1987) future studies should incorporate scales commonly utilized in the marketing research literature. The very premise of a reliable scale is that all questions in the scale can be combined to create one measure. Therefore a comparison of results would either validate the use of this scale in a computerized interviewing situation, or be able to make a statement about effect that collection method has on the mean response to scale questions.

This study, like others in the area, was performed on a cooperative basis where questions were derived for the benefit of interested third parties. While cooperative research has it's benefits, researchers should begin to use more questions where the expected response distributions are known. For example, examination of the acquiescence response set could be performed by including a group of questions where the expected response distribution is quite negative. Another possible method to detect acquiescence is to utilize the ability of the computer to measure response latency, where the fastest possible time to read and make a judgement on the questions could be used to locate the computer equivalent of "random ticking" of responses on a paper questionnaire.

A detailed examination of the effects of variation in presentation is required. Given the inability of changes in the entry task to consistently elicit response biases, it appears that the presentation component, rather than the mechanical entry task of various software protocols may be an important variable. As well, presentation can vary from screen to screen, whereas the entry task is relatively constant throughout the entire questionnaire. In particular the use of various presentation formats in combination with a cursor movement protocol to answer scale questions deserves another examination. Of particular interest would be the effect that 5, 7, and 10 point scale questions have on the variance of responses. As well, a post-hoc analysis in this study showed that presenting questions one at a time was not a problem, but this contention requires more rigorous testing. It may also be interesting to compare the results of a protocol that allowed respondents to back up and re-answer previous questions, verses one that did not.

A conceptual framework for the effect of question content appears in order. Almost all studies to date have shown variation due to content, yet even in traditional

forms of research the relationship between bias and content is not clearly understood. Indications from this study are that less salient subjective questions may be more prone to response effects than judgmental questions. A potential starting point for this framework may be a distinction based on: a) degree of saliency, b) evaluation of personal importance, c) judgement of attributes (both tangible and abstract), and d) personal judgment of awareness levels.

# CHAPTER 10

## SUMMARY

This study has attempted to organize cross-disciplinary research into a conceptual framework of factors that contribute to the validity and reliability of data collected by computerized questionnaires. This framework can be used to more effectively identify and measure specific response effects that may be associated with this relatively new data collection methodology. Two of the variables, characteristics of the entry task, and the respondent's previous computer experience were examined in a field experiment to determine the role that each played in creating or contributing to response effects.

The study demonstrated that computerized completion of questionnaires elicited four response effects: an increase in the range of response to scale questions, a variable effect on the means of scale questions, a tendency to select only one answer for multiple response questions, and longer answers to open-ended questions. By using two different software packages the study was also able to demonstrate that response effects are as much a function of variation in characteristics of the computer specific entry task, as they are a function of the computer interviewing situation in general. The respondent's previous experience with computers was shown to have only a marginal effect on the size of response effects, and this effect was only evident for one of the two answering protocols.

The results indicate that the variables of presentation and question content are areas worthy of future examination. The field of computer science is looked to for leadership in the examination of presentation, particularly as it relates to the principle of sound interface and screen design. A conceptual framework for question content appears to be the first step towards an identification of specific content characteristics that increase the likelihood of eliciting response effects in the computer interviewing situation. To date, only factual questions have not shown any susceptibility in this regard.

The study also examined the results of a methodology which used computerized questionnaire software at an unattended kiosk. While the limitation of this unattended

kiosk to adequately sample the population was clearly shown, the derivation of suitable weighting mechanisms can potentially overcome this limitation. As well, the technology is quite suitable for use as quality of service "barometers", or in pre-testing and exploratory research, and should find a role to play in the research systems of tourism businesses, marketing research organizations and academic institutions.

Finally, the reader should be convinced of the irony that research attempting to discover response effects, is itself, very prone to artifact. The requirement for strict controls in future studies goes without saying

# REFERENCES


Allen, David F. (1987), "Computers Verses Scanners: An Experiment in Nontraditional Forms of Survey Administration", Journal of College Student Personnel, May, pp. 266-273.

Askwal, Susan. (1985), "Computer Supported Reading vs Reading Text on Paper: A Comparison of Two Reading Situations", International Journal of Man-Machine Studies, v. 22, pp. 425-439.

Bahner, Lesley. (1987), "Long Self Administered Questionnaires" in Proceedings of the Sawtooth Conference on Computer Interviewing, Perceptual Mapping & Conjoint Analysis, 1987.

Barron, Mark. (1987), "The Effects of Computer-conducted Verses Councillor-conducted Initial Intake Interviews on Client Expectations", Dissertation, U. of Southern Mississippi.

Beard, J.W. & Peterson, T.O. (1987), "A Taxonomy for the Study of Human Factors in Management Information Systems", Human Factors in MIS, B. Shneiderman (ed.) (in press).

Bernard, H., Kilworth, P. & Sailer, L. (1981), "Information Accuracy in Social Network Data V", Presented at Sunbelt Social Networks Conference, Tampa, Florida, 1981.

Bradburn, Norman. (1983), "Response Effects", Handbook of Survey Research, ed. Rossi, P., J. Wright, A. Anderson, Academic Press, New York, 1983.

Brewer, B.W. (1985), "Randomized Response Techniques: the Canberra Drug Survey", in Survey Interviewing: Theory and Techniques, Terence Beed and Robert Stimson, eds., George Allen and Unwin, Sydney, 1985.

Cannell, C.F. (1985), "Overview: Response Bias and Interviewer Variability in Surveys", in Survey Interviewing: Theory and Techniques, Terence Beed and Robert Stimson, eds., George Allen and Unwin, Sydney, 1985.

Canoune, Howard L. & Edward Leyhe. (1985), "Human Verses Computer Interviewing", Journal of Personality Assessment, Feb 1985, 49(1), pp.103-106.

Cantril H., & Research Associates. (1944), Gauging Public Opinion, Princeton, N.J., Princeton University Press.

Carr A., A. Gosh & R.J. Ancill. (1983), "Can a Computer Take a Psychiatric History?", Psychological Medicine, 13, pp.151-158.

Converse, Jean M. (1986), Survey Questions: Handcrafting the Standardized Questionnaire, Sage Publications: Beverly Hills, CA.

Couch, A. & K. Kenniston. (1960), "Yeasayers and Naysayers, Agreeing Response Sets as a Personality Variable", Journal of Abnormal Psychology, 60, March 1960, pp.171-174.

Crowne, D.P. & D. Marlowe. (1964), The Approval Motive. New York:Wiley, 1964.

Cushman, W. H. (1986), "Reading From Microfiche, a VDT, and the Printed Page: Subjective Fatigue and Performance," Human Factors, 28, pp.63-73.

Dillman, Don. (1978), Mail and Telephone Surveys: The Total Design Method, John Wiley & Sons, New York, 1978.

Duffy, John C & Jennifer Waterton. (1984), "Underreporting of Alcohol Consumption in Sample Surveys: The Effects of Computer Interviewing on Fieldwork", British Journal of Addiction, September 1984, 79(3), pp.303-308.

Eaves, Ronald & Earl Smith. (1986), "The Effect of Media and Amount of Micro-computer Experience on Examination Scores", Journal of Experimental Education, 55(1), Fall 1986, pp.23-26.

Eckhouse, John. (1986), "Interviewing by AT", PC World, July 1986, 4(7), pp. 286-288.

Erdman, H., M. Klien, & J. Greist. (1983), "The Reliability of Computer Interviews for Drug Use/Abuse Information", Behavior Research Methods and Instrumentation, (15), pp.66-68.

Erdman, H., M. Klien, & J. Greist. (1985), "Direct Patient Computer Interviewing", Journal of Consulting Psychology, Dec, 53(6), pp. 760-773.

Fiske, Ted. (1989), Radix Software Inc., Calgary, Alberta.

Franzkowiak, Michael & Pierre Korber. (1985), "Video Questionnaire: Telematique et Marketing", Revue Francaise du Marketing (France), 101, Jan/Feb 1985, pp. 73-82.

Gorn, J. & Scott Frazer. (1988), "Electronic Data: Marketing Answers Through Behavioral Research", Singapore Marketing Review, vol. III.

Green, P.E., P.K. Kedia & R.S. Nikhil (1985), Electronic Questionnaire Design and Analysis with CAPPA, Scientific Press, Palo Alto, CA.

Greist, John H., M. Klien, H.P. Erdman, & J. Bires (1987) "Comparison of Computer and Interviewer Administered Version of the Diagnostic Interview Schedule", Hospital and Community Psychiatry, Dec. 38(12), pp. 1304-1311.

Greenwood, H. Paul, C.A. Higgins, & Tony Dimnik. (1987), "The Micro Survey Technique", Journal of Systems Management, 38 (6), June, pp. 13-19.

Groves, R. M. (1979), "Actors and Questions in Telephone and Personal Interview Surveys", Public Opinion Quarterly, 43(2), Summer, pp. 190-205.

Groves, Robert & Nancy Mathiowitz. (1984), "Computer Assisted Telephone Interviewing: Effects on interviewers and respondents", Public Opinion Quarterly, Spring, (48), pp. 356-369.

Hansen, W. & C. Haas. (1988), "Reading and Writing with Computers: A Framework for Explaining Differences in Performance." Communications of the ACM, September, 31(9), pp. 1080-1089.

Harris, R.J. (1975), Primer of Multivariate Statistics. Academic Press: New York.

Higgins, C., T. Dimnik, & H. Greenwood. (1987), "The Diskq Survey Method", working paper, School of Business Administration, University of Western Ontario, 1987.

Hiltz, S. R. (1979), "Computerized Conferencing for Opinion Research", Public Opinion Quarterly, (43), pp. 562-571.

Hurrell, Thomas & Theodore Lombardo. (1984), "Validation of an Automated 16PF Administration Procedure", Journal of Personality Assessment, Dec 48(6), pp. 638-42.

Katz, Larry & J. Thomas Dalby. (1981a), "Computer and Manual Administration of the Eysenck Personality Inventory", Journal of Clinical Psychology, July 37(3), pp. 586-588.

Katz, Larry & J. Thomas Dalby. (1981b), "Computer-Assisted and Traditional Psychological Assessment of Elementary-School-Aged Children", Contemporary Educational Psychology, (6), pp. 314-322.

Kiesler, S. & L. Sproull. (1986), "Response Effects in the Electronic Survey", Public Opinion Quarterly, 1986, (50), pp. 402-413.

Lee,Jo-Ann. (1986), "The Effects of Past Computer Experience on Computerized Aptitude Test Performance.", Educational and Psychological Measurement; 46(3), Autumn, pp.727-33 .

Licker, Paul, P. Newsted & B. Sheehan. (1986), "A Value Added Model of Electronic Mail Utilization", working paper 07-86, University of Calgary, 1986.

Liefeld, John P (1986), "Computer/Personal/Self Administered Questionnaires: Response Effects and Field Experiences", University of Guelph, Department of Consumer Studies, Working Paper 86-106, Aug. 1986.

Liefeld, John P. (1987a), "Field Experience: Computer Interviewing With Consumer Samples", Proceedings of the American Marketing Association, Microcomputer Workshop, University of Hawaii, Nicolaos Synodinos (ed.), October 1987, pp. 129-140.

Liefeld, John P (1987b), "Computer interviewing: Cursor and Response Task Effects", Canadian Journal of Marketing Research, 1987 (6), pp. 20-25.

Liefeld, John P (1988), "Response Effects in Computer-Administered Questioning", Journal of Marketing Research, Nov 1988 (25), pp. 405-409.

Locander, William, S. Sudman & N. Bradburn. (1976), "An Investigation of Interview Method, Threat, and Response Distortion", Journal of the American Statistical Association, 1976, 71(354), pp. 269-275.

Loftus E. F. (1982), "Interrogating Eyewitnesses - Good Questions and Bad", in Question Framing and Response Consistency , ed. R. M. Hogarth, Jossey-Bass, San Francisco, 1982.

Lucas, R.W., P. Mullins, C. Luna, & D. McInroy. (1977), "Psychologists and Computers as Interrogators of Patients with Alcohol Related Illnesses: A Comparison", British Journal of Psychiatry, 131, pp. 160-67.

Mardia, K.V. (1971), "The Effect of Non-normality on Some Multivariate Tests and Robustness of Non-normality in the Linear Model, Biometrika, 58(1), pp. 105-121.

Marketing News (1985a), "Polling Pole Polls People", January 4, p. 44.

Marketing News (1985b), "Stores, Campuses use Poles for Polls", May 24, p. 25 & 32.

McGuire, William. (1969), "Suspiciousness of Experimenters Intent", Artifact in Behavioral Research, ed. by Rosenthal and Rosnow, New York, Academic Press, 1969.

Menezes D. & N.F. Elbert. (1979), "Alternative Semantic Scaling Formats for Measuring Store Image: An Evaluation", Journal of Marketing Research, (16)1 February, pp. 80-89.

Millstein, Susan G. (1987), " Acceptability and Reliability of Sensitive Information Collected via Computer Interview." Educational-and-Psychological-Measurement; 47(2), pp. 523-33.

Minno, Jim. (1986), "Software Replaces Paper in Questionnaire Writing/Software Updated to Meet Growing Market", Marketing News, 20(1), Jan. 3, pp. 66-67.

Muter, P., S. A. Latremouille, W. C. Treurniet & P. Beam. (1982), "Extended Reading of Continuous Text on Television Screens", Human Factors, 24, pp. 501-508.

Nederhof, A.J. (1985), "Methods of Coping with Social Desirability Bias", European Journal of Social Psychology, (15), 1985, pp. 263-280.

Nederhof, A. J. (1987), "De Grote Onbekende: Een Evaluatie de Computer als Dataverzamelingsinstrument in Survey-onderzoek" (An Evaluation of the Computer as a Data Collection Instrument in Survey Research), Sociologishe, 87/2 (34).

Newsted, P. (1985), "Paper Versus On-line Presentations of Subjective Questionnaires", International Journal of Man-Machine Studies, 23, pp. 231-247.

Niman, Grant. (1988), "Demographics of Persons Responding to Statistics Canada Terminal Kiosks at Four Airports". Unpublished spreadsheets, Statistics Canada, Ottawa, 1988.

Okimoto, Glenn. (1988), "Computer Interactive Data Collection at Honolulu International Airport: Unattended Kiosk Interviewing", paper presented at the 1988 Sawtooth Software Conference on Computer Interviewing.

O'Brien, T. & Dugdale V. (1978), "Questionnaire Administration by Computer", Journal of the Market Research Society, v. 20(4), pp. 228-237.

Pershel, Barbara M. (1972), Indexer Consistency in Perception of Concepts and in Choice Terminology. U.S. Department of Health, Education & Welfare (Office of Education): Washington D.C., 1972.

Resmovic, Victor. (1977), "The Effects of Computerized Experimentation on Response Variance", Behavior Research Methods and Instrumentation, 1977, 9(2), pp. 144-47.

Rosenberg, Milton J. (1969), "The Conditions and Consequences of Evaluation Apprehension" in Artifact in Behavioral Research, New York, Academic Press, 1969.

Rosenthal, R. (1967), "Covert Communication the Psychological Experiment", Psychological Bulletin, 1967, (67), pp. 356-367.

Rosenthal, R. & Fode. (1963), "The Effect of Experimenter Bias on the Performance of the Albino Rat", Behavior Science, 1963, (8), pp. 183-189.

Rosenthal, R. & R.L. Rosnow. (1969), Artifact in Behavioral Research, New York, Academic Press, 1969.

Rushinek, A., S. Rushinek & J. Stutz. (1983), "An Empirical Analysis of the Computerization of Interactive Software Systems", Information and Management, (6), 281-287.

Russell, G. K., K. Peace & G. Mellsop. (1986), "The Reliability of a Micro-computer Administration of the MMPI." Journal of Clinical Psychology, January 1986, 42(1), pp. 120-122.

Skinner, H.A. & B.A. Allen. (1983), "Does the Computer Make a Difference? Computerized verses Face to Face, verses Self Report Assessment of Alcohol, Drug and Tobacco Use", Journal of Consulting and Clinical Psychology, 1983, 51, pp. 267-275.

Slack, W.V., G. P. Hicks, C. E. Reed, L. Van Cura. (1966), "A Computer Based Medical History System", New England Journal of Medicine, 274, pp. 194-198.

Sloan, Hugh. (1988), University of Ohio, personal communication.

Sproull, L. (1986), "Using Electronic Mail for data collection in Organizational Research", Academy of Management Journal, 1986, 29, pp.157-169.

Sudman, S. & N. Bradburn. (1974), Response Effects in Surveys: A Review and Synthesis, Aldine Publishing Company, Chicago, 1974.

Sudman, S. & N. Bradburn. (1979), Improving Interview Method and Questionnaire Design: Response Effects to Threatening Questions in Survey Research, Jossey-Bass, San Fransisco, 1979.

Switchenko, D. M. (1984), "Reading From CRT Versus Paper: the CRT - Disadvantage Hypothesis Re-examined", Proceedings of the Human Factors Society -- 28th Annual Meeting, pp. 429-430.

Tabachnick, Barbara G. & Linda Fidell. (1989), Using Multivariate Statistics. Harper & Row Publishers: New York.

Tapscott, D. (1982), Office Automation: A User Driven Method, Plenum: New York.

Teaborg, J., G. Howard & S. Maxwell. (1980), "Evaluating Planned Organizational Change: A Method for Assessing Alpha Beta and Gamma Changes", Academy of Management Review, 1980, 5(1), pp. 109-121.

Tourism Canada. (1983), Canadian Tourism Attitiude and Motivation Study, Tourism Canada: Ottawa, Canada, 1983.

Trollip, S. R. & G. Sales. (1986), "Readability of Computer Generated Fill-Justified Text", Human Factors, 28, pp. 159-163.

Warner, S.L. (1965), "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias", Journal of the American Statistical Association, 1965, 60, pp. 63-69.

Wright, P. & A. Lickorish. (1984), "Investigating Referee's Requirements in and Electronic Medium", Visible Language, 18, pp. 186-205.

# APPENDIX A

Screen Snap Shots of Question Presentation

## *MULTIPLE RESPONSE QUESTIONS*

Paper Completion (SC group)

Which of the following are your reasons for visiting today?
(PICK AS MANY AS APPLY)

☐ For Fresh Air
☐ To Bring the Children
☐ To Bring Out of Town Visitors

Legend
☐ represents boxes which the respondents can tick off.

Maze (NP group)

ACTION: Respondent has selected "Fresh Air" and "To Bring Visitors"

Which of the following are your reasons for visiting today?
(PICK AS MANY AS APPLY)

[1] For Fresh Air
[2] To Bring the Children
[3] To Bring Out of Town Visitors

Please Enter the appropriate numbers: 13_____

Legend
_____ represents an eight character input field.

## APPENDIX A - CONTINUED

Cappa (CM group)

ACTION: Respondent has Selected "Fresh Air" and is about to select "To Bring Out of Town Visitors"

Which of the following are your reasons for visiting today?
(PICK AS MANY AS APPLY)

| | | |
|---|---|---|
| | **X** | **For Fresh Air** |
| | X | To Bring the Children |
| --> | X | To Bring Out of Town Visitors |

Legend

X possible cursor locations

--> present cursor position

**X** selection that has already been chosen

## APPENDIX A - CONTINUED

### *SCALE QUESTIONS*

Paper Completion (SC group)

The attraction should increase the number of type A Exhibits.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

Legend
☐ represents boxes which the respondents can tick off.

Maze (NP group)

ACTION: Respondent has not yet made a selection.

The attraction should increase the number of type A Exhibits.

[1] Strongly Agree
[2] Agree
[3] Neutral
[4] Disagree
[5] Strongly Disagree

Please Enter your Response: _

Legend
_ one character input field.

## APPENDIX A - CONTINUED

Cappa (CM group) - Category 1 per section 8.4.1

ACTION: Respondent has select "disagree".

The attraction should increase the number of type A Exhibits.

| Strongly Agree | | | | Strongly Disagree |
|---|---|---|---|---|
| 𝕏 | 𝕏 | 𝕏 | **X** | 𝕏 |

Legend
𝕏 possible cursor positions.
**X** a selection that has already been chosen

Cappa (CM group) - Category 2 per section 8.4.1

ACTION: Respondent has selected "Very Good".

How would you rate the overall quality of Type B exhibits?

| | | |
|---|---|---|
| | 𝕏 | Excellent |
| --> | **X** | Very Good |
| | 𝕏 | Good |
| | 𝕏 | Fair |
| | 𝕏 | Poor |

Legend
𝕏 possible cursor locations
--> present cursor position
**X** a selection that has already been chosen

## APPENDIX A - CONTINUED

Cappa (CM group) - Category 3 per section 8.4.1

ACTION: Respondent is neutral to a price increase and strongly disagrees with reducing hours.

To deal with a reduction in operating revues
how would feel about the following alternatives:

| | Strongly Agree | | | | Strongly Disagree |
|---|---|---|---|---|---|
| Increase Price | 𝕏 | 𝕏 | **X** | 𝕏 | 𝕏 |
| | Strongly Agree | | | | Strongly Disagree |
| Reduce Hours | 𝕏 | 𝕏 | 𝕏 | 𝕏 | **X** |

Legend
𝕏 possible cursor positions.
**X** a selection that has already been chosen

# APPENDIX B

## Questionnaire (Disguised)

The University of Calgary is performing a survey of visitors to this Attraction. We would like some basic information about yourself,as well as your feelings about the attraction, along with your attitudes about vacations in general. Thank you for your help and enjoy the rest of your visit.
The questionnaire will take about 10 minutes to complete.

==========

[1] What is your sex?

- ☐ Male
- ☐ Female

[2] Which of the following age groups do you fall into?

- ☐ Under 18
- ☐ 18-25
- ☐ 26-35
- ☐ 36-45
- ☐ 46-55
- ☐ 56-65
- ☐ Over 65

[3] What is your normal place of residence?

- ☐ Calgary
- ☐ Southern Alberta
- ☐ Northern Alberta
- ☐ British Columbia
- ☐ Saskatchewan
- ☐ Other Canada
- ☐ USA
- ☐ Overseas

[4] How many dependent children under the age of 18 do you have?

- ☐ None — If NONE, GOTO QUESTION # [7]
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4 or more

[5] Do you work outside the home?

- ☐ Yes, full time.
- ☐ Yes, part time
- ☐ No — If NO, GO TO QUESTION # [7]

[6] If YES, Which of the following outside services do use to care for your children?

- ☐ Nanny
- ☐ Day Care
- ☐ Baby-sitter
- ☐ Day Home
- ☐ None of the above

[7] What is your combined household income (before tax)?

- ☐ Less than 10,000
- ☐ 10,000 - 19,999
- ☐ 20,000 - 29,999
- ☐ 30,000 - 39,999
- ☐ 40,000 - 49,999
- ☐ 50,000 - 59,999
- ☐ Over 60,000
- ☐ Prefer not to answer.

[8] What is your highest level of education?

- ☐ Some High School
- ☐ Finished High School
- ☐ Some College or University
- ☐ Technical /Trade School or College Diploma
- ☐ University Degree
- ☐ Post Graduate Work

[9] How many times have you attended the Attraction in the past twelve months? (including this visit) _____

[10] Why did you visit the Attraction today?

_____

_____

_____

_____

_____

For the following questions place an **X** in the box that best represents your opinion.

How would you rate the OVERALL QUALITY of the:

| | No Opinion | Excellent | Very Good | Good | Fair | Poor |
|---|---|---|---|---|---|---|
| [11] **TYPE A EXHIBITS.** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| [12] **TYPE B EXHIBITS.** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| [13] **TYPE C EXHIBITS.** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

To what extent do you agree or disagree with the following statements:
With respect to TYPE A EXHIBITS:

[14] The Attraction should increase the number of **Category X** exhibits.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

[15] The Attraction should increase the number of **Category Y** exhibits.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

With respect to Exhibit B:

[16] All displays, both inside and out, should be labelled and described.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

[17] **Calgarians** are well aware of Exhibit B at the Attraction.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

For the next question place an **X** in AS MANY BOXES AS APPLY.

[18] Which of the following are reason(s) that **you** visit the Attraction?

- ☐ To increase my knowledge.
- ☐ To take part in special events.
- ☐ To bring children for a family outing.
- ☐ A good place to bring out of town visitors.
- ☐ For fresh air and exercise.
- ☐ To take part in programs.
- ☐ To support public service efforts.

[19] What thing(s) would you like to see added to the Attraction?

_____

_____

_____

_____

_____

[20] How difficult was it to **obtain information** about programs and activities at the Attraction?

| ☐ | ☐ | ☐ | ☐ | ☐ |
|---|---|---|---|---|
| Haven't Tried | Very Easy | Somewhat Easy | Somewhat Difficult | Very Difficult |

[21] With which of the following Attraction programs are you reasonably familiar?
(YOU MAY CHECK MORE THAN ONE BOX.)

- ☐ Group Tours.
- ☐ Program 1.
- ☐ Program 2.
- ☐ Program 3.
- ☐ Program 4.
- ☐ Annual Memberships.
- ☐ Program 5.

[22] Which of the following public service efforts have you heard of?
(YOU MAY CHECK MORE THAN ONE BOX.)

- ☐ Public Service Effort 1.
- ☐ Public Service Effort 2.
- ☐ Public Service Effort 3.
- ☐ None of the above.

[23] Do you currently own an annual pass?

☐ Yes ☐ No If NO, GO TO QUESTION # **[25]**

FOR ANNUAL PASS HOLDERS:

[24] Which of the following are the **most important** reason(s) in your decision to buy a pass. (YOU MAY CHECK MORE THAN ONE BOX.)

☐ The annual admission pass is a more economical way to visit the Attraction.

☐ Discounts on purchases.

☐ Reason 1.

☐ To get Information on programs and activities

☐ I believe in supporting the Attraction.

☐ Reason 2.

GO TO QUESTION # **[26]**

For PERSONS who DO NOT have an annual pass:

[25] Which of the following are your reason(s) for not joining? (YOU MAY CHECK MORE THAN ONE BOX)

☐ Didn't know that annual pass existed.

☐ Not interested.

☐ Annual pass is too expensive.

☐ Out of town visitor.

☐ Do not visit the Attraction enough to make it worthwhile.

☐ Hadn't really considered it.

As you may be aware, in 1989 operating revenues will be reduced. Please indicate your feelings about the following possible methods to deal with this situation.

| | | | | | |
|---|---|---|---|---|---|
| [26] **Increase admission prices.** | ☐ Strongly Agree | ☐ Agree | ☐ Neutral | ☐ Disagree | ☐ Strongly Disagree |
| [27] **Increase annual pass. prices.** | ☐ Strongly Agree | ☐ Agree | ☐ Neutral | ☐ Disagree | ☐ Strongly Disagree |
| [28] **Pay for parking.** | ☐ Strongly Agree | ☐ Agree | ☐ Neutral | ☐ Disagree | ☐ Strongly Disagree |
| [29] **Reduce hours of operation.** | ☐ Strongly Agree | ☐ Agree | ☐ Neutral | ☐ Disagree | ☐ Strongly Disagree |

[30] Do you know how much of revenues come from gate admission?

- ☐ Don't know.
- ☐ About 30%.
- ☐ About 60%.
- ☐ About 90%.

[31] Do you have an other comments or suggestions for the Attraction?

_____

_____

_____

_____

_____

This Attraction is visited by Calgarians and out of town visitors. In order to better service both groups, we need to know your feelings about **VACATION PLEASURE TRAVEL.** Please indicate how much you agree or disagree with the following statements.

[32] I usually choose vacation places where I have been before.

| Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

[33] For me, money spent on travel is well spent.

| Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

[34] I think it's worth paying more to get luxuries and extras on a vacation.

| Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

[35] I don't have to travel to enjoy a vacation.

| Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

[36] I sometimes use a travel agent to help me select a vacation destination.

| Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

As some people will be completing this questionnaire using a personal computer we need to know a little bit about your experience with computers.

[37] How much "hands-on" experience have you had with any type of computers? DO NOT INCLUDE EXPERIENCE WITH COMPUTER GAMES (such as PacMan).

- ☐ None — IF NONE GO TO QUESTION # **[39]**
- ☐ Used computers once or twice
- ☐ Occasional
- ☐ Regular (once or twice a week)
- ☐ Extensive (daily)

[38] Place an **X** beside each of the following computer related activities that you have performed.

- ☐ Used a word processor
- ☐ Used spreadsheet software
- ☐ Mainframe application (ie. data entry/data lookup)
- ☐ Programming
- ☐ Ad-hoc Query or Report Generation

[39] How many educational course have you taken about computers? _____

Please indicate how much the following statements apply to yourself.

[40] I have a good understanding of how computers work.

| Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

[41] I feel apprehensive about using computers.

| Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

[42] I make an effort to read about changes that are taking place in the field of computers.

| Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

Thank you for participating, and enjoy the rest of your stay.