INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

**Can Computer Surveys Reduce Response Effects?
A Cross-Cultural Assessment**

by

Hongmei Liu

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirement for the degree of

Master of Arts

Department of Psychology

Carleton University

Ottawa, Canada

December, 1996

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-22093-1

The undersigned recommend to the Faculty of Graduate Studies
and Research acceptance of the thesis:

"Can Computer surveys reduce response effects?
A cross-cultural assessment"

submitted by Hongmei Liu
in partial fulfillment of the requirements for the degree of Master of Arts

Thesis Supervisor

Chair, Department of Psychology

Carleton University
December, 1996

Acknowledgments

I would like to express my deep appreciation to my advisor Dr. Warren Thorngate for the help he gave me. Without the intellectual instructions, patient corrections, and warm encouragement from Dr. Thorngate, I would not have completed the thesis for my Master's Degree. During the past two years, I have learned not only the methods of doing psychological research from Dr. Thorngate, but also his spirit of hard work and the knowledge of the new techniques and latest developments in the related academic fields.

I would also like to thank Dr. Dick Dillon and Dr. Jo Wood for their intellectual advice on my experiment. Their understandings of the cultural and language barriers I faced was very supportive, and allowed me to complete the course of my programme.

I wish to appreciate the Faculty of Social Science and Psychology Department of Carleton University for granting me the opportunity and funds to complete my Master's programme in Psychology.

I also want to thank many people I met since I came to Canada, especially my best friends Barbara Carroll, Catherine Borshuk, Xiaohui Chen, Jing Liu, and Fatemeh Bagherian for their kind help and warm support.

Finally, I am deeply grateful to my parents and two sisters for their endless support and encouragement during my time in Canada.

Abstract

Response effects are errors in sample surveys that reduce the validity of survey questionnaires. Previous research has shown that the size and direction of response effects are affected by factors such as the respondent's personality traits, cultural and linguistic background, and by the survey administration procedure. With the growing utilization of computer administered surveys and the increasing use of surveys across various ethnic respondents with different cultures, there is growing interest in examining the effects of computers and cultures on response effects. The present study compared how the response effects (i.e., social desirability, extreme response style, and the form of open questions) are affected by computer administration and traditional paper-and-pencil administration of surveys with native Canadian and with Hong Kong Chinese. Results indicated that the representation of respondents' social desirability was not affected by either the mode of survey administration or the ethnic and language background of the sample. Extreme response style was not affected by survey administration mode either, but affected by the respondent's cultural background. The answers to open questions also represented the cultural difference between the two ethnic sample groups and between the different mode of survey administrations. Implications of these results are discussed in terms of the future surveys with computer administration and/or for multi-cultural respondents.

Table of Contents

	Page
Acknowledgments	i
Abstract	ii
Table of Contents	iii
Introduction	1
Historical Overview	2
The Procedures of Sample Surveys	5
Sampling	5
Questionnaire Design	8
Questionnaire Data Collection	15
Measurement Errors and Biases	19
Social Desirability	21
Extreme Response Style	25
Open Questions	28
Response Effects in Cross-Cultural Surveys	31
Summary	35
Methods	36
Subjects	36
Survey Instrument	36
Research Design	40

Results	45
Social Desirability	45
Extreme Response Style	46
Open Questions	48
Discussion	57
Social Desirability	57
Extreme Response Style	59
Open Questions	62
Reference	67
Appendix I	73

Introduction

In the past several decades, sample surveys have become one of the most important research tools of social sciences. They are used in a wide variety of academic research and associated applied fields. By providing researchers with descriptive data, sample surveys can help us test our theoretical hypotheses of social psychological processes, document people's attitudes and opinions, provide information and prediction of political campaigns, guide developmental trends of consumer market. Sample surveys can not only give us knowledge about different social issues, but also affect the public policy decisions.

Sample surveys should allow the collection of information through standardized questions on systematically sampled individuals. Survey procedures consist of three basic techniques: (1) drawing an unbiased sample from the studied population; (2) gathering comparable information for every respondent within a survey by a standardized interview or self-administered questionnaire; (3) determining the relations among variables, describing distributions of the sample, then estimating characteristics of the population by advanced statistical analysis. These techniques help to ensure that sample surveys obtain reliable and valid outcomes. Yet biases often spoil the results. There are many types of survey biases. The present study attempts to investigate the nature and extent of response effects in sample surveys that are related to how the survey is administered (computer versus paper-and-pencil) and to whom (Canadians versus Hong Kong Chinese).

Historical Overview

Surveys, as well as censuses, were born out of the establishment of modern state bureaucracies. Nineteenth-century surveys were similar to censuses, attempting to cover the entire populations of communities (see Rossi, Wright, & Anderson, 1983). With the appearance of the consumer marketing research after World War I, researchers began to investigate consumers' comparative preference by selecting a small number of subjects and bringing them into the psychophysical laboratory for standardized testing. This is the root of modern sample surveys. At that time, however, the techniques of question writing, sampling, interviewing, and data analysis were quite primitive.

In 1936, several organizations conducted polls to predict the Roosevelt-Landon presidential election in the United States. Gallup and Crossley each interviewed 1500 respondents using sex- and age-quota sampling, while the *Literary Digest* mailed questionnaires to millions of telephone subscribers with a small proportion (15%) of the questionnaires returned. Gallup and Crossley successfully predicted the election outcome by a relatively small number of representative respondents sampled from the national-wide population. The *Literary Digest* used a large number of unrepresentative respondents sampled from a special classified population, i.e., telephone subscribers, failed to predict the election outcome (Gallup & Robinson, 1938; Squire, 1988).

At the end of 1930s, academic institutions entered the field of survey research. A group of universities established their own survey organizations. The involvement of

universities in sample surveys accelerated the development of survey techniques for sampling and statistic analysis (see Rossi, Wright, and Anderson, 1983).

During World War II, sample surveys were used by a number of American government agencies to assess a variety of topics including farm productivity, attitudes about price control of consumer goods, attitudes of soldiers before combat and their behaviour in combat, attitudes toward leadership and social control, and even the moral attitudes of pilots (Stouffer, etc., 1949). By the end of the 1950s, sample surveys had become an irreplaceable research tool in academic social science and were broadly used by public and private institutions in the United States.

The development of electronic computers in the 1960s stimulated a great deal of progress in sample survey methodology. By the end of the 1960s, statistical packages such as SPSS (the Statistical Package for the Social Sciences) and SAS (the Statistical Analysis System) were available for survey analysis (Stevens, 1990).

The development of sample survey methodology continued into the 1970s. The application of statistical packages moved from mainframe computers to personal computers, beginning a new era in data processing because personal computers are widely available so that users can conveniently run SPSS/PC+ and SAS PC (Stevens, 1990). Yet sample surveys remained a labour intensive activity. The increased wage rates for interviewers caused the costs for surveys using face-to-face interviews to grow rapidly. In addition, the rising urban crime rates made residents less likely open doors to strangers, including interviewers, and more people were employed full-time so fewer people were at home during the daytime. For these reasons, response rates for face-

to-face interviews declined sharply. Traditional household sample surveys were no longer feasible. After the procedures of mail surveys and telephone interviewing were developed, more and more sample surveys adopted mail and telephone interviewing techniques as well as computer assisted telephone interviewing (CATI) instead of face-to-face interviews (Groves & Kahn, 1979).

Taking fuller advantage of microcomputers, many computer-assisted data collection systems (CADCS) have been applied to sample surveys. The common characteristic of CADCS is that the data collection process in sample surveys is assisted by computer. This has been the main trend of the development of survey research since mid-1980s. CADCS has been regarded as a substitution for the traditional paper-and-pencil questionnaires because of its potential advantages: increased speed and reduction in costs of survey data collection, and improvement in data quality (Saris, 1991). The most widely researcher used types of CADCS include computer-assisted telephone interviewing (CATI), computer-assisted personal interviewing (CAPI), and the computerized self-administered questionnaire, also called computerized questionnaire (CAQ). Since the mid-1980s, the development of computer networks has stimulated the application of CAQ. Questionnaires can be delivered in the network and sent to an individual's personal computer via a modem. In the same way, the collected answers can be automatically returned to the researcher (Kiesler & Sproull, 1986; Rosenfeld, etc., 1989).

The growth of CADCS has inspired the growth of recent studies that focus on comparing the efficiency of different kinds of CADCS to traditional survey methods. The present study follows this research tradition.

The Procedures of Sample Surveys

The purpose of the present study is to investigate the effects of questionnaire administration and respondents' cultural backgrounds on response effects and open questions. First of all, it is necessary to have an overall review of the procedure for conducting sample surveys and the potential biases within each process.

Sample survey procedures are composed of three major processes: sampling, questionnaire design, and data collection. Each process applies its own theories and techniques and may also have potential biases that reduce the accuracy of the results. A good survey design combines proper methods for each process while controlling biases.

Sampling

Before learning what is meant by sampling, it is necessary to know what a population is. A population is the aggregate of elements, and elements are the basic units that comprise the population (Kish, 1967; Frankel, 1983). A population has specific boundaries to determine which elements are in it and which are not. For example, consider the population of "all female Canadians who are 18 years old or above before December 31, 1995". The boundaries of this population are sex (female), residential area (Canada), and age (18 years old or above).

A sample refers to a subset of a population. Thus, a sample is distinguished from a census which gathers information on every individual in a population. Sample surveys deal

with one or more sampling methods for selecting and observing a part of the population. The two basic categories of sampling methods are probability sampling and non-probability sampling. Each of the categories includes a number of specific methods with the corresponding rules to determine how a sample is drawn (Li, 1981).

In probability sampling, every element in the population has a fixed, non-zero probability of being sampled. Under this sampling design, the sample outcomes can be used to make statistical inferences about the population. Simple random sampling, the most basic type of probability sampling, is attained using a table of random numbers applied to a numbered list of all the elements in the population. Other probability sampling methods include systematic sampling, stratified sampling, cluster sampling, and multistage sampling.

Non-probability sampling, on the contrary, includes all sampling methods in which the probability of each element being selected from a population is unknown. Thus, all sampling processes that do not qualify as probability sampling methods are categorized as non-probability sampling. Specific non-probability methods include quota sampling, which is to ensure a given proportion of the elements in the sample has certain characteristics (e.g., set 40% females and 60% Catholics as quotas for sampling); availability sampling, that is, the elements in the sample are easily available to the interviewer; snow-ball sampling, which relies on the respondents referring other respondents to the researcher. Compared to probability samples, one disadvantage of non-probability samples is that their values technically can not be used to make inferences about the corresponding population.

Sampling errors always exist because no sample is likely to produce results that are precisely the same as those for the entire population from which the sample was drawn. It is always possible to draw by chance a particular group of individual elements who happen to be higher or lower than the population data on the measurement. Sampling errors are random biases since such differences between the sample data and the population data are caused totally by random chance.

The degree of sampling error is affected by both the sample size and the variation among the individuals within the population. There is always some chance that a sample includes some atypical members who are different from the most typical individuals in the population. When the sample size is large, the influence of the atypical members within the sample will be small. Thus the larger the proportion of the entire population included in the sample, the lower the sampling error with probability sample.

Variation among the individuals in the population also affects the degree of sampling error. Suppose that all the individuals in the population are perfectly alike in every aspect. Then only one individual is needed for the sample to be interviewed. If no two individuals are exactly the same in the population, a relatively large sample is required in order to have a high representation of the whole population. Therefore, when a sample is drawn from a population with much variation among the individuals, the sample will be more variable than it would be from the population with little variation. Thus the more similar the individuals of a population, the less variation in the population, and the less sampling error of the sample.

Sample errors can also be caused when the sampling process does not follow the sampling probability. As a result, the intended sample designed by a researcher and those actually sampled are not identical. Some practical reasons that may change the probability of sampling include the following. (1) In the personal interviewing surveys, interviewers have a personal choice on houses. They may prefer the first-floor apartments to high-storied apartments, or may avoid houses with dogs. (2) Interviewers may have no incentive to call back respondents who are first not at home. Thus, those people who are often at home, such as housewives and unemployed people, have a much higher opportunity to be selected. (3) Interviewers may make little effort to enlist cooperation of respondents who first refuse to cooperate, thus causing a relatively high non-response rate. These possible errors may modify the probability of sampling and hence make the sample a non-probability sample.

Questionnaire Design

Measurement has been defined as “the assignment of numerals to objects or events according to rules” (Stevens, 1951). For example, respondents’ education can be categorized in terms of graduation degree levels and measured in a number of categories, with numerical values assigned to each category specifically as (1) Secondary school diploma, (2) College diploma, (3) Bachelor degree, (4) Master degree or higher. The task of measuring instruments in sample surveys is to obtain the necessary measurements from respondents. Questionnaires are the most commonly used instruments in surveys. With a

standardized questionnaire, each respondent in a sample is asked the same set of questions. Answers to these questions are therefore comparable.

One criterion of a good questionnaire is high reliability. A reliable questionnaire is one in which two or more respondents with the same characteristics or one respondent over time should answer a question in the same way with item measured does not change.

Validity, another criterion of a good questionnaire, refers to the correspondence between respondents' answers and what the questionnaire is intended to measure. There is usually no criterion to directly estimate the validity of a measurement, that is, whether or not a questionnaire measurement of a concept actually measures the concept. But validity can be reflected and evaluated by two aspects: internal validity and external validity.

Internal validity refers to the internal consistency of items within a questionnaire. If a number of questions are used to measure the same concept, then answers to these questions should be highly correlated. External validity can be estimated by checking the relationships of a measurement against other measurements of the same concept. It is supposed that, if a measurement indeed measures the intended concept, then it should correlate with other measurements which measure the same thing. There are at least three ways of increasing validity. First, make the questions as reliable as possible through careful questionnaire design. An unreliable measurement will consequently reduce the validity of the measurement. In order to increase measurement validity, it is essential to increase reliability by such means as avoiding ambiguity of wording, making the forms and presentations of the questions standardized, using questions that will mean the same thing to every respondent. Second, pretesting of the questionnaire may explore the problems

that the research did not solve in the design of the survey instrument. After the test of a group of potential respondents, the researcher may know whether the questionnaire instructions and the questions are clear. Third, make use of the existing research literature about how to measure what is to be measured. The previous survey instruments on the same research topics can give good suggestions about the new measurement.

The questions in a questionnaire can be broadly classified into two types: closed questions and open questions. Closed questions provide a set of response categories for each question statement. Respondents can select an answer from the response set which is most appropriate to their situation. For example, "How would you evaluate your job? -- very dissatisfied; dissatisfied; neutral; satisfied; very satisfied" and "How likely would you be to vote in the coming election? -- would not vote; might not vote; might vote; would vote". Open questions do not provide categorized responses but instead let respondents give replies in their own words. Examples include the questions "How do you describe your feelings about the current federal government?", "What do you think a successful marriage is?" and "What are your suggestions on the provincial tax reform?"

The numerical values assigned to response categories in measurements have different meanings. The numerical level of responses can be classified into four levels: nominal, ordinal, interval, and ratio. So four types of data can be collected for statistical analysis. (1) Nominal level. Objects or events are classified into unordered categories. Numbers used for these categories are merely symbols without any numeric properties. A simple example is the sex variable whose answers can be listed as 1 male, 2 female.

(2) Ordinal level. Objects or events are placed in ordered categories along a single dimension from lower to higher or in the opposite order. The numerals being assigned to the categories delineate their ranks. Arithmetic operations for these numerals are not meaningful because equal distance among the ranks can not be assured. An example of the ordinal level of measurement is the response to the question: "How would you evaluate your current income -- 1 very satisfied, 2 satisfied, 3 fair, 4 dissatisfied, 5 very dissatisfied".

(3) Interval level. Numbers attached to the categories represent an equal distance between the ordered categories. Interval measures have all the properties of nominal and ordinal level. However, since they do not have an absolute zero, multiplication operation with the numbers are consequently meaningless. In fact, interval measures are quite rare.

Fahrenheit temperature scale is one of the few common examples. The intervals between two Fahrenheit degrees, 50 and 90 degrees for example, are comparable. But it is meaningless to conclude that the temperature of 80 degrees is twice as hot as 40 degrees.

(4) Ratio level. Numbers assigned to objects have an absolute zero. Therefore, ratio measures are suitable for all arithmetic operations. They have all the properties of the other measurements and are the highest level of measurement. Typical questions providing ratio measures include questions about age, weight, family size, and income.

The above four levels of measurements have been represented by different scales. A scale is a representation of the response categories with which respondents can arrange their attitudes. Scales can be coded by numbers. Therefore scales can turn the response categories into the corresponding levels of numeric codes for statistical analysis. With the

application of scales, the distribution of respondents can be described, and the positions of sampled individuals or groups can be compared. Four commonly used scaling techniques are listed below.

1, The Likert scale, also called the summated rating scale, is one of the most widely used scaling methods in sample surveys. A Likert scale is an ordinal level measurement which consists of a set of statement items. Each statement item is followed with an agreement scale represented by a set of numbered categories. These numbers reflect the degree of agreement or disagreement of a respondent's attitude on each statement. The agreement scales may have two categories (agree -- disagree). Five to seven categories are commonly adopted. Each category is scored with a numeric weight. A five category scale, for example, might have the following list: (1) Strongly disagree; (2) Disagree; (3) Neutral; (4) Agree; (5) Strongly agree. The scoring can also be represented in reverse order in which 1 indicates strongly agree and 5 indicates strongly disagree. Sometimes the scores are summed after every statement has been answered. Then the sum of the scale scores is assumed to be the respondent's attitude score.

The Likert scale is popular because of its flexibility, simple format, and easy composition and interpretation. The statement length of each item can be only a few words or several lines long. One set of instructions and scale representation can serve a large number of items. Once the respondents understand the instructions, they can complete the item set quickly and easily. More important, a summated score can be obtained. Instead of using the results of each individual item, the score of the set of

statement items can be summed as an overall value. However, though the distance between “strongly agree” and “agree” is numerically equal to the distance between “agree” and “neutral”, it may not be psychologically equal. Thus, as an ordinal level measurement, the Likert scale is not applicable for arithmetic operation.

2, Guttman’s cumulative scale, with two choices (e.g., approval -- disapproval) for each item, is composed of a set of logically related items. Each item measures a progressively higher degree of intensity to an attitude object. A positive response to any one item implies a positive response to all items of lower intensity. An example of the Guttman scale is Bogardus’ social distance scale (Bogardus, 1925), designed to measure the attitudes toward ethnic groups. Respondents are required to check whether or not they will accept an average member of an ethnic group as: (1) A relative by marriage; (2) A close friend; (3) A neighbour; (4) A colleague; (5) A citizen of the country. The result will indicate how close they will accept the member of an ethnic group. It is assumed that a respondent who answers “yes” to (1) is also likely to answer “yes” to all the other items. The scores to these items will be summed. The more the total score, the higher degree of the respondent’s acceptance of that ethnic group. The cumulative scale is easy to interpret. However, it is often difficult to find the set of items which have a perfectly logical relationship.

3, Osgood’s semantic differential scale presents a concept by a series of scales (Alreck & Settle, 1985). Each scale has two opposite adjectives with points between them to describe the concept, usually with 7 points. Examples of the two opposite adjectives include “good” and “bad”, “hot” and “cold”, “moist” and “dry”. Respondents are

instructed to rate the concept by selecting a point on each interval according to their immediate impressions or feelings toward the meaning of the concept. It is assumed that what is measured is the connotative meaning of the concept, that is, the implied meaning or the connotations to the respondent. For example, the concept “Green Peace” may be rated on the scale as follows:

	Green Peace							
Good	1	2	3	4	5	6	7	Bad
Weak	1	2	3	4	5	6	7	Strong
Active	1	2	3	4	5	6	7	Passive

Respondents in this example are required to describe their impression on “Green Peace” organization by rating on the three 7-point bipolar adjectival scales. If “Green Peace” is considered “strongly active”, the third scale should be marked “1”; if considered “slightly active”, then it should be marked “3”. “Four” indicates the “neutral” category. The choice of rating depends on how much a respondent feels for the activity of “Green Peace”.

The advantage of the semantic differential scale is the easy operation, in which a number of concepts can be rated on a set of scales easily and quickly. But it requires that the concept can be rated on the scales, and each scale is labelled by two opposite adjectival to describe the concept.

The validity of a questionnaire may be damaged by a number of questionnaire design problems. Vague or complex wording may make the questions ambiguous (Fowler, 1993). The order of questions may affect respondents’ answers on the questions, which is called

order effect (Bradburn, 1983). Moreover, the major systematic problems that may invalidate questionnaire results are response effects. Response effects are biases that occur when respondents give answers not directly based on the question content, but instead related to the behavioural characteristics of themselves or the situation or interaction of the two. Typical response effects include social desirability, extreme response style, acquiescence (Yea-saying), and carelessness. The purpose of the present study is to investigate some of these potential response effects.

Questionnaire Data Collection

Questionnaire data collection can be represented by three procedures: self administration, group administration, and interviewer administration. Self administration requires a respondent to read and answer questions in the absence of an interviewer, so respondents require necessary reading and writing skills. Self-administered surveys can be conducted either by mail or in the household which each has some advantages and shortcomings.

Mail surveys cut the costs of hiring interviewers and travel by mailing questionnaires to respondents. A self-addressed stamped envelope usually accompanies the questionnaire for return mail. Respondents are given instructions for how to complete the questionnaire. To be a self-administered survey, questions should be designed in simple forms with short length and clear sequence so as to maintain the interest of respondents. However, the generally low response rate of mail surveys has stymied researchers for years (Dillman etc., 1974; Fox etc., 1988). The researcher sometimes should prepare for a second or even a third mailing to remind the respondents to answer and return the questionnaire

(Heberlein & Baumgartner, 1978). Additionally, potentially incorrect mailing addresses may cause sampling bias. Even so, the relatively low cost of mail surveys still attract researchers to conduct surveys, especially large-scale or national-wide surveys, by mail when they are applicable.

The household technique for self administration surveys can be simply described as a drop-off and pick-up procedure. The task of the interviewer is to deliver the questionnaire to the sampled individual's residence and set a date for returning to pick it up, or alternatively to ask the respondent to mail it back after completion. In this way, interviewers may have a chance to answer inquiries of respondents. The response rates for household surveys tend to be as high as the personal interview particularly when interviewers persist in picking up the questionnaires, but the cost increases at the same time. The household method is more applicable when the respondents are concentrated in a limited number of resident areas.

In a group administration survey, a group of respondents answer questionnaires together. Thus, group administration is possible when the sampled individuals are willing to come to a given location at the same time. Usually the sampled individuals have already existed in a kind of organization such as school students, factory workers, team members of a sports club. The questionnaire administrator's task focuses on organizing respondents and clarifying questions. Respondents should complete the questionnaire independently. Thus, group administration is similar to self administration. This mode is economical and is a quick way to complete the process of data collection.

In interviewer administration, interviewers are trained and supervised to read the questions and note respondents' answers so that there is direct feedback between interviewers and respondents. The questionnaires can be longer and more complex than those in self administration because interviewers can help to control the sequences of questionnaires in addition to the instructions. Two commonly used techniques of interviewer administration are the personal interview and the telephone interview.

The personal interview, also called a face-to-face interview, is a traditional way of data collection. It tends to be the most costly and time consuming type of administration because interviewers go to each respondent and communicate with him/her in person. Usually large-scale probability sampling surveys do not employ personal interviews when sampled individuals are spread across a wide geographical area. But the response rate and data quality of personal interview is higher than any other method, in part because the number of refusals and omitted questions decrease (Jordan, Marcus, & Reeder, 1980; Krysan, Schuman, Scott, & Beatty, 1994).

An alternative technique for interviewer administration is the telephone interview. The telephone interview has been widely used since the creation of random-digit-dial sampling by which the area codes and telephone numbers are randomly selected by computer or another randomizing process. Telephone interviews usually cost less money and time than personal interviews. Questionnaires in telephone interviews can be longer than those in mail surveys, but shorter questionnaires are recommended. Otherwise respondents may become fatigued or distracted when they are required to stay on the phone for too long. The choice of scales in a telephone interview should also be simplified (e.g., yes-no, agree-

disagree) instead of lengthy answer categories due to the difficulty of all categories being held in the respondent's short-term memory. Telephone interview surveys are speedier than sample surveys conducted either by mail or by personal interview. As a result, they are frequently used for surveys such as public opinion polls and election polls.

With the development of microcomputer technology, computers have been widely employed in the process of collecting data in sample surveys. There are three main computer administered data collection systems (CADCS):

- (1) Computer-assisted telephone interviewing (CATI). In this system, an interviewer sits in front of a computer and talks with the respondent on the telephone. Through the phone, the interviewer reads the questions shown on the monitor screen to the respondent and types the respondent's answers into the computer. It has the advantages of telephone interview while combining the interviewing and inputting data process into one step.
- (2) Computer-assisted personal interviewing (CAPI). The interviewer takes a portable computer to the respondent's residence. The interviewing process is the same way as in CATI except that the interviewer conducts the interview face-to-face instead of over telephone. As with CATI, CAPI is also a type of interviewer administration in which the survey is administered by the interviewer.
- (3) The computerized self-administered questionnaire, is often called the computer-assisted questionnaire (CAQ). In this system, questionnaires are presented on the computer screen. Respondents are required to read questions and input answers themselves on the computer. Thus the traditional paper-and-pencil questionnaire is

replaced the computer questionnaire.

Additional CADCS systems include touch-tone data entry (TDE) and voice recognition entry (VRE). In both systems, respondents listen to recorded questions read by the computer and give answers by telephone. Such techniques are only useful for relatively short and simple surveys.

Compared with the traditional paper-and-pencil questionnaires, CADCS systems are more efficient, speedy, and less costly. They also improve the data quality by simplifying the data inputting process (Saris, 1991).

Measurement Errors and Biases

Like all the other scientific measurements, those from sample surveys are not error free. According to Bradburn's (1983) categorization, there are two main types of error in surveys: sampling error and nonsampling error. Sampling error refers to the random, chance variance of sample estimates around the true population value. Sampling errors are random biases of sampling, and are assumed to "cancel each other out" in the long run.

Nonsampling error comes from the procedures of survey measurement. Major nonsampling errors include nonresponse, missing data, and respondents' misinterpretations of a question. Nonresponse refers to the failure to obtain a respondent's reply because he/she is not at home or refuses to participate the survey, or the answered questionnaire is lost or not returned by the respondent. Missing data are caused by skipped questions, illegible responses, refusals to answer some items, etc., which make a respondent's reply incomplete. Misinterpretations of a question occur for four major reasons: (1) a respondent's misunderstanding of the question, perhaps because of the interviewer's

unclear presentation of the question; (2) a respondent's errors of memory about what or when something happened; (3) a respondent's intention to enhance his/her self-presentation by deliberately falsifying a response in order to make a good impression on the interviewer or to prevent the interviewer knowing his/her real situation; (4) a respondent's habitual tendency to give some kinds of answers such as saying "yes" or using extreme categories in measurement scales.

The errors from misinterpretations are called response biases or response sets or, more generally, response effects. Respondents produce response effects when they give answers not directly based on the question content, but instead related to the behavioral characteristics of themselves and situation. When the question is about family income, for example, respondents might understand the definition of family income in different ways than the researcher such as that whether it includes insurance payments or inheritances. Respondents also might forget about some income (e.g., stock interest), or they might deliberately keep secret income from illegal sources or from cash payment that they have evaded tax. They even might inflate their income because of vanity. These different kinds of potential errors are response effects. Among them, the intention of falsifying responses is the most difficult to control, and may seriously affect the validity of survey measurements. This intentional error is called social desirability.

In addition to the motives of respondents, other factors may also cause response effects. These factors include questionnaire administration method, question format (open versus closed), and the respondent's personality or cultural background. Respondents, for

example, may change their answers from “Yes” to “Maybe” if the survey question allows them the “Maybe” alternative. Young respondents may use the extreme ends of rating scales more often than old respondents do. The present study will investigate two important response effects, social desirability and extreme response style, in terms of the comparisons between two different survey administrations and two sample groups which have different cultural and language backgrounds.

Social Desirability

Among a variety of response effects, social desirability (SD) is one of the most common. SD can be simply defined as the tendency of giving the most socially acceptable answers to a question (Oskamp, 1977). In Crowne and Marlowe’s (1964) description, SD occurs when “people conform to social stereotypes of what is good to acknowledge concerning oneself in order to achieve approval from others.” (p27). Phillips and Clancy (1971) more broadly interpret SD as the tendency to deny socially undesirable traits and admit socially desirable traits. In sample survey research, respondents might give the answer to an item not according to their real attitudes or behaviour, but according to what they believe is the socially acceptable attitude or behavior.

Although SD is hard to control, it can be measured and revealed. Edward (1957), for example, created a SD scale from various MMPI scales to assess the degree of respondents’ tendency to give socially desirable responses in self-description. He found that those personality characteristics considered as desirable in society are also claimed by most respondents, while undesirable characteristics are denied by them.

One of the most widely used instruments for measuring SD is the Marlowe-Crowne Social Desirability scale (M-C SD) (Crowne & Marlowe, 1960, 1964). Crowne and Marlowe (1964) summarized SD as the need for social approval and hypothesized that individuals differed in the strength of having this need. Moreover, they assumed that those people with higher need would be more likely to choose socially approved answers and avoid the opposite responses in a questionnaire. Because of its good reliability and validity, M-C SD scale became a popular instrument for assessing the amount of SD in surveys. By using this scale as an instrument, many interesting correlates of SD have been found (Phillips & Seal, 1969; Cooperstock, 1971; Clancy & Gove, 1974; Gove & Geerken, 1977). Phillips and Clancy (1972), for example, supported Crown and Marlowe's assumption that need for social approval did affect people's responses. They further found that respondents who assessed a personality trait as highly socially desirable (e.g., having many friends, being nonprejudiced) reported themselves more frequently on that trait than did people who assessed that trait as having relatively low social desirability.

Clancy and Gove (1974) examined sex differences in the need for social approval using a short version of M-C SD scale. They found that males and females showed little differences in the SD response. This finding supported the conclusion that the higher rate of reporting emotional problems among women reflected the actual gender difference instead of women's lower tendency of SD. This contradicts Phillips and Segal's (1969) and Cooperstock's (1971) argument that the higher rates of women's reports of psychiatric symptoms could be attributed to women's perceptions of such symptoms as

less undesirable than men's perceptions.

The presence of SD in survey research has been revealed in some studies. As Gove and Geerken (1977) point out, however, the accumulated evidence from SD research is still not enough for a final conclusion about the effect of respondents' traits on SD responding in sample surveys.

From the 1980s, the application of computers in sample surveys shifted from data analysis to data collection. The involvement of computer applications in survey research has both the advantages and limitations (Sarvis, 1991; Synodinos & Brennan, 1988). Several studies have been done to examine the validity of the computer administered questionnaires by comparing them with the traditional paper-and-pencil administered questionnaires. One of the famous studies for response effects in computerized surveys was done by Kiesler and Sproull (1986). They compared the amount of SD in a computer administered sample and a paper-and-pencil administered sample using a 5-item short version of the M-C SD scale. Respondents in the computer condition gave fewer socially desirable responses than those in paper-and-pencil condition. This result is supported by a similar study conducted by Martin and Nagao (1989) who found that SD scores were lower in a computer administered survey than in a face-to-face survey. Their research suggests that the highly standardized, anonymous, impersonal and nonjudgemental nature of computers might reduce socially desirable responding with self-report assessment, especially when the interview contained potentially socially sensitive items.

Counter evidence was also found by some researchers. Lautenschlager and Flaherty (1990) compared SD among computerized, individual paper-and-pencil, group paper-and-

pencil surveys when respondents were required to report their identities to the researcher and when their anonymity was assured. They used Paulhus's (1986) Balanced Inventory of Desirable Responding (BIDR) to assess SD. There was more SD responding in the computer condition than in both the individual and group paper-and-pencil conditions. The test also indicated that respondents in the anonymous condition presented less socially desirable responding than did those in identified condition.

There is also a third kind of result concerning the effects of computer administration on SD. A number of studies have claimed that SD shows no systematic difference between computer-assisted and paper-and-pencil surveys, even though respondents might prefer to answer questionnaires on a computer (Whitener & Klein, 1995; Davis & Cowles, 1989; Finegan & Allen, 1994; Millstein, 1987; Booth-Kewley, Edwards, & Rosenfeld, 1992).

From the above studies, it can not be simply concluded that computer surveys increase, decrease, or have no effect on SD responding. The contradictory results can be attributed to many factors such as respondents' individual characteristics, different interface design of computer questionnaires, various types and lengths of scale versions, and so on. More importantly, there might be other factors which affect SD. Examples include whether the respondent completes the questionnaire individually or in group (Lautenschlager & Flaherty, 1990; Whitener & Klein, 1995), whether the computerized questionnaire can be freely navigated to previous and following pages (Whitener & Klein, 1995), and respondents' varied experience with and anxiety about using a computer

(Finegan & Allen, 1994; Davis & Cowles, 1989). These factors may influence the validity of the measurement and should be explored in future research._

Extreme Response Style

Extreme response style (ERS), also called extremity response bias, is generally defined as the tendency of respondents to prefer the end-points of rating scales (e.g., 1 and 7 on a seven-point scale). Other interpretations of ERS include: "The proportion of rating scales a respondent answers in extreme intervals" (Greenleaf, 1992, p330), and the standard deviation of the respondent's answer from the middle-point of the rating scale (Warr & Coffman, 1970; Kiesler & Sproull, 1986). Respondents exhibiting ERS give extreme answers to rating-scale questions not because of question contents, but because of their response biases. ERS can be problematic in rating-scale sample surveys because ERS will skew the frequency distribution of the choice categories toward the tails, then increase the standard deviations for the variables, and subsequently increase within-group variance. The inflation of such error will reduce the chance of rejecting the null hypothesis, that is, increase type II error (i.e., accepting of the null hypothesis when it is false) (Hamilton, 1968; Hui & Triandis, 1985; Heide & Gronhaug, 1992).

Previous studies focusing on the relationships between ERS and different kinds of personal characteristics have obtained various results. Some researchers have discovered that ERS is independent of specific item contents and is stable over different measurements and over time (Arthur, 1966; Merrence, 1970; Bachman & O'Malley, 1984; Warr & Coffman, 1970; Greenleaf, 1992; Zax, Gardiner, & Lowy, 1964). Marin et al (1992) and Shulman (1973) demonstrated that less educated respondents use the

extreme intervals more frequently than more educated ones. Bachman and O'Malley (1984) also reported a racial difference between black and white respondents on ERS: compared to white respondents, blacks were more likely to use extreme categories in 5-point Likert scales, especially the positive end, i.e., the "mostly agree" extreme. Crandall (1982) assessed ERS by a 6-point Likert scale and investigated the relationship of ERS and social interest. Social interest is defined to be "genuine interest and concern for the welfare of others as well as oneself" (p82). The result showed an inverse relation between social interest and ERS scores. Biggs and Das' (1973) study indicated that ERS differences reflected respondents' different cognitive styles. Specifically, respondents who preferred selecting extreme intervals had a cognitive style of internality (extroverted, dogmatic, and analytical); respondents who preferred central intervals had a style of externality (being introverted, non-dogmatic, and leading to fact-rote learning).

The above conclusions are challenged by other research. Hui and Triandis (1985) indicated that response effects, including ERS, were more similar across adjacent time periods than across lengthy time periods, which implies instability of ERS over time and across individuals. Other research (van der Kloot, Kroonenberg, & Bakker, 1985) indicated that extreme responses varied by specific stimuli, scales and their interaction. This is consistent with Albaum and Murphy's (1988) report that the differences of ERS are due to scale methods. In their research, respondents were assigned either to answer questions on a 5-point Likert scale directly, or first to indicate agree, disagree, or neither of them and then give the intensity of their responses. The so-called two-stage

measurement showed a greater proportion of extreme responses than the one-stage process. In addition, Norman (1969) found that respondents with high anxiety would have a greater tendency to use extreme categories especially when the stimuli were ambiguous, while Innes (1977) demonstrated a positive relationship between anxiety and 'yea-saying' extreme response.

Another focus of ERS research is on the effect of survey methods. Jordan, Marcus & Reeder (1980) discovered that the telephone survey sample exhibited higher ERS scores than did the household sample. This conflicted with the results of another study (Herzog & Rodgers, 1988) in which ERS differed little between telephone and personal survey samples.

Kiesler and Sproull (1986) examined ERS in computerized survey and paper-and-pencil survey. The result showed no significant difference in ERS attributable to administration method. Unlike the research on social desirability, studies of response style in computer administered surveys are rare. This is somewhat surprising given the increasing use of computer applications in sample surveys.

In summary, most previous research on ERS has focused on testing the validity and stability of ERS as one type of response effect and its relationships with personal characteristics. The conflicting results suggest that some potential factors such as the item contents of scales and the number of categories might affect the prevalence of ERS.

Unlike tests of personality traits and intelligence and even the test of social desirability, there are no standard instruments for assessing ERS as a response bias (Paulhus, 1991). The applied measures for ERS might not be particularly designed for assessing ERS but be

instead developed for other purposes. And most researchers prefer to use their own measurements. Although most items are presented as Likert scales, the different numbers of intervals also affect the results. Hui and Triandis (1985) reported that when other factors are controlled, respondents used extreme intervals in the 10-point Likert scale less frequently than in the 5-point scale.

As Greenleaf (1992) has pointed out, the stability of ERS over time with different measurements can support the conclusion that ERS is a stable and valid response style related to people's personal traits. Respondents with varied cultural backgrounds have different general personalities. It is possible that cultural characteristics also affect the response style, that is, people from different cultural backgrounds have different number of tendencies regarding ERS .

Open Questions

As well as response effects cited above, the form of survey can also affect responses. As noted earlier, question forms are broadly classified into two types: closed and open. Closed questions provide two or more categories for respondents to select. Examples include rating scales and yes-no questions for which the question formats are fixed. In contrast, open questions (also called open-ended questions) have no fixed answer choices but instead ask a reply in the respondents' own words.

The advantages and disadvantages of these two types of questions have been debated by several researchers (e.g., Lazarsfeld, 1944; Sheatsley, 1983; Shuman & Presser, 1981). Free-format questions allow respondents to frame their answers without forced

suggestions from the researchers. The researchers can get fuller and probably more detailed replies that more exactly reflect respondents' feelings or opinions. Fixed-formatted closed questions, in contrast, are often easier to ask, answer, code, and analyse. Thus, closed questions are much more frequently and widely used in large-scale sample surveys. It has been suggested (Sheatsley, 1983) that an open question can be employed when there are too many categories for a closed question (e.g., "Who is your favourite actor?") or in collecting answers that become the basis for closed categories.

Dohrenwend (1965) examined the effect of question forms (closed vs. open) on response effectiveness. She designed a number of criteria to evaluate the effectiveness of answers and compared the answers given to open and closed questions in an interview. Included were criteria such as the length of the answers, the usability of a response when the respondent's answers were not meaningless, ambiguous or irrelevant to the questions, and the amount of self-revelations which were identified by respondents' own opinions, attitudes, feeling and beliefs. Respondents were asked both the open and closed questions by the interviewer. Dohrenwend's findings argued against the presumption that open questions would be less effective than closed questions. In addition, there was no evidence to indicate that open questions could offer replies either with more depth or with more validity. The study also indicated that respondents with higher education answered more open questions than did respondents with lower education. Thus it was concluded that open questions were less effective with respondents at a lower educational level.

The effect of the two question forms (closed vs. open) on the answers to salient questions has been examined by Shuman, Ludwig, and Krosnick (1986). They defined

salient questions as those questions that catch a person's immediate attention. According to their hypothesis, the answer "the threat of nuclear war" is seldom given to a question asking for "the most important problem (MIP)" in open formed questions because the answer lacks salience, while it is more frequently chosen in a closed set of MIP categories because of the real importance of the answer. Their comparison between open versus closed MIP questions did not show a significant effect of question forms on the answers to questions about MIP and salient issues. This result suggests that open questions are preferable when the possible choices for answering the questions are too many.

Because of the difficulties in researching open questions, studies about the effect of survey administrations on open questions are rare. Kiesler and Sproull (1986) compared the total number of words, the number of personality traits of respondents' self description, and the number of personal pronouns in answering to open questions between computer administered and paper-and-pencil administered surveys and found no significant difference when the answers for open questions in the computer survey could not be edited freely in word processor fashion. In a subsequent test, when respondents were permitted to edit the answers for open questions freely in the computer, the freely-edited answers were significantly longer than those in the paper-and-pencil questionnaire and the non-edited computerized questionnaire. In the freely-edited condition, the number of the first-person personal pronouns in open question answers was more than in the non-edited condition. This finding implies more self-centered responses of respondents in the free-edited computerized survey.

Although longer answers to open questions may not be better, it is quite possible that shorter answers may not provide enough necessary information. A user-friendly computer survey with text editing capabilities encourages respondents to give longer answers to open questions. Kiesler and Sproull's (1986) computer survey was done by electronic mail which lacked a text editing function. The present study will conduct a computer survey with newer software in which open questions can be typed and edited freely. This study will test the effect of a computer administered survey on answers to open questions and compare it to the traditional paper-and-pencil method.

Response Effects in Cross-Cultural Surveys

One of the difficult tasks of sample survey design is to construct questionnaires with comparable survey questions across cultures. In order to ensure sample comparability, researchers should develop equivalent questions that convey the same meaning to respondents from different cultural and/or linguistic backgrounds. Furthermore, researchers should understand the extent of different response effects for respondents of different cultures.

Ross and Mirowsky (1984) tested the tendency to give socially desirable responses among Mexican descendants and Anglos; Mexican Americans gave more socially desirable responses than did Anglos. This supported their hypothesis that people in relatively powerless position would prefer to represent a "good face" (p194) to outsiders. In another study (Dyal & Chan, 1985), samples of Hong Kong Chinese students both in Hong Kong and in Canada exhibited more tendency on social desirability than did a Canadian sample.

Studies on extreme response style have also found evidence to suggest that people with different cultural background have different ERS in rating scale surveys. Two studies of Hispanic and non-Hispanic whites have obtained consistent results: Hispanic respondents in US use extreme intervals more frequently than do non-Hispanics (Hui & Triandis, 1989; Main, Gamba, & Marin, 1992). In an earlier study, Zax and Takahashi (1967) compared the tendency to select extreme ratings between American and Japanese respondents. The data revealed a much stronger ERS in the American sample. Chun, Campbell and Yoo (1974) also found a greater tendency for ERS among American respondents compared to Koreans. Another cross-cultural study (Stening & Everett, 1984) including samples from nine countries found remarkable ERS differences among various samples. Among them, Hong Kong respondents under 40 years old with a degree or diploma used mid-point interval with the highest frequency, while samples from Malaysia and Indonesia over 40 years old without a diploma gave the most extreme responses. These differences are attributed to particular effects of cultural factors such as the language used to pose the questions (Stening & Everett, 1989), linguistic styles (Bachman & O'Malley, 1984), judgment styles (Hui & Triandis, 1989), child-rearing methods (Zax & Takahashi, 1967).

Moral and social values might also affect people's response styles in sample surveys. Consider Chinese values, for example, with which the present study is concerned. The fundamental aspects of Chinese value include tolerance of others, harmony with others, humbleness, solidarity with others, non-competitiveness, personal steadiness and stability,

prudence (carefulness), being conservative, protecting “face”, and moderation, that is, following the middle way (Chinese Culture Connection, 1987). These aspects influence people’s attitudes, judgments, and interpersonal relationships.

With the traits of protecting “face” and prudence, it is reasonable that people may have more need for approval and may avoid admitting socially undesirable attitudes. As a result, it can be expected that Chinese people might have a higher score on a SD test. Similarly, the Chinese values such as moderation, being conservative, and humbleness make Chinese avoid extreme attitudes and radical behaviour. It might be reflected by low ERS among Chinese respondents.

Contrary to the Western culture which emphasizes individualism, Chinese culture emphasizes social cohesion and subordination of the individual and teaches people to be in harmony and solidarity with others. Consequently people with Chinese values may not be like those in Western culture who are comfortable expressing individual feelings and opinions. It is thus reasonable to expect that, because of different cultural values, Chinese people will give fewer and shorter replies to open questions than will Western people in sample surveys.

In the present study, a group of Chinese respondents was sampled from among those raised in Hong Kong. One reason is that more and more immigrants in Canada are from Asia, especially from Hong Kong. Among the 220,000 immigrants coming to Canada in 1994, for example, 16% of them were from Hong Kong, and 52% from other Asian areas. Since 1991, the Chinese language has become the second largest minority language in Canada (Serrill, 1995). Another reason to use the Hong Kong sample is that

communication and commercial trade between North America and the Pacific area are becoming more frequent. It is desirable to have knowledge of Hong Kong people's response styles in surveys sampled from that area.

Hong Kong, like Taiwan, Singapore and South Korea, is a typical modern industrialized region that successfully developed the Western pattern of capitalism while keeping its Chinese cultural heritage (Wong, 1986). According to an earlier investigation, approximately 97% of Hong Kong residents are Chinese-speaking native Chinese (Jarvie & Agassi, 1969). They have a close tie with Chinese traditional culture. For example, the term "face" has an important position in Hong Kong social values. People seriously fear "losing face" (being humiliated) and have a strong desire to "keep face" (Agassi & Jarvie, 1969, p38). Meanwhile, having been a colony of Britain for a century, Hong Kong's economical, legal, and bureaucratic systems are copied from the West. Researchers have compared managerial values (Ralston et al., 1992; Wong & Olve, 1995) and moral judgment (Ma, 1989) among mainland China, Hong Kong, and the United States or England respectively. The results have illustrated the mixed attitude and behavioural styles of Hong Kong Chinese who combined the Eastern and Western social characteristic patterns. This finding confirms that it is worthwhile to explore the response biases of Hong Kong respondents in survey research.

Summary

The present study is concerned with the extent of three kinds of response effects in computer and paper-and-pencil surveys among respondents from different cultural and language background. The hypotheses are listed below.

1. It is expected that the social desirability scores are lower in computer administration than in paper-and-pencil administration; and the SD scores among a group of Hong Kong Chinese respondents are higher than among a group of Canadians.
2. It is predicted that the ERS scores have no significant difference in computer administration and paper-and-pencil administration, but that ERS scores of the Hong Kong sample are lower than those of the Canadian sample.
3. The study also predicts that respondents will give longer statements and more suggestions for open questions on a computer questionnaire than on a paper-and-pencil questionnaire, but that the Hong Kong respondents will give shorter statements and fewer suggestions than do the Canadian respondents.

Methods

Subjects

A total of 64 subjects, 32 native Canadians and 32 Hong Kong Chinese, participated in this study. Twenty of them were recruited from the undergraduate students enrolled in Introductory Psychology at Carleton University. The other subjects were first or second year students including five third year students from other courses at Carleton University. As requested in the participation sign-up sheet, the Canadian subjects grew up in Canada and had lived in Canada long enough to obtain their education in Canada since junior high school and with English as their first language. The Hong Kong subjects had been in Canada ranging from 9 months to 5.5 years (average = 3.1 years) and were able to read and write both English and Chinese. Among the 32 Canadians, 16 were males and 16 were females. The 32 Hong Kong subjects were composed by 17 males and 15 females. The average age of the total sample was 22.7, ranging from 19 to 35 years old.

Survey Instrument

There is no generalized regulation for the length of self-administered questionnaires. It is suggested, however, that questionnaire length will affect the response rate and quality of answers. The present questionnaire was intended to take 25 to 35 minutes because half of the subjects' first language is not English, and they may require a little more time than the native Canadian subjects to complete the questionnaire.

The response effect of social desirability was measured by a shortened 24-item version of the 33-item Marlowe-Crowne Social Desirability scale (Crowne & Marlowe, 1964).

According to a pretest by a respondent whose first language is not English, it took 20 minutes to complete the original M-C SD scale. Considering both the length and contents of the scale, 9 of the original 33 items were removed. Items that were not suitable to the subjects in the present study or that contained unusual words were excluded. For instance, the item "Before voting I thoroughly investigate the qualifications of all the candidates" and the item "I never make a long trip without checking the safety of my car" were excluded because some Hong Kong subjects did not have the experience of voting, and some subjects did not have a car. The remaining 24 items took less than 15 minutes to complete in a second pretest.

The contents of the items are the description of ordinary personal and interpersonal behaviors with two scale choices for each item: "true" and "false". Among the 24 items, 11 of them are socially desirable but people usually do not do them (e.g., "I never hesitate to go out of my way to help someone in trouble."). Respondents would represent their endorsement to socially desirable items by selecting "true". The other 13 items are socially undesirable but people usually do them (e.g., "I like to gossip at times."). Respondents demonstrate their rejection of socially undesirable items by responding "false". The two types of items were listed in random order in the questionnaire (see question 8 to question 31 in Appendix I).

Extreme Response Style (ERS) was assessed by the ERS scale created by Greenleaf (1992) because it is a typical Likert scale and the only validated instrument available for assessing ERS. Fourteen items were employed in the present study from the original 16 items. The item "I am a homebody" was removed in order to avoid possible

misunderstanding by the Hong Kong subjects. The item “No matter how fast our income goes up, we never seem to get ahead” was deleted as well because of the altered meaning when it was translated from English into Chinese. The original 6-point scale was replaced by a 7-point scale, adding the “neutral” category. And the interval range was changed from 1 to 6 to from -3 to 3 (see question 32 to question 45 in Appendix I) in which “neutral” was assigned to be 0. The two end-point labels (interval -3 and 3) were “strongly disagree” and “strongly agree”. In the first pretest, it took 10 minutes for the subject whose first language was not English to complete the original scale. The shortened scale took less than 10 minute for all subjects.

The open questions in computer administration and paper-and-pencil administration were different because subjects might simply repeat their answers in the second session if the questions were exactly the same. Considering the total length of the questionnaire, the present study left 8 to 10 minutes for answering open-ended questions, 4 to 5 minutes per question. So, there were two versions of the open-ended questions. Each version had two questions. The two questions in version A were number 46-A “In as many sentences as you wish, please describe the characteristics of your personality.” and numbered 47-A “Below please list what you believe are the major problems facing the world today.” The two questions in version B were numbered 46-B “In as many sentences as you wish, would you please recommend ways to balance the budget at your university.” and numbered 47-B “Below please list the qualities you want to find in your (future) spouse.” (see Appendix I). If a subject answered version A in computer questionnaire, for example,

then he/she would answer version B in paper-and-pencil questionnaire, and vice versa. These four questions had been selected because their topics were equally related to all subjects in the study. Thus, if a subject had refused to respond, it would be more likely due to his/her attitude than to lack of knowledge of the answers.

Additionally, the questionnaire contained five demographic items concerning subjects' major, year at university, gender, age, nationality, and number of years in Canada (see question 1 to question 5 in Appendix I). There were also two items about the respondent's knowledge of computers and a self evaluation of the Hong Kong subjects' English level (see question 6 to question 7 in Appendix I).

In summary, the questionnaire was composed of four sections: personal background (Q1--Q7), the SD scale (Q8--Q31), the ERS scale (Q32--Q45), and the open questions (Q46A--Q47A or Q46B--Q47B). The order of the sections might affect the answers, particularly the length of the open questions. Possible methods for dealing with the order include complex Greco-Latin square, the random order, and the fixed order for each subject. In most previous research, however, open questions are traditionally put in the last part of the questionnaire. Since the present study focused on administration mode and cultural background rather than possible order effects, the present questionnaire listed the four sections in one fixed order (see Appendix I) for all subjects. However, the present experimental design for the treatment of questionnaire administration mode was a repeated measure, that is, the items for the computerized and paper-and-pencil questionnaires were identical. In order to control for the possibility that respondents give answers to the second measure according to their memory from the first measure, the item order within

each section were re-randomized in the computerized and paper-and-pencil questionnaires for the second measure.

Research Design

The present experiment was composed of three sub-experiments which had identical independent variables. The two independent variables were: (1) mode of questionnaire presentations, and (2) ethnic origin/language.

The mode of questionnaire presentation had two levels: the paper-and-pencil questionnaire (PPA) and the computer questionnaire (CA). In both mode, the questionnaires were self-administered. In the PPA condition, the questionnaire was printed by Laser printer in 6 pages. Respondents used a pencil to circle their chosen responses and to write words on the paper. In the CA condition, the questionnaire was presented on a monitor screen by a computer program written in Multimedia ToolBook. Each of the 12 screens contained 2 to 4 questionnaire items. Respondents used the mouse to click on the corresponding buttons to select their chosen responses and to turn to different pages; they inputted answers to open-ended questions by using the computer keyboard. In each administration condition, respondents could scan forward and backward to different items and pages and correct their answers freely. The word font and colour of text were identical in the CA and PPA questionnaire presentations. The difference between them was that the number of items presented on the CA screen each time was less than the number of items per page in PPA questionnaire. Also, in the CA questionnaire, each item could only be given one choice no matter how many categories were listed for the item. In

the PPA questionnaire, however, respondents could not be forced to select a unique category to every item, though they were asked to do so. Each subject was required to complete both PPA and CA questionnaires, so presentation mode was a within-subject variable (see Table 1).

The second independent variable, the respondent's ethnic origin and language of the questionnaire, was composed of three levels: Canadian subjects with an English language questionnaire (C/E), Hong Kong subjects with an English language questionnaire (HK/E), and Hong Kong subjects with a Chinese language questionnaire (HK/C). The Chinese language questionnaire was translated from English according to the direct meaning of each question item. Since there were different subjects in each group, this was a between-subject variable (see Table 1).

Table 1. Assignment of subjects in sub-experiment 1 and 2

	Computer/ Administered (CA)	Paper-and-pencil/ Administered (PPA)
Canadian/English (C/E) n=32	Subject n1---n32	Subject n1---n32
HongKong/English (HK/E) n=16	Subject n33--n48	Subject n33--n48
HongKong/Chinese (HK/C) n=16	Subject n49--n64	Subject n49--n64

The dependent variable in the first sub-experiment which assessed SD scores was the total count of the scores for the 24 items from the shortened Marlowe and Crown Social Desirability scale. Respondents obtained one point if they answered "true" to a socially desirable item or answered "false" to a social undesirable item. Thus the higher the SD score, the more tendency to SD.

In the literature review, three methods of counting the score of Extreme Response Style scale were introduced. Each method has been used in previous research. The present study used all the three methods so that the test of ERS could be compared with all the previous research. Consequently, there were three measures of the dependent variable in the second sub-experiment which assessed the tendency to give extreme responses. One measure was the respondent's standard deviation of the 14 items in the ERS scale. The higher the value of the scale, the more tendency to extreme responses of the respondent. The second measure of the dependent variable was the average deviation of the 14 items from the mid-point, i.e., 0 in the present ERS scale. If a respondent selected interval -2, for example, his/her deviation value on this item was $|-2 - 0| = 2$. The third measure was the number of items on which a respondent circled a -3 or a 3.

The four open questions in two versions in the third sub-experiment were analysed separately. In order to solve the problem of order effect, half of the respondents answered version A in the CA questionnaire and version B in the PPA questionnaire. The other half answered version A in the PPA, and answered version B in the CA. The two dependent variables for the four open questions were: (1) the total number of words for each question, and (2) the number of list items for each question. In question 46-A, specifically, the list items were the characteristics of the respondent's personality. In question 47-A, the items were the major problems facing the world today. In question 46-B, they were the suggestions of balancing the budget for the university. In question 47-B, the list items were the qualities of the (future) spouse.

The methods of inputting Chinese characters on a keyboard are complicated and not standardized, so most Hong Kong respondents did not know how to type Chinese characters on a keyboard. As a result, all HongKong /Chinese group subjects in the CA condition were not required to answer open questions. They only answered open questions in the PPA condition. In addition, since the component unit of the word in English and the character of Chinese was not equivalent, it is meaningless to compare the number of words between English and Chinese language. So the only dependent variable used to compare the HongKong /Chinese to the other groups was the number of list items in each question.

Therefore, the third sub-experiment for open questions were separated into two parts. One was sub-experiment 3A, which was a 2x2 two-way multivariate mixed design. The two independent variables were the treatment of questionnaire presentation (CA and PPA) and the ethnic origin groups of subjects (C/E and HK/E). The questionnaires were presented only in English. The two dependent variables were the number of words and the number of list items for each open-ended question (see Table 2.1). Another was sub-experiment 3B, which was a one-way univariate analysis for three ethnic groups(C/E, HK/E, HK/C) with only paper-and-pencil questionnaire (PPA). The unique dependent variable was the number of list items to each open question. The subjects in Canadian /English and HongKong /English group were the same subjects in the sub-experiment 3A. The additional ethnic group was the HongKong/Chinese group (see Table 2.2).

Table 2.1 Assignment of subjects in sub-experiment 3.
(Sub-experiment 3A)

	Computer/ Administered (CA)	Paper-and-pencil/ Administered (PPA)
Canadian/English (C/E) n=32	Subject n1---n32	Subject n1---n32
HongKong/English (HK/E) n=16	Subject n33--n48	Subject n33--n48

Table 2.2 Assignment of subjects in sub-experiment 3.
(Sub-experiment 3B)

	Paper-and-pencil/ Administered (PPA)
Canadian/English (C/E) n=32	Subject n1---n32
HongKong/English (HK/E) n=16	Subject n33--n48
HongKong/Chinese (HK/C) n=16	Subject n49--n64

Following the identical instructions on how to complete the questionnaires in both PPA and CA modes, respondents completed the questionnaires individually in the lab. Half of the respondents (16 C/E, 8 HK/E, and 8 HK/C) were randomly assigned to complete the questionnaires first in PPA mode, then in CA mode. The other half of the respondents completed them in the opposite sequence. The interval between the PPA and CA modes was 7 days or longer. After completing the questionnaire the second time, respondents were debriefed and all their questions about this experiment were answered.

Results

Social Desirability

The SD score for each respondent was the sum of the 24 ratings of the 24-item SD scale. The means of the SD scores for the two administration mode groups were very close (CA = 11.17, PPA = 10.36). The mean scores for the three ethnic origin/ language groups were: C/E = 10.23, HK/E = 10.62, HK/C = 11.97. The differences were quite small even though the Hong Kong respondents had a relatively higher average score on the SD scale, especially when the questions were represented in Chinese language. It was thus not surprising that the analysis of variance revealed no significant main effect for the mode of administration, $F(2, 122) = 1.90, p = .153$, or for the ethnic origin/ language groups (C/E, HK/E, and HK/C), $F(1, 122) = 1.14, p = .287$. In addition, the interaction effect of these two main effect was not significant, $F(2, 122) = .166, p = .847$. Thus the results did not indicate that Canadian sample and Hong Kong Chinese sample differed in their social desirability and did not indicate that SD was affected by mode of questionnaire administration. An ANOVA also indicated that there was no significant difference between male and female respondents on their SD-scale score, $F(1, 126) = .201, p = .654$.

In order to investigate order effect on social desirability, the present study compared the SD scores for computer questionnaires conducted in the first week with the SD scores for computer questionnaires from the second week but did not find any significant difference. Similarly, no significant difference was found between the SD scores for paper-and-pencil questionnaires conducted in the two weeks. The results indicate that order of presentation did not affect the measurement on social desirability.

Extreme Response Style

The means and standard deviations of the three ERS scores was shown in Table 3. When the scale score was indicated by the sum of deviations of the 14-item scale, the overall mean of the Canadian group was the highest ($C/E = 22.28$), which was much higher than the Hong Kong English group ($HK/E = 18.22$) and the Hong Kong Chinese group ($HK/C = 17.88$). This was consistent with the finding of ANOVA which demonstrated a significant main effect of the ethnic and language condition, $F(2, 122) = 13.19, p < .001$. The main effect of administration mode was not significant ($CA = 20.50$, $PPA = 19.83$), $F(1, 122) = 1.16, p = .284$, nor was the interaction effect, $F(2, 122) = 10.23, p = .627$.

When the ERS score was measured by the respondent's standard deviation, the outcome was equivalent. By using the standard deviation of the 14 scores from the ERS scale as each respondents' ERS score, ANOVA indicated that the ethnic/ language main effect was significant, $F(2, 122) = 5.33, p < .006$, but neither the main effect of administration mode, $F(1, 122) = .003, p = .959$, nor the interaction effect, $F(2, 122) = .161, p = .852$, were significant.

The third method of measuring ERS was by counting the number of extreme categories, "3" and "-3". The result was consistent with those of the previous two methods, that is, the only marked effect was the ethnic origin/ language main effect, $F(2, 122) = 6.39, p < .002$. The main effect of administration mode, $F(1, 122) = .694, p = .406$, and the interaction effect, $F(2, 122) = .179, p = .836$, were not significant.

Table 3. Descriptive statistics for ERS

Measure	CA (computer)		PPA (paper-and-pencil)		Overall	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Y-DEV (ERS by sum of deviation $ x-0 $)						
C/E	22.22	5.11	22.34	4.65	22.28	4.85
HK/E	19.00	4.29	17.43	4.05	18.22	4.18
HK/C	18.56	4.24	17.19	5.11	17.88	4.67
Overall	20.50	4.95	19.83	5.21	20.16	5.08
Y-STD (ERS by standard deviation)						
C/E	1.67	.35	1.71	.38	1.69	.36
HK/E	1.50	.34	1.46	.30	1.48	.32
HK/C	1.50	.39	1.51	.28	1.51	.33
Overall	1.59	.36	1.60	.35	1.59	.36
Y-COUNT (ERS by counting 3&(-3))						
C/E	3.06	2.35	3.03	2.51	3.05	2.41
HK/E	1.94	2.08	1.44	1.50	1.69	1.80
HK/C	2.00	1.55	1.56	1.32	1.78	1.43
Overall	2.52	2.15	2.27	2.15	2.39	2.15

Table 3 shows that the difference in ERS score came from the Canadian group versus the two Hong Kong groups; the two HK groups had a similar averages on ERS (e.g., C/E = 3.05, HK/E = 1.69, HK/C = 1.78 when ERS was scored by counting the extreme categories), and the Post Hoc test specifically demonstrated that the Canadian respondents tended to select more extreme categories than both groups of Hong Kong Chinese respondents. More specifically, the Canadian group circled markedly more 3 (strongly agree) or -3 (strongly disagree) categories than the two Hong Kong groups did. But as with the other measures of ERS, there was no systematic difference between computer administration and paper-and-pencil administration, and no interaction effect of administration and cultural/ language.

The effect of respondents' sex was tested but no significant result was found for any of the three ERS measures. When ERS was scored by counting the 3 and -3 categories,

for example, $F < 1$. In summary, the only systematic difference for ERS was the two ethnic groups.

I also tested for order effects for the computer questionnaires and the paper-and-pencil questionnaires between the two administrations. The non-significant difference indicated that, order did not affect the measurement of ERS.

Open Questions

The analysis of open questions was separated into two independent sections. One section, sub-experiment 3A, was tested by a 2x2 multivariate analysis of variance with two dependent variables: the number of words and the number of list items for each question. The ethnic origin/ language variable only included the Canadian group (C/E) and the Hong Kong English group (HK/E) (see Table 2.1). Hong Kong Chinese group (HK/C) was deleted in this MANOVA but was included in the second section, sub-experiment 3B, in which only one dependent variable, the number of list items, was employed (see Table 2.2). Since the style and content of the four open questions were different, four tests were conducted separately for questions 46-A, 47-A, 46-B and 47-B. In the original questionnaire, one question asking for the respondent's frequency of using a word processor each week was used as a covariate. After the outcomes of the analysis of variance (ANOVA) and the analysis of covariance (ANCOVA) were compared, it was found that the results for the two main effects and the interaction effect were correspondingly equivalent in the two analysis methods. Therefore, only the results of the analysis of variance are reported here.

Consistent with the findings for ERS, the test of MANOVA for question 46-A, the description of one's own personality characteristics demonstrated a significant multivariate difference between the two ethnic groups, Wilks' Lambda = 5.29, $p < .009$. This multivariate difference came from both unique variables, that is, the number of words, $F(1, 44) = 6.15$, $p < .017$, and the number of listed items, $F(1, 44) = 9.31$, $p < .004$. Canadian respondents wrote more words in question 46-A than did the Hong Kong Chinese respondents ($C/E = 82.88$, $HK/E = 43.88$), and the number of listed items of the Canadian respondents was more than the Hong Kong respondents ($C/E = 8.34$, $HK/E = 4.56$) (see Table 4). But the administration mode did not have a significant effect on the answers in question 46-A with Wilks' Lambda = 1.01, $p = .371$, even though the actual average of words in CA mode (85.29) was much higher than the PPA mode (54.46). The interaction effect had no multivariate systematic difference, either, with Wilks' Lambda = 1.53, $p = .228$. However, the descriptive statistics displayed an interesting result for the two dependent variables: the Canadian groups gave answers with many more words in the computer questionnaire than in the paper-and-pencil questionnaire ($CA = 107.56$, $PPA = 58.19$), while the Hong Kong group answered question 46-A with the similar lengths in two mode conditions ($CA = 40.75$, $PPA = 47.00$) (see Table 4).

It thus may be concluded for question 46-A that the Canadian respondents give longer answers to open questions than do Hong Kong respondents, including more words and items. Although the administration mode did not significantly affect the answers, the Canadian respondents tended to answer the question with more words and items in the

computerized questionnaire than in the paper-and-pencil questionnaire, while the Hong Kong respondents did not show such a difference.

Table 4. Descriptive statistics for open question 46-A

Measure	CA (computer)		PPA (paper-and-pencil)		Overall	
	M(cases)	Std Dev	M(cases)	Std Dev	M(cases)	Std Dev
Y- number of words in 46-A						
C/E	107.56 (16)	72.36	58.19 (16)	42.12	82.88 (32)	63.41
HK/E	40.75 (8)	32.88	47.00 (8)	21.79	43.88 (16)	27.14
Overall	85.29 (24)	69.13	54.46 (24)	36.48	69.88 (48)	56.86
Y- number of list items in 46-A						
C/E	8.94 (16)	4.06	7.75 (16)	5.29	8.34 (32)	4.67
HK/E	4.25 (8)	2.49	4.88 (8)	1.25	4.56 (16)	1.93
Overall	7.38 (24)	4.21	6.79 (24)	4.54	7.08 (48)	4.34

Table 5. Descriptive statistics for open question 46-B

Measure	CA (computer)		PPA (paper-and-pencil)		Overall	
	M(cases)	Std Dev	M(cases)	Std Dev	M(cases)	Std Dev
Y- number of words in 46-B						
C/E	84.19 (16)	47.72	62.63 (16)	41.39	73.41 (32)	45.29
HK/E	40.88 (8)	24.69	36.25 (8)	36.33	38.56 (16)	30.10
Overall	69.75 (24)	45.89	53.83 (24)	40.99	61.79 (48)	43.79
Y- number of list items in 46-B						
C/E	3.44 (16)	1.79	3.20 (15)	2.18	3.32 (31)	1.96
HK/E	3.25 (8)	1.83	4.00 (8)	4.14	3.63 (16)	3.12
Overall	3.38 (24)	1.76	3.48 (23)	2.94	3.43 (47)	2.38

The overall results from MANOVA for question 46-B were quite similar to those of question 46-A. The interaction effect with Wilks' Lambda = .61, $p = .548$, and the main effect of administration mode with Wilks' Lambda = 1.42, $p = .253$, were not statistically significant. However, the descriptive statistics (see Table 5) continued to demonstrate that respondents tended to answer the question with more words on average in the computer condition than in the paper-and-pencil condition (CA = 69.75, PPA = 53.83), and that this

tendency was greater for Canadians ($CA = 84.19$, $PPA = 62.63$) than for Hong Kong Chinese ($CA = 40.88$, $PPA = 36.25$). The only multivariate significant difference was between the two ethnic groups with Wilks' Lambda = 4.48, $p < .017$, and this multivariate difference came from the unique variable of the number of words for the current question ($C/E = 73.41$, $HK/E = 38.56$), $F(1, 43) = 7.34$, $p < .010$. The number of items listed by two groups, that is, the suggestions for balancing the budget of the university, had no systematic difference. In conclusion, the Canadian respondents replied to question 46-B with more words than did Hong Kong respondents, but the number of listed items in the answers were basically equivalent between the two groups.

Question 47-A asked the major problems facing the world today. It might not be as easy to answer this question as to answer the other three open questions. The response showed no multivariately significant difference between the Canadian group and the Hong Kong group with Wilks' Lambda = 1.87, $p = .167$, neither difference for the administration mode with Wilks' Lambda = 2.14, $p = .131$, or for the interaction effect, Wilks' Lambda = .76, $p = .476$. Even though, the actual values of the average words for two groups ($C/E = 51.59$, $HK/E = 27.53$) and the means of list items ($C/E = 4.72$, $HK/E = 3.93$) described a consistent tendency as the results of the previous two questions, in which the Canadian respondents answered the question with more words and more list items than did the Hong Kong respondents. Again, the Canadians wrote more using computers than writing by hand, both words ($CA = 73.13$, $PPA = 30.06$) and list items ($CA = 5.13$, $PPA = 4.31$). But the Hong Kong Chinese demonstrated a mixed result in which their average words were slightly higher in computer mode ($CA = 34.88$, $PPA =$

19.14) while the list items in PPA mode was higher (CA = 3.75, PPA = 4.14) (see Table 6). It implies that Hong Kong respondents had a more concise replies in paper-and-pencil questionnaire although this difference was not significant. The administration mode had no significant effect on the Hong Kong sample for this question.

Table 6. Descriptive statistics for open question 47-A

Measure	CA (computer)		PPA (paper-and-pencil)		Overall	
	M(cases)	Std Dev	M(cases)	Std Dev	M(cases)	Std Dev
Y- number of words in 47-A						
C/E	73.13 (16)	65.35	30.06 (16)	26.54	51.59 (32)	53.72
HK/E	34.88 (8)	37.71	19.14 (7)	19.37	27.53 (15)	30.62
Overall	60.38 (24)	59.65	26.74 (23)	24.68	43.92 (47)	48.57
Y- number of list items in 47-A						
C/E	5.13 (16)	2.94	4.31 (16)	1.08	4.72 (32)	2.22
HK/E	3.75 (8)	1.49	4.14 (7)	.69	3.93 (15)	1.16
Overall	4.67 (24)	2.60	4.26 (23)	.96	4.47 (47)	1.97

Table 7. Descriptive statistics for open question 47-B

Measure	CA (computer)		PPA (paper-and-pencil)		Overall	
	M(cases)	Std Dev	M(cases)	Std Dev	M(cases)	Std Dev
Y- number of words in 47-B						
C/E	48.94 (16)	57.27	53.81 (16)	50.37	51.38 (32)	53.11
HK/E	31.25 (8)	18.48	31.25 (8)	18.06	31.25 (16)	17.65
Overall	43.04 (24)	48.12	46.29 (24)	43.26	44.67 (48)	45.30
Y- number of list items in 47-B						
C/E	8.69 (16)	4.60	8.33 (15)	3.85	8.52 (31)	4.19
HK/E	5.63 (8)	2.00	5.25 (8)	2.05	5.44 (16)	1.97
Overall	7.67 (24)	4.15	7.26 (23)	3.61	7.47 (47)	3.86

Question 47-B, which was about the qualities the respondent wants for his/her (future) spouse, generally showed the same tendency as the previous findings. The only multivariate significance was the ethnic main effect, with Wilks' Lambda = 3.72, $p < .033$.

But the difference was contributed by the list items ($C/E = 8.52$, $HK/E = 5.44$), $F(1, 43) = 7.36$, $p < .010$. It should be pointed out that the original average words to this question in the computer questionnaire was slightly fewer than in paper-and-pencil questionnaire ($CA = 43.04$, $PPA = 46.29$) (see Table 7).

Table 8. Descriptive statistics for open questions on main effect of ethnic origin/ language

Measure	M (cases)	Std Dev
Y- number of list items in 46-A		
C/E (Canadian/English)	8.34 (32)	4.67
HK/E (HongKong/English)	4.56 (16)	1.93
HK/C (HongKong/Chinese)	3.25 (8)	3.01
Overall	6.54 (56)	4.37
Y- number of list items in 46-B		
C/E (Canadian/English)	3.32 (31)	1.96
HK/E (HongKong/English)	3.63 (16)	3.12
HK/C (HongKong/Chinese)	2.88 (8)	1.13
Overall	3.35 (55)	2.25
Y- number of list items in 47-A		
C/E (Canadian/English)	4.72 (32)	2.22
HK/E (HongKong/English)	3.93 (15)	1.16
HK/C (HongKong/Chinese)	5.88 (8)	1.25
Overall	4.67 (55)	1.93
Y- number of list items in 47-B		
C/E (Canadian/English)	8.52 (31)	4.19
HK/E (HongKong/English)	5.44 (16)	1.97
HK/C (HongKong/Chinese)	4.13 (8)	2.30
Overall	6.98 (55)	3.84

The descriptive test for sub-experiment 3B represents the averages of list items for open question 46-A ($C/E = 8.34$, $HK/E = 4.56$, $HK/C = 3.25$) (see Table 8). The test for sub-experiment 3B employed an ANOVA to compare the number of list items across the three groups (C/E , HK/E , & HK/C). The difference among the three ethnic/ language groups for question 46-A was significant with $F(2, 53) = 8.43$, $p < .001$. This was

consistent with the findings for question 46-A in sub-experiment 3A. Post Hoc test displayed that the significant difference was between the Canadian group with each Hong Kong group.

In addition to the descriptive and inferential statistics for the four open questions, crosstabs were calculated for content analysis on question 46-A and 47-B. The findings could be helpful to understand the cultural reasons that affected the ethnic differences on ERS and open questions.

In question 46-A, the description of self personality characteristics, the most frequently appearing ten words were listed (see Table 9). After the test of Pearson Chi-square statistic, or the binomial test when the cell number is too small, it was found that systematically more Canadian respondents liked to describe themselves as “friendly” ($p < .039$), “humorous” ($p < .001$), and “caring” ($p < .035$). The only word the Hong Kong respondents used significantly more than the Canadians was “lazy” ($p < .008$).

Table 9. Distribution and frequency of words used in question 46-A (N=64)

Words	C/E	HK/E+HK/C	Overall	Percent* in N (%)
Friendly	16	8	24	37.5
Helpful	9	4	13	20.3
Humorous	12	1	13	20.3
Extrovert	7	5	12	18.8
Caring	7	1	8	12.5
Lazy	0	7	7	10.9
Independent	3	3	6	9.4
Introvert, shy	3	3	6	9.4
Intelligent	4	1	5	7.8
Creative	4	0	4	6.3

* Percent = (Overall / N) × 100.

Question 47-B was about the qualities desired of one's (future) spouse. The 14 most frequently cited qualities are listed in Table 10. Six of them were significantly different between the two ethnic samples using a Pearson Chi square test. Among them, Canadian respondents wished their (future) spouse to be "humorous" or "funny" ($\chi^2 = 27.71, p < .000$), "honest" ($\chi^2 = 4.95, p < .026$), "intelligent" or "smart" ($\chi^2 = 9.69, p < .002$), "(physical) attractive" ($\chi^2 = 16.31, p < .000$), and "trustworthy" ($\chi^2 = 7.82, p < .005$). The only special quality for the Hong Kong respondents was "loving my parents" ($\chi^2 = 13.28, p < .000$).

Table 10. Distribution and frequency of words used in question 47-B (N=64)

Words	C/E	HK/E+HK/C	Overall	Percent in N (%)
Good hearted, kind, nice	10	14	24	37.5
Humorous, funny	21	1	22	34.4
Good look/size	9	11	20	31.3
Honest	13	5	18	28.1
Loving me	7	11	18	28.1
Intelligent, smart	14	3	17	26.6
Caring	10	6	16	25.0
Attractive (e.g. physical)	13	0	13	20.3
Trustworthy	11	2	13	20.3
Understanding	7	6	13	20.3
Loving children	7	4	11	17.2
Loving my parents	0	11	11	17.2
Supporting	4	5	9	14.1
Cook, laundry	1	5	6	9.4

In question 47-B, the highest frequency of the word used by the Canadians was "humorous" ($n = 21$), which also frequently appeared in question 46-A ($n = 12$) (see Table 9). It can be concluded that the sense of humor was regarded as a very important personality trait in Canadian culture. Meanwhile, the Canadian respondents preferred to

depict themselves positively as “friendly” or “easygoing” ($n = 16$), “helpful” ($n = 9$), “caring” ($n = 7$), “intelligent” ($n = 4$), “creative” ($n = 4$) (see Table 9). The mentioned negative traits of the Canadians were such as “stubborn”, “too demanding”, “bad temper”, and “casual”. The Hong Kong Chinese seldom evaluated themselves to be “humorous” ($n = 1$), “intelligent” ($n = 1$), or “creative” ($n = 0$). On the contrary, they listed more negative aspects of their characteristics. The highest frequently used words was “lazy” ($n = 7$). Others included “impatient”, “greedy”, “less confident”, “jealous”, and “bad mood”.

Discussion

Social Desirability

Previous research has found that, when compared to paper-and-pencil questionnaire, computerized surveys result in fewer item completion errors, more correct coding, and speedy data analysis (Kiesler & Sproull, 1986; Rosenfeld, etc., 1989). The present study found that the computer and the paper-and-pencil modes of administration did not show significant difference in social desirability. This finding is consistent with much of the published research literature (Davis & Cowles, 1989; Millstein, 1987; Booth-Kewley, Edwards, & Rosenfeld, 1992; Finegan & Allen, 1994; Whitener & Klein, 1995). But it is contrary to the results of Kiesler and Sproull (1986) and Martin and Nagao (1989) who found a lower score of SD scale in the computer administration, as well as differs to the finding of Lautenschlager and Flaherty (1990) who found more SD responding in the computer mode. The contradictions may reflect a number of procedural and methodological differences between the studies. For instance, different studies employed different SD scales; the number of items of the same SD scale varied among the different research which might affect the reliability and validity of the assessment; the interface of the computerized questionnaire in the current study was much more similar to the paper-and-pencil questionnaire; the current study provided a private environment to respondents, so that anonymity conditions which may affect social desirability bias were equivalent in both the computer and the paper-and-pencil administration conditions. Furthermore, knowledge of computer applications was far more common in 1996 than in 1986 and 1989 when Kiesler's research was done; respondents in the present study were university

students who used computers frequently and might be more skilled and feel more comfortable on computer applications than people did ten years ago.

Comparisons between the Canadian sample and the Hong Kong Chinese on social desirability did not show significant differences, either. This contradicts the findings of Dyal and Chan (1985) who found more SD among Hong Kong Chinese students, and contrary to the result of Ross and Mirowsky's (1984) study in which the Mexican Americans gave more SD responses than the Anglos did and the researchers attributed it to the relatively powerless positions of the Mexican Americans in the society.

The Hong Kong respondents in the present study were international students or new immigrants who had been in Canada for no more than five years (mean = 3.09 years, range from 9 months to 5.5 years). Because of the current immigration laws of Canada, new immigrants from Hong Kong usually are likely to have financial investments in Canada or be professionals. Thus, most of these students' parents are likely middle-class professionals or have business in Canada and/or Hong Kong. Therefore the Hong Kong Chinese sample might not feel powerless in the society as the Mexican Americans did and thus not satisfy Ross and Mirowsky's powerless condition for high SD. Additionally, the Hong Kong sample in the present study was influenced by both (1) Chinese culture and social values and (2) Canadian culture and social values, and might have become more "Canadian" in expressing their real thoughts. Also, the anonymous experimental condition allowed the Chinese participants to worry less about keeping a "good face" and this may have lowered their tendency to respond with socially desirable answers.

Although researchers are often disappointed by insignificant differences, there are advantages associated with the present negative findings. The results indicate that researchers can select the computer administered surveys instead of the traditional paper-and-pencil administered surveys without worrying that it will lead to more socially desirable responses. The present study indicates that, as the financial and logistical considerations allow, computerized surveys are recommended because of their many previously mentioned advantages. And when the survey sample includes Hong Kong Chinese, the bias from social desirability should not be a source of concern.

Extreme Response Style

The results of the present study on Extreme Response Style (ERS) supported the two hypotheses. One was that the administration mode did not affect respondents' ERS. This finding was consistent with Kiesler and Sproull (1986) who found no systematic difference on extremity of responses between computerized and paper-and-pencil surveys. The second hypothesis concerned the cultural effects of ERS. In the present study, the Canadian respondents demonstrated significantly more ERS than did the Hong Kong Chinese. This result is similar to those of several cross-cultural studies which found significant differences on ERS among various ethnic groups (Hui & Triandis, 1989; Marin, Gamba, & Marin, 1992; Zax & Takahashi, 1967; Chun, Campbell, & Yoo, 1974; Stening & Everett, 1984; Bachman & O'Malley, 1984).

Previous studies provide evidence that ERS is a stable response style related to respondents' personal traits such as age, education level, and race (Shulman, 1973; Stening & Everett, 1984; Marin, Gamba, & Marin, 1992). The findings in the present

study support their conclusions for cross-ethnic differences on ERS. However, reasons for the obtained differences are unclear. According to Stening and Everett's (1984) explanation, it was due to the different languages of the questionnaires, which is similar with Bachman and O'Malley's (1984) argument which attributes it to different linguistic styles. In the present study, the Hong Kong sample was assessed on ERS scale both in English and Chinese language, and the result between two Hong Kong groups with different languages was quite similar. Thus it appears that language is not the direct cause of ethnic differences in ERS. Zax and Takahashi's (1967) discussion, which attributes the racial differences in ERS to child-rearing methods, was not useful in accounting for the present research findings, because the relationship between child-rearing method and respondents' ERS tendency was unknown. Hui & Triandis (1989) argue that ethnic differences are due to respondents' judgment styles, but they do not address why people with different cultural backgrounds demonstrate different judgment styles on multi-point Likert scales.

As previously cited, the present study assumed that it was respondents' different cultural backgrounds that led to their systematic differences in ERS. The Hong Kong respondents were educated in a Chinese cultural environment. Chinese culture and social values emphasize traits such as humility, prudence and moderation, which means following the middle way (Chinese Culture Connection, 1987). The influence of Chinese values likely spreads into Chinese personality characteristics, attitudes, judgment style and interpersonal relationships.

Thus, the “middle way” could be reflected in Likert scale midpoints and hence low ERS. The results from the open questions in the present study also provided evidence for the effect of the Chinese culture and social values. For example, fewer individuals in the Hong Kong group than in the Canadian group described themselves to be intelligent, creative, friendly or nice, or having a sense of humour. More Hong Kong respondents listed at least one negative trait such as lazy, perhaps because Chinese values regard humility as one of the most important virtues. Since moderation is a central aspect of Chinese culture, it is not surprising that the Hong Kong Chinese gave answers low in ERS.

Based on the results, it can be advised that ERS is not related to the survey administration method. When the advantages of computer administration in the process of survey conduction is considered, computer administered surveys are thus recommended. It should be noticed that when a survey sample is composed of respondents who are from a variety of ethnic backgrounds, especially of Chinese respondents, the examinations of attitude differences between the ethnic groups should be made cautiously, because the responses may not only reflect respondents' attitudes, but also reflect their cultural characteristics. For instance, if both the choice “somewhat disagree” and “disagree” are provided simultaneously, the Chinese respondents may likely use “somewhat disagree” instead of “disagree” to express their disagreement. When they choose “disagree”, the degree of their disagreement may be greater than that of the Canadian respondents. Thus the same answer may not be psychologically equivalent, since Chinese people tend to represent their attitudes moderately. The real differences of an attitude between the two

ethnic groups may be covered if the conclusion is simply made according to the same response data. Consequently, it is suggested to explore alternative scoring methods in future research when the survey questionnaire use multi-point scales such as Likert scale.

Open Questions

Because of the difficulties of research on open-ended questions, previous studies on this topic are rare. The present study also faced some of these difficulties. First, it is not feasible to use repeated-measures designs for open questions, because respondents may simply repeat their replies if they were asked to answer the same open questions twice. Respondents may also feel angry or bored by writing the same answers to the same open questions a second time. In consultation with my prospectus committee, the present study employed different open questions in the computer administered and the paper-and-pencil administered questionnaires. The researcher of the present study hoped the answer styles across open questions remained the same so that the answers were comparable, since what is of concern in the open answers are the number of words and list items in each answer. However, respondents may also give the same rating on the rating scales in a repeated-measure experiment. In the future research, it would be interesting to compare the written answers for open questions asked twice and the different rating scales for measuring the same concept.

The second difficulty of research on open questions concerned coding the answers. Some respondents wrote the answer in a continuous paragraph instead of in relatively clear list of items. And some answers in the list item styles actually had more than one trait

or suggestion in one item. The present study counted the list items based on the identical standards for all respondents, that is, the number of traits or suggestions according to the real meaning, not just simply counted the items the respondents provided or the number of sentences in the paragraph.

Similar to the results for SD and ERS, administration mode did not have a statistically significant effect on open questions. But in one question (Q47-A), the average number of words in the computer questionnaire answers was markedly greater than in the paper-and-pencil questionnaire answers. And in two of the three other questions, the average words displayed the same tendency, though without statistical significance. The number of list items did not indicate this tendency. It seems that in the computerized survey, respondents tend to write more words for each list item and more words for the whole question than in the paper-and-pencil survey. This is consistent with Kiesler and Sproull's (1986) finding that respondents wrote more words in the computerized open questions than in the paper-and-pencil when they could freely edit the answers in the computerized questionnaire. It is therefore suggested that when more words are desired in answers to open questions, computer questionnaires should be used. But the computer questionnaire has no more superiority on obtaining more list items than does the traditional paper-and-pencil questionnaire.

The present study also tried to explore the ethnic differences in answers to open-ended questions. Among the results of the four open-ended questions, three questions indicated an overall significant difference on words and list items between the two ethnic groups. Furthermore, among the non-significant results for the two unique dependent variables,

Canadian groups wrote more words and list items in the answers. Only in question 46-B the numbers of list items are very close between the two ethnic groups. So there does seem to be a tendency for Canadian respondents to write more words on average than Hong Kong respondents for open questions.

But what is the reason for ethnic differences in answers to open-ended questions? The most likely reason might be language differences. In the present study in which English and Chinese language were involved, it was difficult to compare the number of words between the two languages because the word units in English and Chinese are not identical. But the number of list items in the answers were more easily comparable. Canadian groups listed significantly more items than did the two Hong Kong groups completing English or Chinese language questionnaires. The similarity of the two Hong Kong groups indicates that the ethnic difference was not due to language difference but to something else. I think the most likely possibility is, again, cultural values. Since individualism is one of the important points in Western culture, Canadian people are likely to express individual feelings more freely than are Chinese people, and they are more likely to write their opinions. On the contrary, Chinese people, include the Hong Kong Chinese, are influenced by the Chinese cultural traits such as harmony with others, humility, prudence (carefulness), and so on. So it is not as acceptable to express individual opinions or thoughts. As a result, Hong Kong Chinese respondents would give fewer words and items in open questions than would the Canadian respondents. As with ERS, this finding suggests that in sample survey research with respondents composed of different cultural

backgrounds, it might be misleading simply to use quantitative measures such as the average of word numbers for open-ended questions. Alternative better scoring methods are necessary to be explored in the future research.

The findings from the content of the answers for the open-ended questions implies that there are differences on social and cultural values between Canadian culture and Chinese culture. Canadian people are more likely to depict themselves to be friendly, caring, intelligent, honest, attractive, and most important, to have a sense of humour. They are more likely to describe themselves positively, while the Hong Kong Chinese seem to have more humility and moderation. Hong Kong respondents describe themselves with fewer personality traits and relatively more negative traits (see Table 4). This is consistent with the social norms of the Chinese culture noted previously. This implication supports the findings and explanations of the present research on extreme response style and open-ended questions.

In summary, the present study demonstrated how survey administration method and cultural /language differences affect measurement on SD, ERS, and open-ended questions. It found that response effects including social desirability and extreme response style would not additionally bias computer administered surveys compared to the paper-and-pencil administered surveys. But it found a significant difference between respondents from different cultural backgrounds on ERS and open questions which implied caution in scoring and statistical comparison of cultural groups in future survey research. There are also some problems and limitations in the present study. For example, the sample size was relatively small, the contents of open questions were hard to keep identical and equivalent

for every respondent; there was no proper scoring method for open questions in different languages. It might be the task of the future research to solve these problems so as to make the research more accurate.

Reference

- Album, G., & Murphy, B. D. (1988). Extreme response on a Likert scale. *Psychological Reports*, 63, 501-502.
- Alreck, P., & Settle, R. (1985). *The Survey Research Handbook*. Homewood: Richard D. Irwin Inc.
- Anderson, A., Basilevsky, A., & Hum, D. Measurement: Theory and techniques. In P. Rossi, J. D. Wright, A. Anderson (Eds.). *Handbook of survey research*. New York: Academic Press.
- Arthur, A. Z. (1966). Response bias in the semantic differential. *British Journal of Social and Clinical Psychology*, 5, 103-107.
- Bachman, J. G., & O'Malley, P. O. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly*, 48, 491-509.
- Bailey, K. D. (1978). *Methods of social research*. New York: Free Press
- Biffs, J. B. & Das, J. P. (1973). Extreme response set, internality-externality, and performance. *British Journal of Social and Clinical Psychology*, 12, 99-210.
- Bogardus, E. S. (1925). Measuring social distances. *Journal of Applied Sociology*, 9, 299-308.
- Booth-Kewley, S., Edwards, J. E., & Rosenfeld, P. (1992). Impression management, social desirability, and computer administration of attitude questionnaires: Does the computer make a difference? *Journal of Applied Psychology*, 77, 562-566.
- Bradburn, N. M. (1983). Response effects. In P. H. Rossi, J. D. Wright, A. Anderson (Eds.). *Handbook of Survey Research*. New York: Academic Press.
- Chinese Culture Connection. (1987). Chinese values and the search for culture-free dimensions of culture. *Journal of Cross-Cultural Psychology*, 18, 143-164.
- Chun, K. T., Campbell, J. B., & Yoo, J. H. (1974). Extreme response style in cross-cultural research: A reminder. *Journal of Cross-Cultural Psychology*, 5, 465-480.
- Clancy, K. & Gove, W. (1974). Sex differences in mental illness: An analysis of response bias in self-reports. *American Journal of Sociology*, 80, 205-216.

- Cooperstock, R. (1971). Sex differences in the use of mood-modifying drugs: An explanatory model. *Journal of Health and Social Behaviour*, 12, 238-244.
- Crandall, J. E. (1982). Social interest, extreme response style, and implications for adjustment. *Journal of Research in Personality*, 16, 82-89.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354.
- Crowne, D. P., & Marlowe, D. (1964). *The approval motive*. New York: Wiley.
- Davis, C., & Cowles, M. (1989). Automated psychological testing: Method of administration, need for approval, and measures of anxiety. *Educational and Psychological Testing*, 49, 311-320.
- Dohrenwend, B. S. (1965). Some effects of open and closed questions on respondents' answer. *Human Organization*, 24, 175-184.
- Dillman, D., Carpenter, E., Christensen, J., & Brooks, R. (1974). Increasing mail questionnaire response: A four state comparison. *American Sociological Review*, 39, 744-756.
- Dyal, J. A., & Chan, C. (1985). Stress and distress: A study of Hong Kong Chinese and Euro-Canadian students. *Journal of Cross-Cultural Psychology*, 16, 447-466.
- Edward, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden Press.
- Finegan, J. E., & Allen, N. J. (1994). Computerized and written questionnaires: Are they equivalent? *Computer in Human Behaviour*, 10, 483-496.
- Fox, R., Crask, M., & Kim, J. (1988). Mail survey response rate: A method analysis of selected techniques for increasing response. *Public Opinion Quarterly*, 52, 467-491.
- Fowler, F. J. (1993). *Survey Research Methods*. Sage Publications.
- Frankel, M. (1983). Sampling theory. In P. Rossi, J. D. Wright, A. Anderson (Eds.). *Handbook of survey research*. New York: Academic Press.
- Gallup, G., & Robinson, C. (1938). *A guide to public opinion polls*. Princeton: Princeton University Press.

- Greenleaf, E. A. (1992). Measuring extreme response style. *Public Opinion Quarterly*, 56, 328-351.
- Groves, R., & Kahn, R. (1979). *Surveys by telephone*. New York: Academic Press.
- Hamilton, D. H. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, 69, 192-203.
- Heide, M., & Gronhaug, K. (1992). The impact of response styles in surveys: A simulation study. *Journal of the market research society*, 34, 215-230.
- Heberlein, T. A & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaire: a quantitative analysis of the published literature. *American Sociological Review*. 43, 447-462.
- Herzog, A. R., & Rodgers, W. L. (1988). Interviewing older adults: Mode comparison using data from a face-to-face survey and a telephone re-survey. *Public Opinion Quarterly*, 52, 84-99.
- Hui, C. H. & Triandis, H. C. (1985). The instability of response sets. *Public Opinion Quarterly*, 49, 253-260.
- Hui, C. H. & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*. 20, 296-309.
- Innes, J. M. (1977). Extremity and "Don't know" sets in questionnaires. *British Journal of Social and Clinical Psychology*, 16, 9-12.
- Jawie, I. C., & Agassi, J. (1969). A study in westernization. In I. C. Jawie, & J. Agassi (Eds.). *Hong Kong: A Society in Transition*. 129-163.
- Jordan, I. A., Marcus, A. C., Reeder, L. G. (1980). Response styles in telephone and household interviewing: A field experiment. *Public Opinion Quarterly*, 44, 210-222.
- Kiesler, S., & Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly*, 50, 402-413.
- Kish, L. (1967). *Survey Sampling*. New York: John Wiley & Sons. Second Edition.
- Krysan, M., Schuman, H., Scott, L. J., & Beatty, P. (1994). Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly*, 58, 381-399.

- Largner, T. S. (1962). A twenty-two item screening score of psychiatric symptoms indicating impairment. *Journal of Health and Human Behaviour*, 3, 269-276.
- Lautenschlager, G. J. & Flaherty, V. L. (1990). Computer administration of questions: More desirable or more social desirability? *Journal of Applied Psychology*, 75, 310-314.
- Lazarsfeld, P. F. (1944). The Controversy over detailed interviews-an offer for negotiation. *Public Opinion Quarterly*, 8, 38-60.
- Li, P. S. (1981). *Social Research Methods*. Toronto: Butterworth.
- Ma, H. K. (1989). Moral orientation and moral judgment in adolescents in Hong Kong, Mainland China, and England. *Journal of Cross-Cultural Psychology*, 20, 152-177.
- Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of Cross-Cultural Psychology*, 23, 498-509.
- Martin, C. L. & Nagao, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology*, 74, 72-80.
- Merrens, M. M. (1970). Generality and stability of extreme response style. *Psychological Reports*, 27, 802.
- Merrens, M. M. (1971). Personality correlates of extreme response style: A function of method of assessment. *Journal of Social Psychology*, 85, 313-314.
- Millstein, S. G. (1987). Acceptability and reliability of sensitive information collected via computer interview. *Educational and Psychological Measurement*, 47, 523-533.
- Oskamp, S. (1977). *Attitudes and opinions*. Englewood Cliffs: Prentice-Hall.
- Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleitner & J. S. Wiggins (Eds.). *Personality assessment via questionnaires*. Berlin: Springer-Verlag.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P.R. Shaver, & L. S. Wrightsman (Eds.). *Measures of personality and social psychological attitudes*.

- Phillips, D. & Segal, B. (1969). Sexual status and psychiatric symptoms. *American Sociological Review*, 34, 58-72.
- Phillips, D. L., & Clancy, K. (1970). Response biases in field studies of mental illness. *American Sociological Review*, 35, 503-515.
- Phillips, D. L., & Clancy, K. j. (1971). Some effects of "social desirability" in survey studies. *American Journal of Sociology*, 77, 921-940.
- Ralston, D. A., Gustafson, D. J., Elsass, P. M., Cheung, F., & Terpstra, R. H. (1992). Eastern values: A comparison of managers in the United States, Hong Kong, and the People's Republic of China. *Journal of Applied Psychology*, 77, 664-671.
- Ross, C. E. & Mirowsky, J. (1984). Social-desirable response and acquiescence in a cross-cultural survey of mental health. *Journal of Health and Social Behaviour*, 25, 189-197.
- Rossi, P. H., Wright, J. D., & Anderson, A. B. (1983). Sample surveys: History, current practice, and future prospects. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.). *Handbook of survey research*. New York: Academic Press.
- Saris, W. E. (1991). *Computer-assisted interview*. Newbury Park, CA: Sage.
- Serrill, M. S. (1995). Special report a nation blessed, a nation stressed. *Time*, 146, 20-41.
- Sheatsley, P. B. (1983). Questionnaire construction and item writing. In P. Rossi, J.D. Wright, A. Anderson(Eds.). *Handbook of survey research*. New York: Academic Press.
- Shulman, A. (1973). A comparison of two scales on extremity response bias. *Public Opinion Quarterly*, 37, 407-412.
- Shuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.
- Shuman, H., Ludwic, J., & Krosnick, J. A. (1986). The perceived threat of nuclear war, salience, and open questions. *Public Opinions Quarterly*, 50, 519-536.
- Squire, P. (1988). Why the 1936 *Literary Digest* poll failed. *Public Opinion Quarterly*, 52, 125-133.
- Stening, B.W., & Everett, J.E. (1984). Response style in a cross-cultural managerial study. *The Journal of Social Psychology*, 122, 151-156.

- Stevens, S. S. (1951). Mathematics, measurement and psychophysics. In S.S. Stevens (ed.) *Handbook of Experimental Psychology*. New York: Wiley.
- Stouffer, S. A., etc. (1949). *The American Soldier: Combat and Its Aftermath*. New Jersey: Princeton University Press.
- Synodinos, N. E., & Brennan, J. M. (1988). Computer interactive interviewing in survey research. *Psychology & Marketing*, 5, 117-137.
- van der Kloot, W. A., Kroonenberg, P. M., & Bakker, D. (1985). Implicit theories of personality: Further evidence of extreme response style. *Multivariate Behavioural Research*, 20, 369-387.
- Warr, P. B. & Coffman, T. L. (1970). Personality, involvement, and extremity of judgment. *British Journal of Social and Clinical Psychology*, 9, 108-121.
- Whitener, E. M., & Klein, H. J. (1995). Equivalence of computerized and traditional research methods: The roles of scanning, social environment, and social desirability. *Computers in Human Behaviour*, 11, 65-75.
- Wong, G. Y. Y., & Olive, N. (1995). Acquisition of managerial values in the People's Republic of China and Hong Kong. *Journal of Cross-Cultural Psychology*, 26, 255-275.
- Wong, S. L. (1986). Modernization and Chinese culture in Hong Kong. *China Quarterly*, 106, 306-325.
- Zax, M., Gardiner, D. H., Lowy, D. G. (1964). Extreme response tendency as a function of emotional adjustment. *Journal of Abnormal and Social Psychology*, 69, 654-657.

APPENDIX I

Questionnaire A

Instructions: This questionnaire consists of 47 items. Each item is followed a few choices. For each item, please circle the choice that suits you best. Be sure to answer every question. Remember that the information provided here will be kept confidential.

1. Your major: _____
2. Year of study at the university: _____
3. Age: _____
4. Sex: (1) Male (2) Female
5. Indicate what ethnic group you belong to:
 (1) Native Canadian
 (2) Hong Kong Chinese
 (3) Other: _____
- *. If your are from Hong Kong, how long have you been in Canada? _____
6. How often do you use a word processor per week?
 (1) Less than once
 (2) 1 to 2 times
 (3) 3 to 4 times
 (4) 5 times or more
7. How do you rate your English proficiency?
 (1) poor
 (2) fair
 (3) good
 (4) excellent

Listed below is a number of statements concerning personal attitudes and traits. Please read each item and decide whether the statement is true or false as it pertains to you. Circle the choices you select.

8. No matter who I'm talking to, I'm always a good listener.
(1) True (2) False
9. I am always careful about my manner of dress.
(1) True (2) False
10. At times I have really insisted on having things my own way.
(1) True (2) False
11. I have never felt that I was punished without cause.
(1) True (2) False
12. When I don't know something I don't at all mind admitting it.
(1) True (2) False
13. I have never intensely disliked anyone.
(1) True (2) False
14. I can remember "playing sick" to get out of something.
(1) True (2) False
15. I would never think of letting someone else be punished for my wrongdoings.
(1) True (2) False
16. I have never deliberately said something that hurt someone's feelings.
(1) True (2) False
17. I never hesitate to go out of my way to help someone in trouble.
(1) True (2) False
18. There have been occasions when I took advantage of someone
(1) True (2) False
19. I am sometimes irritated by people who ask favours of me.
(1) True (2) False

20. There have been times when I felt like rebelling against people in authority even though I knew they were right.
(1) True (2) False
21. There have been times when I was quite jealous of the good fortune of others
(1) True (2) False
22. It is sometimes hard for me to go on with my work if I am not encouraged.
(1) True (2) False
23. I have never been upset when people expressed ideas very different from my own.
(1) True (2) False
24. On a few occasions, I have given up doing something because I thought too little of my ability.
(1) True (2) False
25. I sometimes feel resentful when I don't get my way.
(1) True (2) False
26. I'm always willing to admit it when I make a mistake.
(1) True (2) False
27. I like to gossip at times.
(1) True (2) False
28. I sometimes try to get even, rather than forgive and forget.
(1) True (2) False
29. If I could get into a movie without paying and be sure I was not seen, I would probably do it.
(1) True (2) False
30. I am always courteous, even to people who are disagreeable.
(1) True (2) False
31. On occasion I have had doubts about my ability to succeed in life.
(1) True (2) False

Using the scale below as a guide, read the following items and decide to what extent it is characteristic of your feelings and behaviour. Rate the degree to which each item is true to you. Then circle the number you select. Please do not omit any item.

For example: I like to chat with friends.

If you agree with this statement, you may circle the number (1) or (2) or (3). The bigger the number, the more you agree with the statement.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2) (3)	agree		

32. I work very hard most of the time.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2) (3)	agree		

33. Everyone should use a mouthwash to help control bad breath.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2) (3)	agree		

34. I will probably have more money to spend next year than I have now.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2) (3)	agree		

35. My days seem to follow a definite routine--eating meals at the same time each day, etc.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2) (3)	agree		

36. I eat more than I should.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2) (3)	agree		

37. Television is my primary form of entertainment.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2) (3)	agree		

38. I like to feel attractive to members of the opposite sex.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

39. Advertising insults my intelligence.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

40. A college education is very important for success in today's world.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

41. I try to avoid foods that are high in cholesterol.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

42. When I see a full ashtray or wastebasket, I want it emptied immediately.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

43. TV commercials place too much emphasis on sex.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

44. I like to visit places that are totally different from my home.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

45. Investing in the stock market is too risky for most families.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

Please answer the following questions on your own words.

46. In as many sentences as you wish, would you please recommend ways to balance the budget at your university?

47. Below please list qualities you want to find in your (future) spouse.

Thank you for your cooperation !

Questionnaire B

Instructions: This questionnaire consists of 47 items. Each item is followed a few choices. For each item, please circle the choice that suits you best. Be sure to answer every question. Remember that the information provided here will be kept confidential.

1. Your major: _____
2. Year of study at the university: _____
3. Age: _____
4. Sex: (1) Male (2) Female
5. Indicate what ethnic group you belong to:
 (1) Native Canadian
 (2) Hong Kong Chinese
 (3) Other: _____
- *. If you are from Hong Kong, how long have you been in Canada? _____
6. How often do you use a word processor per week?
 (1) Less than once
 (2) 1 to 2 times
 (3) 3 to 4 times
 (4) 5 times or more
7. How do you rate your English proficiency?
 (1) poor
 (2) fair
 (3) good
 (4) excellent

Listed below is a number of statements concerning personal attitudes and traits. Please read each item and decide whether the statement is true or false as it pertains to you. Circle the choices you select.

8. I never hesitate to go out of my way to help someone in trouble.
(1) True (2) False
9. It is sometimes hard for me to go on with my work if I am not encouraged.
(1) True (2) False
10. I have never intensely disliked anyone.
(1) True (2) False
11. On occasion I have had doubts about my ability to succeed in life.
(1) True (2) False
12. I sometimes feel resentful when I don't get my way.
(1) True (2) False
13. I am always careful about my manner of dress.
(1) True (2) False
14. If I could get into a movie without paying and be sure I was not seen, I would probably do it.
(1) True (2) False
15. On a few occasions, I have given up doing something because I thought too little of my ability.
(1) True (2) False
16. I like to gossip at times.
(1) True (2) False
17. There have been times when I felt like rebelling against people in authority even though I knew they were right.
(1) True (2) False
18. No matter who I'm talking to, I'm always a good listener.
(1) True (2) False

19. I can remember "playing sick" to get out of something.
(1) True (2) False
20. There have been occasions when I took advantage of someone.
(1) True (2) False
21. I'm always willing to admit it when I make a mistake.
(1) True (2) False
22. I sometimes try to get even, rather than forgive and forget.
(1) True (2) False
23. When I don't know something I don't at all mind admitting it.
(1) True (2) False
24. I am always courteous, even to people who are disagreeable.
(1) True (2) False
25. At times I have really insisted on having things my own way.
(1) True (2) False
26. I would never think of letting someone else be punished for my wrongdoings.
(1) True (2) False
27. I have never been upset when people expressed ideas very different from my own.
(1) True (2) False
28. There have been times when I was quite jealous of the good fortune of others.
(1) True (2) False
29. I am sometimes irritated by people who ask favours of me.
(1) True (2) False
30. I have never felt that I was punished without cause.
(1) True (2) False
31. I have never deliberately said something that hurt someone's feelings.
(1) True (2) False

Using the scale below as a guide, read the following items and decide to what extent it is characteristic of your feelings and behaviour. Rate the degree to which each item is true to you. Then circle the number you select. Please do not omit any item.

For example: I like to chat with friends.

If you agree with this statement, you may circle the number (1) or (2) or (3). The bigger the number, the more you agree with the statement.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2)	(3)	agree	

32. When I see a full ashtray or wastebasket, I want it emptied immediately.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2)	(3)	agree	

33. Television is my primary form of entertainment.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2)	(3)	agree	

34. I try to avoid foods that are high in cholesterol.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2)	(3)	agree	

35. Advertising insults my intelligence.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2)	(3)	agree	

36. Investing in the stock market is too risky for most families.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2)	(3)	agree	

37. Everyone should use a mouthwash to help control bad breath.

Strongly			Neutral			Strongly
disagree	(-3)	(-2)	(-1) (0) (1) (2)	(3)	agree	

38. TV commercials place too much emphasis on sex.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

39. A college education is very important for success in today's world.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

40. My days seem to follow a definite routine--eating meals at the same time each day, etc.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

41. I like to visit places that are totally different from my home.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

42. I work very hard most of the time.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

43. I like to feel attractive to members of the opposite sex.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

44. I will probably have more money to spend next year than I have now.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

45. I eat more than I should.

Strongly Neutral Strongly
disagree (-3) (-2) (-1) (0) (1) (2) (3) agree

Please answer the following questions on your own words.

46. In as many sentences as you wish, please describe the characteristics of your personality.

47. Below please list what you believe are the major problems facing the world today.

Thank you for your cooperation !