1

**PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT**

PRECISIONSM RESOLUTION TARGETS

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Vous filez votre référence

Vous filez votre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Nonsampling Survey Error: Interviewer Induced Error and Question Related Error

**by
Dinah M. Gumprich 159751**

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

Master of Arts

Department of Political Science

**Carleton University
Ottawa, Ontario**

**© Dinah M. Gumprich
1995**

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file / Votre référence

Your file / Votre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

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

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

ISBN 0-612-08944-4

Canada

Name Nirah M Gumerich

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Political Science - General

SUBJECT TERM

0615

U-M-I

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture	0729
Art History	0377
Cinema	0900
Dance	0378
Fine Arts	0357
Information Science	0723
Journalism	0391
Library Science	0399
Mass Communications	0708
Music	0413
Speech Communication	0459
Theater	0465

EDUCATION

General	0515
Administration	0514
Adult and Continuing	0516
Agricultural	0517
Art	0273
Bilingual and Multicultural	0282
Business	0688
Community College	0275
Curriculum and Instruction	0727
Early Childhood	0518
Elementary	0524
Finance	0277
Guidance and Counseling	0519
Health	0680
Higher	0745
History of	0520
Home Economics	0278
Industrial	0521
Language and Literature	0279
Mathematics	0280
Music	0522
Philosophy of	0998
Physical	0523

Psychology	0525
Reading	0535
Religious	0527
Sciences	0714
Secondary	0533
Social Sciences	0534
Sociology of	0340
Special	0529
Teacher Training	0530
Technology	0710
Tests and Measurements	0288
Vocational	0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language	
General	0679
Ancient	0289
Linguistics	0290
Modern	0291
Literature	
General	0401
Classical	0294
Comparative	0295
Medieval	0297
Modern	0298
African	0316
American	0591
Asian	0305
Canadian (English)	0352
Canadian (French)	0355
English	0593
Germanic	0311
Latin American	0312
Middle Eastern	0315
Romance	0313
Slavic and East European	0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy	0422
Religion	
General	0318
Biblical Studies	0321
Clergy	0319
History of	0320
Philosophy of	0322
Theology	0469

SOCIAL SCIENCES

American Studies	0323
Anthropology	
Archaeology	0324
Cultural	0326
Physical	0327
Business Administration	
General	0310
Accounting	0272
Banking	0770
Management	0454
Marketing	0338
Canadian Studies	0385
Economics	
General	0501
Agricultural	0503
Commerce-Business	0505
Finance	0508
History	0509
Labor	0510
Theory	0511
Folklore	0358
Geography	0366
Gerontology	0351
History	
General	0578

Ancient	0579
Medieval	0581
Modern	0582
Jack	0328
African	0331
Asia, Australia and Oceania	0332
Canadian	0334
European	0335
Latin American	0336
Middle Eastern	0333
United States	0337
History of Science	0585
Law	0398
Political Science	
General	0615
International Law and Relations	0616
Public Administration	0617
Recreation	0814
Social Work	0452
Sociology	
General	0626
Criminology and Penology	0627
Demography	0938
Ethnic and Social Studies	0631
Individual and Family Studies	0628
Industrial and Labor Relations	0629
Public and Social Welfare	0630
Social Structure and Development	0700
Theory and Methods	0344
Transportation	0709
Urban and Regional Planning	0999
Women's Studies	0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture	
General	0473
Agronomy	0285
Animal Culture and Nutrition	0475
Animal Pathology	0476
Food Science and Technology	0359
Forestry and Wildlife	0478
Plant Culture	0479
Plant Pathology	0480
Plant Physiology	0817
Range Management	0777
Wood Technology	0746
Biology	
General	0306
Anatomy	0287
Biostatistics	0304
Botany	0309
Cell	0379
Ecology	0329
Entomology	0353
Genetics	0349
Limnology	0793
Microbiology	0410
Molecular	0307
Neuroscience	0317
Oceanography	0416
Physiology	0433
Radiation	0821
Veterinary Science	0778
Zoology	0472
Biophysics	
General	0786
Medical	0760

EARTH SCIENCES

Biogeochemistry	0425
Geochemistry	0996

Geodesy	0370
Geology	0372
Geophysics	0373
Hydrology	0388
Mineralogy	0411
Paleobotany	0345
Paleoecology	0426
Paleontology	0418
Paleozoology	0985
Palynology	0427
Physical Geography	0368
Physical Oceanography	0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences	0768
Health Sciences	
General	0566
Audiology	0300
Chemotherapy	0992
Dentistry	0567
Education	0350
Hospital Management	0769
Human Development	0758
Immunology	0982
Medicine and Surgery	0564
Mental Health	0347
Nursing	0569
Nutrition	0570
Obstetrics and Gynecology	0380
Occupational Health and Therapy	0354
Ophthalmology	0381
Pathology	0571
Pharmacology	0419
Pharmacy	0572
Physical Therapy	0382
Public Health	0573
Radiology	0574
Recreation	0575

Speech Pathology	0460
Toxicology	0383
Home Economics	0386

PHYSICAL SCIENCES

Pure Sciences	
Chemistry	
General	0485
Agricultural	0749
Analytical	0486
Biochemistry	0487
Inorganic	0488
Nuclear	0738
Organic	0490
Pharmaceutical	0491
Physical	0494
Polymer	0495
Radiation	0754
Mathematics	0405
Physics	
General	0605
Acoustics	0986
Astronomy and Astrophysics	0606
Atmospheric Science	0608
Atomic	0748
Electronics and Electricity	0607
Elementary Particles and High Energy	0798
Fluid and Plasma	0759
Molecular	0609
Nuclear	0610
Optics	0752
Radiation	0756
Solid State	0611
Statistics	0463
Applied Sciences	
Applied Mechanics	0346
Computer Science	0984

Engineering	
General	0537
Aerospace	0538
Agricultural	0539
Automotive	0540
Biomedical	0541
Chemical	0542
Civil	0543
Electronics and Electrical	0544
Heat and Thermodynamics	0548
Hydraulic	0545
Industrial	0546
Marine	0547
Materials Science	0794
Mechanical	0548
Metallurgy	0743
Mining	0551
Nuclear	0552
Packaging	0549
Petroleum	0765
Sanitary and Municipal	0554
System Science	0790
Geotechnology	0428
Operations Research	0796
Plastics Technology	0795
Textile Technology	0994

PSYCHOLOGY

General	0621
Behavioral	0384
Clinical	0622
Developmental	0620
Experimental	0623
Industrial	0624
Personality	0625
Physiological	0989
Psychobiology	0349
Psychometrics	0632
Social	0451

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

**NONSAMPLING SURVEY ERROR: INTERVIEWER INDUCED ERROR
AND QUESTION RELATED ERROR**

submitted by

Dinah Gumprich, B.A. Hons.

in partial fulfilment of the requirements
for the degree of Master of Arts

Chair, Department of Political Science

Thesis Supervisor

Carleton University
November 2, 1995

Abstract

Nonsampling error will be examined theoretically and empirically in this M.A. thesis. Chapter 1 focuses specifically on interviewer influence within nonsampling error. This chapter discusses the respondent's role in interviewer error, interviewer training and questioning methods, vocalization effects, and the demographic characteristics of interviewers. Chapter 2 analyzes the influence of questions in nonsampling error. This chapter focuses on question design, context, wording, scales, response options, order, and length. Chapter three consists of an empirical examination of data amalgamated by David A. Northrup and Anne E. Oram. This chapter puts the theoretical information of the previous two chapters to work in searching for evidence of nonsampling error. Question wording, order, and the influence of interviewer gender are examined specifically in this chapter. The recognition of nonsampling error is vital to the confirmation and validation of survey analysis, and ultimately, to the correction of nonsampling error in future surveys.

Table of Contents

Abstract ii

Introduction 1

Sampling vs. Nonsampling Error 2

Reliability and Validity 6

Chapter 1

Interviewer Influence and Survey Error 18

Respondent Behaviour and Motivation 19

Interviewer Training, Questions, and Questioning Methods 22

Vocalization Effects 26

Interviewer Characteristics: Age, Race, and Gender 27

Interviewer Age 27

Interviewer Race 28

Interviewer Gender 34

Conclusion 36

Chapter 2

Question Related Influence and Survey Error 39

Questionnaire Design, Type and Effects 40

Question Comprehension 44

Question Context 46

Question Wording: Abstraction and Frequency 50

Scales — Measurement and Options 61

Middle Responses 63

Don't know - No Opinion 65

Closed ended and Open-ended Questions 66

Question Order and Response Category Order 67

Question Length 72

Conclusion 73

Chapter 3

Statistical Analysis of Nonsampling Error 75

The Hypothesis 77

The Variables 78

The Dependent Variables 79

Statistical Analysis — Frequency Distribution Analysis 83

Statistical Analysis — Crosstabs 87

Statistical Analysis — Multiple Regression 92

Statistical Analysis — Logistic Regression 113

Statistical Analysis — Discriminant Analysis 121

Conclusion 132

Conclusion 136

Appendix: Frequency Distributions for the Dependent Variables 143

Bibliography 155

List of Tables

**Crosstabs Results: Measures of Association and Significance for Dependent Variables
and Order/Wording Variables 89**

Multiple Regression: Correlation Matrix Results 101

Multiple Regression: Standardized Beta Weights 103

Logistic Regression: Logistic Regression Coefficients 115

Discriminant Analysis: Results 122

**Discriminant Analysis: Significance and Standardized Canonical Discriminant
Function Coefficients 125**

Introduction

While not often discussed in the reporting of survey results, survey error comprises a significant and important part of the understanding of statistical measures. In fact, an examination of survey error is a vital part of all quantitative social science research. It is instrumental in placing statistics in perspective in the larger field of research, and crucial in preventing distortion in the application of survey results. Survey error encompasses a lively and complex debate in the research community. Yet despite this, a discussion of survey error is often purposefully avoided in the public arena when statistics are presented. Since error can make "concrete" statistical results appear somewhat "soft", statisticians and professionals may avoid looking at error so as to construct a facade of strength and certainty in their numbers. If error is discussed in this sphere, it is often only in the context of random or unavoidable error, or problems with the sampling distribution. The problems of sampling distributions could technically be avoided, and are fairly clear to understand and discuss in terms of future potential solutions.

In contrast, nonsampling error can be vague both in defining problems and finding solutions. This does not diminish its importance. In fact, it should elevate the importance of understanding the underlying issues of nonsampling error. This introductory discussion will include defining the issues of nonsampling error, validity, reliability and variance. It will also focus specifically on the impact of the interviewer in several aspects of quantitative research. I have conducted statistical analysis for several years under the direction of Professor Scott Bennett. I have often been dismayed by the lack of discussion of error of any sort in many statistical research examinations. While some surveys do mention survey error briefly, they often elude the issue as it pertains to their particular analysis. If survey analysts do mention error, their discussion remains hinged on sampling difficulties and more definable types of error. With many critics cynical of purportedly

manipulated survey results, especially around election periods. I have grown curious as to the influence of nonsampling error. Specifically, the areas of interviewer influence and question order and wording contain the potential to change entirely the interpretation of survey results, and how they should be applied to issues within the government and public policy. Upon further investigation, I discovered that much theoretical research has been done on nonsampling survey error. Unfortunately, little empirical research has been conducted, especially on a wide scale, involving many different aspects of the issue.

I decided to write my M.A. Thesis in Political Science on nonsampling error because of the importance of polling to Canadian politics, as well as to the growing market research industry, and public policy application of surveys results. I feel that in order for these areas to use surveys correctly, the existence of nonsampling error must be documented and analyzed, so that this type of error can be accounted for, if not solved, in the application of survey results. As well, I feel it important to include both a theoretical and empirical study in this examination of nonsampling error. Theory provides a clear understanding of the issues involved in nonsampling error, and empirical examination will support the continuation of further statistical tests of large scale data on this issue. It will hopefully demonstrate concrete evidence of this error, so that pollsters will begin to realize the importance of recognition of nonsampling error. This introduction will begin to clarify the issues involved in nonsampling error. Later chapters will focus specifically on interviewer influence and question related error, respectively. Without this foundation of an understanding of error, statistical results will otherwise be left in the shadow of uncertainty.

Sampling vs. Nonsampling Error

Statistical error can be divided up into sampling error and nonsampling error, and

each of these categories contains a random and non-random component. Random sampling error occurs when the researcher makes a conscious decision to study a particular portion of the population, rather than whole population.¹ It is not a mistake per se, but, instead, a part of the design of the research project undertaken intentionally by the researcher. The researcher fully understands the consequences and impact of such a choice. Random sampling error is accepted since researchers not only understand it well, but find it enables them to control costs. The remedy for this error clearly is not to study solely a specific subset of the population, but instead a randomly selected group representative of the general population. For example, a researcher may choose to study the incidence and type of drug addiction that occurs in the general population. Instead, the researcher only interviews drug addicts recovering from addiction, and having entered a treatment facility within the last three months. In this case, sampling error could occur. If the researcher instead chooses to select randomly respondents from the general population, then sampling error would not necessarily be as likely to occur. Sampling error is well founded in the statistical research community. An understanding of sampling error, its consequences and remedies, is therefore fairly straightforward, and is undertaken intentionally with full comprehension on the part of the researcher.

Nonsampling error, in contrast, occurs because of mistakes or unintentional errors on the part of the researcher, the survey team, or the respondents. Judith Lessler argues that nonsampling errors arise in the development and execution of survey methods. They may be the result of definitions that are misunderstood, problems in the statistical computations, or significant nonresponse. She further asserts that researchers choose to use a particular survey design with a complete understanding of the existence of errors in

1. Judith T. Lessler, and William D. Kalsheek, *Nonsampling Error in Surveys*, (New York: John Wiley and Sons, 1992), 5.

it.² Still, in nonsampling error difficulties, researchers may not be fully aware of the extent of these errors, or the degree to which they will impact on the survey results. While researchers may acknowledge the presence of nonsampling errors, they may not seek to determine the strength of the error. Instead they argue that since no absolute standard of truth exists in social science attitude research, the complete understanding of error can never be completely uncovered, and should therefore just be accepted as it exists in the design.

When quantitative social science research is done with error in mind, it is often *only* to discover the extent of such error, and not for other attitudinal information it may uncover. Researchers looking at nonsampling error seek better ways of writing survey questions, reducing nonresponse, and improving training and consistency of interviewers and respondents.³ Such researchers point out that a lack of qualification in staff training, or the desire to maximize cost efficiency will result in a poorer quality of data, and hence increase nonsampling error in survey results.⁴ Or, researchers may attempt to increase the sample size without a resulting increase in budget, so an overworked staff will inevitably make increased mistakes. In addition, researchers may fail to understand the intentions of the sponsor and of the survey design, and will fail to meet the objectives of what the survey originally intended to measure. Respondents may suffer fatigue effects from overly long questionnaires, or may feel compelled to answer questions for which they have no interest or feelings of responsibility only to avoid admitting ignorance.⁵ To some degree, each of these problems will cause nonsampling error.

2. Lessler and Kalsbeek, *Nonsampling Error in Surveys*, 5.

3. Robert Groves, *Survey Errors and Survey Costs*, (New York: John Wiley & Sons, 1989), 3.

4. Lessler and Kalsbeek, *Nonsampling Error in Surveys*, 17-23.

5. *Ibid.*

Researchers seek to learn how to control these errors in experiments with the survey design and other factors in data collection. They find that they can control the survey; at least insofar as presenting each respondent with an identical or controlled instrument. Other effects are more difficult for these researchers to control uniformly. Interviewer actions, for example, despite training and instructions, can vary widely. Further still, researchers have almost no control over the mood of the respondents and their desire to cooperate, the respondents' reactions to interviewer and questionnaire characteristics, or to political and time effects that occur over the span of a research project.⁶ Avoidance of such errors in almost all surveys is virtually impossible, even when researchers purposefully seek to control for these effects. Nonsampling errors will affect the biases and variances of the survey estimators. Barbara Bailer argues that nonsampling errors present much more of a serious issue than sampling errors. She further asserts that "in large-scale surveys, it is not uncommon that the contribution of interviewer error to total variance is larger than the contribution of sampling error."⁷ Nonsampling error is therefore a very important issue, even if only a certain segment of researchers attempt to discover its importance in the greater scheme of survey research.

As mentioned previously, error occurs because the search for truth is controversial in survey research. Conceptual and empirical definitions of social science concepts are fraught with disagreement. The understanding of these concepts is also further hindered by systematic and random errors in any type of quantitative observation.⁸ Concepts must

6. Lessler and Kalsbeek, *Nonsampling Error in Surveys*, 277-8.

7. Barbara Bailer and Paul P. Biemer, "Some Methods for Evaluating Nonsampling Error in Household Censuses and Surveys," in *W.G. Cochran's Impact on Statistics*, eds. Poduri S.R.S. Rao and Joseph Sedransk, (New York: John Wiley & Sons, 1984), 254.

8. Richard P. Bagozzi and Claes Fornell, "Theoretical Concepts, Measurements, and Meaning," in *A Second Generation of Multivariate Analysis, Volume 2: Measurement and Evaluation*, ed. Claes Fornell, (New York: Praeger Publishers, 1982), 25.

therefore be put in context with spurious meaning that originates from biases in survey procedures. Bagozzi and Fornell advocate that spurious meaning occurs as a result of nonrandom errors in measurement of the focal concepts of the research, and is systematically included throughout the research. This has the potential to affect systematically variation in the survey results, and therefore influence detrimentally the relations between concepts or variables in the research.⁹ In bivariate relationships, measurement errors can give the illusion of a stronger or weaker link than otherwise exists. Error in multivariate relationships can present even more complexities, where it can lead to inferring relationships that are incorrect in both direction and magnitude.¹⁰ In attitudinal research, designed to gauge opinions and predispositions, problems associated with error and its affect on variable relationships may be much more severe.¹¹ Error can be quite problematic for survey research. It can distort the relationships between variables, and possibly lead to a series of misunderstandings about the conclusions of survey results.

Reliability and Validity

At this point, it may be useful to examine how concepts of reliability and validity play a role in the minimization of such error, especially in measurement error. Reliability remains important because of the controversy over the search for truth in attitude research. Question constructs may only touch on the true understanding of latent variables.¹² These variables may be difficult to measure using normal questioning methods, and it may be

9. Richard P. Bagozzi and Claes Fornell, "Theoretical Concepts, Measurements, and Meaning," 31.

10. Frank M. Andrews, "Construct Validity and Error Components on Survey Measures: A Structural Modeling Approach," *Public Opinion Quarterly* 48 (1984), 409-410.

11. Herbert B. Asher, "Some Consequences of Measurement Error in Survey Data," *American Journal of Political Science* 18 (1974), 484-5.

12. Duane F. Alwin and Jon A. Krosnick, "The Reliability of Survey Attitude Measurement," *Sociological Methods and Research* 20 (1991), 140.

even more problematic to keep such questions consistently measuring the same latent variable across surveys. Reliability attempts to retain linkage between observations and the theoretical constructs behind the "true" variable.¹³ It also seeks to determine if the survey and its procedures will yield consistent and measurable results.¹⁴ These definitions also tie reliability with validity. Yet difficulties still exist in measures testing reliability. If reliability in a survey is low, the statistical test may indicate not low reliability, but instead only that there exists no significant difference because of an increase in variance caused by error, instead of good variance caused by the survey construct.¹⁵ It may not always therefore be possible to infer that a survey question is reliable in measuring a particular concept when using a statistical test.

Some researchers attempt to improve the chances of having a reliable indicator by judging a concept using multiple indicators. For example, instead of just one question measuring a particular attitude, a series of questions, eventually to be compiled into an index measuring the same attitude, is presented in the survey instrument. If this is used, the researcher must pay careful attention to assure that all the items address the same concept and measure it in comparable methods, even if different response scales are used.¹⁶ This may be referred to by some researchers as *parallel measures*. The researcher attempts to uncover some underlying characteristics of an attitude in several questions, each containing the same degree of precision.¹⁷ This method may be quite successful. Even if problems arise whereby some of the measures cannot be combined, the researcher

13. Alwin and Krosnick, "The Reliability of Survey Attitude Measurement," 141.

14. Marlene E. Henerson, Lynn Lyons Morris and Carol Taylor Fitz-Gibbon, *How to Measure Attitudes*, (New York: Sage Publications, 1978), 132.

15. *Ibid.*, 154.

16. Andrews, "Construct Validity and Error Components," 415.

17. Groves, *Survey Errors and Survey Costs*, 329.

has several measures to fall back on that could be used to judge a particular attitude.

Nonetheless, using composite scores can unintentionally worsen reliability if they are used. The lack of reliability of one of the component items can compound in combination with other items,¹⁸ thereby threatening the reliable measurement of the intended attitude. The reliability of individual items must therefore be clearly understood and examined prior to their inclusion in a composite index. Such an index will correlate the reliability of survey responses, and create an upper limit of the magnitude of correlations or relationships with other variables in the study.¹⁹ Conventional views indicate that reliability must be present to infer validity in the research.

Validity explores whether the survey instrument is an appropriate measure for the concepts which the researcher intends to explore.²⁰ Validity reflects *construct validity*, which compares the effectiveness of the measure as observations are related to the researcher's theoretical construct.²¹ In testing effects, validity is not as crucial as reliability. Nonetheless, while reliability may be more important for these studies, validity should not be dismissed altogether. Validity is generally understood not to be inferred from reliability alone. It contains its own unique requirements.

Still, what is important in these studies is not what the survey is attempting to measure, but whether or not the survey measures attitudes consistently, and how certain forces impact on that measurement. This is essential in the understanding of nonsampling error, because if an instrument is not reliable, it would be virtually impossible to distinguish nonsampling error from other methodological flaws in the survey. Validity is

18. Alwin and Krosnick, "The Reliability of Survey Attitude Measurement," 143.

19. *Ibid.*

20. Henerson, Morris and Fitz-Gibbon, *How to Measure Attitudes*, 132.

21. Andrews, "Construct Validity and Error Components," 411.

not necessarily as important in this research, since what the survey measures is secondary to its effects. In that case, reliability and validity can be thought of as two separate concepts. Conventional views generally do not separate the concepts.²² This makes having a reliable but not valid measure problematic in the conventional understanding of reliability and validity. In this research, it would appear more important that researchers measure a concept well, but less important that researchers measure what they thought they were measuring.

How these results vary is a function of the variance of the survey. Some variance is normal in a survey, and explained by the natural distribution of responses. In fact, researchers seek to have varied responses, because otherwise the survey would not distinguish anything among respondents. Answers normally vary because of differences in how people view occurrences, objects, and themselves. For example, they may have differing perception of their health,²³ though in reality, their overall health may vary little. A retired athlete may feel that she is losing muscle tone, while a new devotee to aerobics may feel that she is building muscle. Compared to each other however, they may be of equal health and muscle tone. How respondents report their socio-psychological or physical traits can be problematic. It may not therefore be best to accept a people's statements of their traits or attitudes as the best indicator of those traits or attitudes.²⁴ Responses to questions can therefore be a matter of perspective. Attitudes are complex, and in order to minimize problems, researchers may be better off to classify respondents themselves, rather than to have the respondent place themselves in a series of answers.

22. Scott Bennett, *The Quantitative Approach in Political Science: An Introduction*, (Queenston: The Edwin Mellen Press, 1986), 130.

23. Andrews, "Construct Validity and Error Components," 413.

24. Henerson, Morris and Fitz-Gibbon, *How to Measure Attitudes*, 12.

Responses may also differ because of the different ways that people understand scales and response categories. This measurement error may be because of methods variance. Furthermore, in a series of questions with this problem, providing that the respondent answers consistently, the method effects will compound, generating covariation between the variables. This can either strengthen or weaken correlation between concepts in the survey. This problem is a serious one because as long as the potential for differing interpretations of survey items exists, it can occur in virtually all survey questions.²⁵ The differences in respondent interpretations is only one part of the larger picture of nonsampling survey error. It is often discussed in relation to the formation and execution of survey questions; how their wording and context impact on respondent perceptions. Socio-demographic factors, such as education, are often deemed to be key predictors of respondent perceptions and reactions to survey research.

Respondents may be reacting to different stimuli when they give survey responses. Andrews presents three bases by which the respondent may be influenced to answer a question and hence create measurement error, which do not include socio-demographic factors. First, respondents may have different underlying feelings about the concepts being measured in the survey. Second, the respondents may be reacting to the scales being used for data collection in order to measure their attitudes in the survey instrument. Third, other problems, such as memory lapses, misinterpretation or misunderstandings by the respondents or the interviewers can account for a major source of correlated error.²⁶ Measurement error can therefore potentially occur at many different stages or during the survey process. All of these sources of measurement error assume that respondents are trying their best to present the most clear representation of their (firm) opinions to the

25. Andrews, "Construct Validity and Error Components," 413.

26. *Ibid.*, 418-19.

interviewer. This is a fairly strict assumption in survey research.

Further problems occur when respondents do not have a firm or clear opinion in reaction to the attitudes that the survey is attempting to measure. Alwin argues that if respondents do not have a firm opinion or knowledge about the issues at hand, they may in fact behave randomly in answering survey questions. Alwin presents evidence from Rugg and Cantril (1944) that "response errors were often less likely if persons 'had standards of judgement resulting from stable frames of reference,' as opposed to situations 'where people lack reliable standards of judgement and consistent frames of reference.'"²⁷ In addition, people may feel obligated to present an opinion to a researcher, even if they do not have one. Alwin states that people believe opinionated individuals to be more intelligent and respected, and so therefore the respondent ought to offer some sort of attitude. Respondents may be reacting to a desire to adhere to what they perceive to be interviewer expectations, as well as to project a positive self-image. When this occurs, respondents are again offering essentially random "attitudes."²⁸ This will increase the random variation in particular variables. Because of these problems, critics argue that measurement error is not the fault of the respondent, but of imperfect instruments, and should not be blamed on respondents faced with social barriers. In Andrew's study,

the survey design characteristics account for 66 percent to 72 percent of the variance in the dependent variables. This is an important finding, for it shows that much of the variation in measurement quality can be explained, and hence measurement quality is subject to prediction in other surveys and perhaps to improvement.²⁹

Nonetheless, respondent ambiguity remains a problem in attitude measurement. It can be

27. Alwin and Krosnick, "The Reliability of Survey Attitude Measurement," 144.

28. *Ibid.*

29. Andrews, "Construct Validity and Error Components," 428.

both the problem of the respondent and the measurement instrument. Survey responses must attempt to account for many varying facts of a complex opinion. They must resolve ambiguity in respondents' opinions, and allow the respondent to make a clear, unhesitating choice among all the alternatives. As Alwin argues, this could translate into a degree of random choice, since many response categories may not directly relate to attitudinal cues.³⁰ Therefore, responses could be ambiguous, or the feelings of the respondent could be ambiguous, whereby choosing a response that encompasses all of the respondent's feelings would be a fairly hard task to accomplish.

Part of the problem with response categories is that the question or the response alternatives do not directly relate to common attitudes of respondents. They lack ecological validity. Groves asserts that researchers pressure respondents to provide answers as a part of a survey measurement that may fail to capture their interests or experiences. In addition, "by sweeping problematic answers into categories such as "don't know" or "refused," survey researchers ignore signals that negotiations about meaning have broken down."³¹ Closed ended questions become problematic if the interviewer incorrectly places the respondents answer into an inappropriate category,³² intentionally or accidentally. The respondent may therefore not be able by virtue of the survey instrument to choose the most appropriate response.³³ This further enhances the possibility of random measurement error. These issues will be further discussed in Chapter 2.

A greater degree of education may help respondents to recognize their attitudes,

30. Alwin and Krosnick, "The Reliability of Survey Attitude Measurement," 146.

31. Groves, *Survey Errors and Survey Costs*, 369.

32. Floyd J. Fowler Jr., *Survey Research Methods, Second Ed.*, (New York: Sage Publications, 1993), 108.

33. Alwin and Krosnick, "The Reliability of Survey Attitude Measurement," 147.

and therefore lessen the possibility that random measurement error may occur. Education may allow people to synthesize complex issues, and to apply their opinions to varying measurement instruments. In contrast however, further education may increase a respondent's tolerance for ambiguity, and reduce their willingness to form opinions quickly on complex issues.³⁴ Nonetheless, a more educated respondent is usually found to reduce error, and give more reliable answers, which is consistent with arguments surrounding socio-demographic factors. This discussion about a respondent's level of education contains many sociological assumptions. It does not however, invalidate the need for respondents of many different educational backgrounds, as would occur within the whole population. Alwin asserts that when less educated respondents are asked to place themselves on an ideological scale, such as in issues of party identification, the results are much more reliable than on measures that assess attitudes towards policies, social groups, or measures expressing political efficacy or alienation.³⁵ While it is important to know which measures are most vulnerable, this does not tell the whole story about the influences on both the measures and the respondents.

Interviewers may account for a portion of error that occurs in data collection. This is difficult to assess however, so the impact of this error is not often known in total. Biemer argues that variation in responses caused by interviewers is difficult to separate from variation caused by sampling or response variance. Even if one is able to distinguish such variation in the sample, this may be characteristic of the individual interviewer's sample, and not of the interviewer herself. Or, if it is caused by the interviewer, it may not be applicable to other survey samples. Interviewer assignments must therefore be controlled, and ideally, should be representative of the total survey sample.³⁶ Otherwise, the

34. Alwin and Krosnick, "The Reliability of Survey Attitude Measurement," 154.

35. *Ibid.*, 173.

researcher will continue to have difficulty separating sampling effects from interviewer influence if the interviewers are continually assigned on a nonrandom basis. Studies of this however, have not necessarily been forthcoming. In fact, classical studies done of interviewer effects have generally been based on few interviewers, as low as two to five in the studied sample.³⁷ Ideally more interviewers should be studied, and their samples should be representative of the entire survey sample, yet this is difficult to establish and maintain in a world of fiscal restraint. Larger studies have been performed since, and will be discussed in Chapter 1.

Despite the recognized problems of interviewer effects on respondent answers, common indicators of performance that can be used to predict interviewer effects are not used consistently among research studies.³⁸ It may be difficult to compare studies therefore, if interviewers are not compared using the same measures to judge the extent of their influence on respondents. One common statistic though, generally used in many examinations of such error is the interclass correlation, generally referred to as rho (ρ_{int}). This measures ranges in value from 0 to 1.0. If the result indicates a value of 0, the interviewer is not likely to be influencing responses to a particular question. The interviewers are therefore standardized in generating responses and producing results consistent with a lack of influence. If ρ_{int} produces higher values, more interviewer related error is likely occurring for that measure. Interviewers are impacting on respondent answers much more than in the previous case. Yet even low values can be problematic. For example, "[e]ven though most values are 0.02 or less, if an interviewer takes 50 interviews

36. Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, *Measurement Error in Surveys*, (New York: John Wiley & Sons, 1991), 260.

37. J.J. Feldman, Herbert Hyman and C.W. Hart, "A Field Study of Interviewer Effects on the Quality of Survey Data," *Public Opinion Quarterly* 15 (1951-52), 734-5.

38. Groves, *Survey Errors and Survey Costs*, 381.

on average, the effect of a p_{int} of 0.02 is to increase the standard errors by 40 percent over what they would be with no interviewer-related error."³⁹ While this examination of interviewer effects will not discuss other researchers' specific statistical results, realizing how small errors can compound and result in large problems may result in a better understanding of the magnitude of this problem.

Attempting to produce studies to examine this problem can also be problematic. Reinterviewing respondents may produce unintended effects on the part of both the respondent and the interviewer. If an identical instrument is used a second time, two results may occur. Groves indicates that the respondent may give the identical result with the intention of wishing to appear consistent, or the respondent may purposefully give a different response in order to appear more open-minded on attitudinal measures. On the side of the interviewer, knowledge of the respondent's previous answer may induce the correlation of error across the two surveys.⁴⁰ Interviewing may therefore be seen as a destructive process, where previous results may affect new survey results if the same respondents, interviewers and measurement instruments are used.

Using only one survey across different respondents to gauge interviewer effects may also be problematic. The researcher is forced to look at other ways by which to obtain the same expected values for more than one interviewer. This however, ties back in to the problem of the representativeness of the sample, and the effects may be deemed to be only applicable to that particular sample of a particular interviewer. Results may not then be inferred about the general population, or about a general class of interviewers.⁴¹ Critics

39. Groves, *Survey Errors and Survey Costs*, 262.

40. *Ibid.*, 325.

41. Robert H. Hanson and Eli S. Marks, "Influence of the Interviewer on the Accuracy of Survey Results," *Journal of the American Statistical Association* 53 (1958), 636.

may then be able to charge that a negative result does not indicate a relationship between the instrument and interviewer effects, but instead may just be the result of a large sampling variance. Therefore results from these studies are not without difficulties internally, and have continued problems externally in application to the future decline of interviewer effects.

These results may therefore merely demonstrate that both the estimates of interviewer impact, and the measurement instrument itself, are unreliable. It is possible, according to Feldman, that based on some of these facts and an erroneous understanding of what they indicate, researchers could make flawed judgements as to which interviewers are preferable. If this occurs, researchers could end up with consistently bad staff members, which would negate the reliability and validity of their results.⁴² Therefore, it is a difficult and complicated task to arrive at a correct estimation and interpretation of interviewer effects. Furthermore, this is not considered among most attitudinal researchers to be of a high priority, and "few are willing to jeopardize the quality of a survey by altering procedures in order to measure what may be relatively small effects."⁴³ The cost may be prohibitively high, and the rewards not immediately forthcoming. Interviewer induced error therefore continues to exist despite the efforts of some researchers to minimize its impact on survey results.

Nonsampling error in surveys can be very problematic for standard error and survey estimates. Unfortunately, it is not frequently studied in general, or studied specifically by those companies who possess the budgets to do so with better methodology and measurement control. Many statistical firms choose to ignore the potential

42. Feldman, Hyman and Hart, "A Field Study of Interviewer Effects," 756.

43. Clyde Tucker, "Interviewer Effects in Telephone Surveys," *Public Opinion Quarterly* 47 (1983), 85.

consequences of nonsampling error, and instead to take their survey results at face value. The study of error, or particular attention to error in general surveys must be made to appear not as a flaw or a diminishment of the importance of the research findings, but instead as an enhancement and confirmation of the statistical results. After all, if a researcher can confirm that her results are not caused by any type of nonsampling error, the results would appear even stronger than if the shadow of error tainted the full acceptance of the results. Error plays a vital role in the variance, reliability and validity of statistical results. In this information age, it is better understood than avoided, because the potential for harm grows as each research study is produced. The sources of this error, and the issues involved in its study should now be further examined.

Chapter 1

Interviewer Influence and Survey Error

Interviewers are a serious source of potential Nonsampling error in survey research. Interviewers can affect respondents' reporting of behaviour and attitudes not only by their overt actions, but merely by their presence. The physical traits of the interviewer, or the perceived socio-demographic attributes of the interviewer could cause the respondent to react based on the desire for social respectability and acceptance. Or, variation in interviewer behaviour, such as how each interviewer reads survey questions, could impact on the variance between surveys. Any of these problems could result in serious flaws in the survey research, and could destroy the general applicability and usefulness of the survey results. Interviewers must therefore aim to be consistent in performing their interviews, as well as in attempting to construct representative samples of respondents. Yet when these problems occur, and are not recognized or reported, the survey results could be misunderstood and misapplied in many important areas, such as in marketing research, policy approval, or government action.

Interviewer error is difficult to recognize, problematic to analyze consistently, and impossible to solve uniformly. This should not deter further analysis. Interviewer error and respondent reactions should be closely examined so that the possibility exists that if they are well understood, some problem solving can begin. This chapter will discuss the respondent role in interviewer error, interviewer training and questioning methods, vocalization effects, and the demographic characteristics of interviewers. The study of interviewer effects is a vital part of the examination of nonsampling error, and should continue to be explored as long as interviewers, and surveys, continue to play an important role in the information age.

Respondent Behaviour and Motivation

Researchers first begin by defining the problems of interviewer influence. Groves states that interviewer effects can occur in four different ways. First, respondents may react to the socioeconomic and demographic characteristics of the interviewer. Second, the interview may intentionally or unintentionally alter the survey by failing to ask some of the questions, or by failing to enumerate some of the responses. Interviewers may also incorrectly record respondent answers. Third, interviewers can change the questions by emphasizing different words or by stressing different concepts in the survey. Fourth, the interviewer may respond in different ways to respondent requests for help, and thereby influence the respondent's answer. Groves further argues that these errors are variable, and are not easily replicated or held constant in magnitude across respondents and interviewers.¹ Varying effects are not only difficult to judge, but practically impossible to replicate for additional studies. The total impact of the interviewer can therefore be difficult to estimate for each individual study. Nonetheless, a better understanding of how these errors occur will ultimately yield better research studies.

The result of interviewer effects is to increase the standard error in the survey estimates.² Interviewers are increasingly associated not only with their role in rate of response, but also in how they influence respondent answers. Some advocate that interviewer effects can be minimized with better training and increased supervision. This should theoretically increase reliability and consistency in measurement.³ This only tells part of the story. Different motivation for the respondents to give answers to the interviewers are at work that continually modify the information the researchers receive.

1. Groves, *Survey Errors and Survey Costs*, 359.

2. Fowler, *Survey Research Methods*, 121.

3. *Ibid.*

Fowler argues that two different motivations compete. The first dictates that respondents find satisfaction in giving interviewers their own opinions, and in presenting complete, concise and accurate information. On the other hand, a society increasingly sensitive to the approval of others may make the respondent more likely to attempt to satisfy the researcher with what appears to be appropriate answers for the research topic at hand. Fowler calls this the information-giving game and the ingratiation game. He argues that a concentration of questions on a particular topic, as well as limited responses, give the respondent cues as to what the interviewer is attempting to uncover. This not only gives the respondent an idea of the purpose of the interview, but also of what should be a "normal" or "acceptable" response.⁴ This relates to the idea of the desire of the respondent to appear socially respectable. An open-ended question would not necessarily give the respondent the same direction of thinking. Respondent motivation is therefore just as important as the interviewer's intentions or effects on the survey instrument.

Reinecke adds three reasons that explain why people respond in certain ways during survey interviews. The first two explanations were originally given by Fishbein. He states that a person follows an internal rule of what constitutes norms, or rules of behaviour. In addition, Fishbein originally argues that people follow ideals of what they think is appropriate according to society. Finally, Reinecke adds that individuals act as they feel is correct and appropriate behaviour according to the situation.⁵ All of these reasons are important. Each highlights the internal motivation of respondents, and how they are related to both individual and societal norms. In fact, the desire of individuals to

4. Kurt W. Back and Kenneth J. Gergen, "Idea Orientation and Ingratiation in the Interview: A Dynamic Model of Response Bias," *American Statistical Association: Proceedings of the Social Statistics Section* (1963), 284.

5. Jost Reinecke and Peter Schmidt, "Explaining Interviewer Effects and Respondent Behavior: Theoretical Models and Empirical Analysis," *Quality and Quantity* 27 (1993), 224.

appear socially respectable is an important factor for how people respond to survey questions. This is often termed 'social desirability'. In Andrews' study,

Respondents who scored relatively high on this concern had a modest tendency to give data that were below average in validity and above average in residual error.⁶

Social desirability can therefore have an impact on the survey results. If this result is compounded in a large scale survey, it could have the potential to impact significantly on the overall outcome of the statistical results.

The social desirability factor may be influenced by several other factors. The respondent may be influenced by the characteristics of the interviewer. The perceived status of the interviewer can make respondents who possess a tendency to want to appear social desirable to adjust their behaviour accordingly in the survey interview.⁷ Or, a personal style of interviewing, for example, where interviewers pose leading questions, may result in a lessened validity for the study.⁸ The artificial situation of the survey interview, respondents' perceptions of the questions, and the interviewers themselves can influence how respondents' answer.

People may become anxious in the formal sphere of personal survey interviews. Their orientation toward social desirability and acceptable public behaviour kicks in, and they react subconsciously with these factors in mind. A questionnaire not administered by an interviewer may lessen these effects. A questionnaire cannot "smile, frown, or raise an eyebrow, which an interviewer in some cases will do, in spite of honest efforts to avoid influencing the respondent."⁹ Interviewer effects that occur, willfully or by accident,

6. Andrews, "Construct Validity and Error Components," 434.

7. Reinecke and Schmidt, "Explaining Interviewer Effects and Respondent Behavior," 229.

8. Groves, *Survey Errors and Survey Costs*, 427.

should be examined carefully, because interviewers are still an important part of the information generation process.

Interviewer Training, Questions, and Questioning Methods

Questions and responses may also be influenced by interviewers. Their opinions on the questions, as well as on the research in general, can influence how respondents in turn perceive the research, and therefore how they will react to the survey questions.¹⁰ Originally, some researchers believed that it was the nature of the question that influenced the respondent.¹¹ The vital question for these researchers was whether or not a survey item was factual or nonfactual. Census data, for example, was thought to be fairly insensitive to the effects of interviewers. Not only are the questions absolute, but they can be verified. Hanson argues that while this may be true in some cases, in other instances, "facts" are largely a matter of opinion, such as in the case where a respondent is asked if she was looking for work last week.¹² Tucker asserts that researchers now realize that it is not the nature of the question that affects respondents, but the respondents' reaction to sensitive topics that are introduced with leading openings.

As well, open-ended questions and subjective evaluations of the respondents by interviewers are suspected of increasing error.¹³ In Andrews' study, respondents who enlist the aid of interviewers for clarification or repetition of questions give responses that are lower in validity and higher in residual error.¹⁴ This may relate not only to respondent

9. Henerson, Morris and Fitz-Gibbon, *How to Measure Attitudes*, 27.

10. Feldman and Hart, "A Field Study of Interviewer Effects," 742.

11. Clyde Tucker, "Interviewer Effects in Telephone Surveys," *Public Opinion Quarterly* 47 (1983), 85.

12. Hanson and Marks, "Influence of the Interviewer," 639.

13. Tucker, "Interviewer Effects in Telephone Surveys," 85.

14. Andrews, "Construct Validity and Error Components," 434.

problems in understanding fully the survey questions, but in fact reflect the style of interviewing the interviewer uses.

If a respondent is unsure about her answer, an interviewer using a personal style of questioning will magnify the possibility of biased responding. In fact, Groves reports that when interviewers are trained to use a personal style of questioning, they are more likely to pose more of the survey questions in a suggestive manner.¹⁵ This will further distort the data. Yet sometimes interviewer behaviour that purposefully reacts with the respondent will increase the reliability, and decrease errors in survey methods. For example, Fowler discusses a situation where interviewers are instructed to praise respondents who answer slowly, and respondent reporting consequently improves. Yet as Fowler indicates, these methods are difficult to check routinely. Thus interviewers are left on their own with the task of instructing the respondent on the tasks of the survey, "and not surprisingly, interviewers differ as to how they do it—in ways that affect data."¹⁶ Standardized instruction programs could help to diminish the differences among interviewers, though differences could continue to exist solely as a function of the interviewers characteristics or personality.

Interviewers are not perfect, and clearly differ in the way they follow the instructions of the survey method. Interviewers may differ in the way they apply their training to their job situations, such as in asking questions and in how they record data.¹⁷ Variability therefore exists, and this variability continues to be a cause for concern in survey estimates and error. Groves estimates that anywhere from five to sixty percent of all

15. Groves, *Survey Errors and Survey Costs*, 421.

16. Fowler, *Survey Research Methods*, 118.

17. Barbara Bailer, Leroy Bailey, and Joyce Stevens. "Measures of Interviewer Bias and Variance," in *A Second Generation of Multivariate Analysis, Volume 2: Measurement and Evaluation*, ed. Claes Fornell, (New York: Praeger Publishers, 1982), 337.

survey questions are reworded by interviewers in the process of the survey interview.¹⁸ While this may seem innocuous, it can have a great potential impact on biased responses or increased variance in the survey results.

In addition, variability may increase if questions are changed. One common mistake made in survey interviews is for the interviewer to omit a possible response category. Groves reports that in one survey experiment, where interviewers are allowed to decide for themselves if they will give a respondent the choice of answering affirmatively to a "don't know" or "neutral" category, interviewer-related error is recorded as very high, especially when the neutral category is offered. Almost thirty percent more people, in fact, chose that specific category when it was offered.¹⁹ This shows that even unintentional omissions or rewording of questions can have a great impact on survey estimates. Interviewers are certainly not neutral participants in the survey process.

More experienced interviewers are often credited with posing fewer problems in relation to interviewer effects than less experienced interviewers. In addition, closed-ended questions may yield fewer interviewer effects than open-ended questions.²⁰ In fact, Fowler argues that answer recording ought to be standardized so as to minimize the impact of variation between interviews and interviewers. Open-ended questions leave open the possibility that the interviewer will fail to record the respondent's answers verbatim, or to summarize or inadequately capture the message that the respondent gives during the process. Open-ended questions may also make interviewer probing necessary, such as where an individual is unfamiliar with the topic, or is unaccustomed to making

18. Groves, *Survey Errors and Survey Costs*, 264.

19. *Ibid.*, 265.

20. Leslie Kish, "Studies of Interviewer Variance for Attitudinal Variables," *Journal of the American Statistical Association* 57 (1962), 93.

generalizations in application to a particular problem.²¹ Or in the case of ambiguous questions, the respondent may further call upon the need for social desirability, and change her answer to suit the researcher, as previously discussed. These dilemmas in open-ended questioning can cause a great deal of error where the interviewer inadvertently influences a respondent's answer.

Open-ended questions also bring into play interviewer expectations. Interviewers are not only inconsistent in recording survey interviews verbatim, they are also "more likely to record the statements that were consistent with their expectations, and more likely to omit statements that were inconsistent with their expectations."²² Inconsistent recording is therefore not just a function of sloppy interviewing, but a subconscious effort on behalf of interviewers to present surveys consistent with their own preconceived expectation of the survey results. Interviewer expectation can have a positive side, despite this apparent source of error. Interviewers who expect to be able easily to persuade respondents to be interviewed, according to Singer, do succeed significantly in achieving a better response rate. According to Singer, this could suggest that interviewer attitudes are a more important source of influence in comparison to the survey instrument itself, at least in relation to response rates.²³ Bailar confirms this observation in a study that shows a significant variance between interviewers working in the same area.²⁴ Interviewers can therefore have both positive and negative effects; negative in the possibility that poor training or expectations influence the interviewers to omit or change responses, but positive in that they are able to influence respondents to participate, thereby increasing the

21. Kish, "Studies of Interviewer Variance," 97.

22. Groves, *Survey Errors and Survey Costs*, 266-7.

23. Eleanor Singer, Martin R. Frankel and Marc B. Glassman, "The Effect of Interviewer Characteristics and Expectations on Response," *Public Opinion Quarterly* 47 (1983) 81.

24. Bailar, "Measures of Interviewer Bias and Variance," 343.

sample size and the potential for more reliable and generalizable results.

Vocalization Effects

Interviewers can have an effect on respondents not only in visual cues, or in written observations, but also in their vocalizations. Interviewers can do more than convince respondents to participate; the mere intonation of their voice can have the potential to influence the respondents' answer. Barath and Cannell demonstrate in a fascinating study the effects of the interviewer's voice. In their study, interviewers present long lists of questions where the possible responses are a series of "yes" and "no" answers. During some interviews, the interviewers lower their voices at the end of the sentence, so that the usual voice raising at the end of a question is absent, and in other interviews, simply pose the questions the regular way.

Barath and Cannell find that in the cases where the interviewers raise their voices in the form of a normal question, the frequency of positive responses is significantly higher compared to the voice-dropping group. They further find that the voice-rising interviewer appears to encourage more valid reporting, whereas the voice-dropping group inhibits accurate reporting. Voice-rising and voice dropping produce response bias either encouraging "yea-saying" or "nay-saying", and therefore either contribute to the overreporting or underreporting of certain attitudes or behaviour.²⁵ This study is clearly important, because voice intonation, such as in asking a series of questions, often occurs without the interviewer intending to do so, and without the realization of what this does for the survey estimates. This also shows that survey construction can have an impact on responses. Had the survey been altered, whereby "yes-no" questions were alternated with

25. Arpad Barath and Charles F. Cannell, "Effect of Interviewer's Voice Intonation," *Public Opinion Quarterly* 40 (1976), 373.

other response type questions, this effect could be minimized. Nonetheless, this is an important finding for both interviewer effects, and survey construction with regard to nonsampling error.

Interviewer Characteristics: Age, Race, and Gender

Other influences of the interviewer however, are not easily changed. For example, several studies demonstrate that the characteristics of the interviewer will affect survey respondents.²⁶ Feldman argues that while this has been clearly demonstrated in experimental studies, it remains unproven in the real life situations of survey interviewing. In his opinion, the results of these studies have little practical application since they are conducted with few interviewers, and have a negligible statistical impact.²⁷ Nonetheless, Feldman stated his opinion in 1952, and clearly the passage of time has not demonstrated the validity of his convictions. Fowler certainly realizes the importance of studying the demographic characteristics of the interviewers and respondents as essential to survey research.²⁸ Three of the demographic characteristics that ought to be discussed are age, race, and the sex of the respondent. Each of these can have a particular impact on respondents, and should be carefully examined.

Interviewer Age

The age of the respondent, as well as the age of the interviewer, can have an effect on the dynamics of the survey interview. Tucker argues that older respondents are more susceptible to interviewer effects.²⁹ Perhaps they are not as aware as their younger

26. Tucker, "Interviewer Effects in Telephone Surveys," 85.

27. Feldman, Hyman and Hart, "A Field Study of Interviewer Effects," 737.

28. Fowler, *Survey Research Methods*, 111.

29. Tucker, "Interviewer Effects in Telephone Surveys," 85.

counterparts of the possibility of influence, or as on guard subconsciously against question manipulation. Older respondents may be more open to suggestion, and willing to go along with an interviewer because they are eager to enjoy the interviewer's company. Older respondents do not rate quite as agreeable as younger respondents in Feldman's study. In that same study, younger interviewers rate all age groups as less agreeable compared to the rating of the same groups by older interviewers.³⁰ Singer advocates that older interviewers are more successful at persuading and obtaining interviews. This however he explains, may be less an advantage, than extreme youth is a penalty.³¹ The age of the interviewer and the respondent can therefore have an impact on the interview. Perhaps some of this may be explained by the generational perceptions that arise during an interview. Age effects may therefore be generational effects. Or, different ages may have different attitudes about contact with individuals, and may therefore influence the possibilities of success in obtaining an interview. In any case, clearly age remains an important demographic factor to study in attempting to determine the influence of interviewers in nonsampling error.

Interviewer Race

Several studies have been done that attempt to factor into interviewer influence the race of the interviewer. One important observation in this discussion is that

[e]ven if one cannot prove that the answers under a given interview condition are more *accurate*, one can still show that the overall pattern of responses would have been *different* if the racial composition of the interviewer staff had differed, and therefore that the race of the interviewer should be taken into account in the survey design and analysis.³²

30. Feldman, Hyman and Hart, "A Field Study of Interviewer Effects," 745.

31. Singer, Frankel and Glassman, "The Effect of Interviewer Characteristics," 74.

Discussions of the race of the interviewer have become increasingly important as the composition of interviewers changes over the years. This demographic change, could mean for example that the assertion that blacks have become increasingly conservative between 1950 and 1980, could be the result of more blacks being interviewed by whites, than any actual change in ideology.³³ Several studies have shown that the responses of both black and white respondents change given the race of the interviewer. What may be important is the match of characteristics of the interviewer and the respondents. This match may be a principal source interviewer effects that contribute to nonsampling error.

Hatchett's study finds that respondents produce more liberal opinions or pro-black opinions when the interviewer is black. He also demonstrates that the acceptance of racial intermarriage varies forty-five percent depending on the race of the interviewer.³⁴ Why do these differences occur? Are the opinions of the respondents genuinely changing, or is there another factor at work? Hatchett argues that this response changing occurs because respondents are trying to avoid offending the interviewer of the "opposite" race. He states that the respondent "accommodates," "in order to get through the racial interaction with minimal tension."³⁵ This occurs for both black and white respondents and interviewers.

The polite atmosphere of the interview will minimize any desire on the part of the respondent to offend the interviewer, if she happens to be of another race. Anderson argues that blacks will be more candid with black interviewers, and will be more reluctant

32. Barbara Anderson, Brian D. Silver, and Paul R. Abramson, "The Effects of the Race of the Interviewer on Race-Related Attitudes of Black Respondents in SRC/CPS National Election Studies," *Public Opinion Quarterly* 52 (1988), 293.

33. *Ibid.*, 295.

34. Shirley Hatchett, and Howard Schuman, "White Respondents and Race-of-Interviewer Effects," *Public Opinion Quarterly* 39 (1975/6), 525.

35. *Ibid.*, 527.

to express negative attitudes toward whites. Blacks are able to discuss their feelings more frankly with black interviewers because they feel that this would not offend a black interviewer — whom they assume at least has the possibility of feeling the same way. Anderson further reports that black respondents are equally likely to tell a white interviewer that they are close to neither blacks nor white, as to state that they feel close to both races.³⁶ In addition, blacks will be more likely to profess closeness to whites if they are interviewed by a white interviewer.³⁷ Yet Anderson does not use this evidence to mean that responses given to black interviewers are more valid or truthful than those given to white interviewers.³⁸ Clearly both measures must be taken into account, in order to gain a complete understanding of both the respondent's answer and the interaction between respondent and interviewer.

Race, education and political participation may all be factors that are linked. This may be used to partially discount some findings that focus on race and interviewer effects. For example, Anderson found that blacks interviewed by blacks tend to possess lower overall levels of formal education than the blacks that are interviewed by whites.³⁹ This may have been because of the geographic areas to which interviewers are assigned. If a white interviewer is assigned to a wealthier, predominantly white neighbourhood, the respondents she encounters, in any race, will likely have a higher level of education. Anderson argues that while the effects of white and black interviewers on black respondents are significant and numerous. Because of the educational tie however, these findings may in fact be argued to be an artifact of respondent characteristics. How do other characteristics, such as political and policy attitudinal variables, respond to the race of the

36. Anderson, Silver, and Abramson, "The Effects of the Race of the Interviewer," 308.

37. *Ibid.*, 305.

38. *Ibid.*, 292.

39. *Ibid.*, 300.

interviewer?

The race of the interviewer can in fact have an impact not only on the political attitudes of the respondent, but also ultimately on the actions of the respondent. Anderson argues that blacks interviewed by blacks will exhibit more interest, concern and a sense of civic duty, than blacks interviewed by whites. What is most interesting is that Anderson asserts that after this interview, black respondents will be more likely to go out and vote if interviewed by a black interviewer rather than a white interviewer. While these may be important civic motivators, blacks may also be more likely to report falsely to black interviewers that they actually voted than to white interviewers.⁴⁰ Blacks may not want to admit to those of the same race that they failed in some sense to fulfill their civic duties. Yet with less of a sense of responsibility to another race, they may be more likely to tell the truth. This may actually balance with the times when black respondents are more likely to muffle their true opinions in order to maintain a sense of decorum with white interviewers.

In policy research, respondent answers may also be affected by the race of the interviewer. Anderson finds that black respondents are more likely to tell a black interviewer rather than a white interviewer that they feel civil rights leaders are pushing too slowly with reforms. In addition, Anderson finds that blacks interviewed by black interviewers are more likely to favour spending increases on welfare and food stamps.⁴¹ Blacks may be more affected by these policies, and will be more likely to divulge their opinions to those who may be similarly affected. Welfare and foodstamps do have a social stigma, and blacks may not want to admit affiliation — even by approval — to white interviewers. These types of sensitive issues may be the only area where interviewers have

40. Anderson, Silver, and Abramson, "The Effects of the Race of the Interviewer," 291.

41. *Ibid.*, 293.

effects on black respondents. Anderson argues that previous research provides evidence that interview-related effects are more noticeable in questions related to expressions of personal feelings toward members of another race.⁴² Effects, Anderson concludes, will come "into play in situations in which there are unambiguous references to race and a perceived threat to the maintenance of a polite conversation between the respondent and the interviewer."⁴³ In situations where respondents do not possess other characteristics which would appear to make them prefer one policy choice to another, Anderson argues that the race-of-interviewer effects are not spurious.⁴⁴ In personal interviews, the race of the interviewer can therefore have a significant effect in certain types of questions. Race effects need to be understood because if they are ignored and survey results are taken at face value, this may have implications for the understanding of blacks' attitudes toward federal aid programs.⁴⁵ Race effects clearly occur in some instances in personal interviews. It is also important to look at race effects in a more commonly used interview today: the telephone interview.

It would be expected that telephone interview respondents would be less likely to suffer from race-of-interviewer effects. First, they are unable to see the interviewer, and are therefore unable to recognize the interviewer's race by sight. As well, the psychological distance across telephone lines may reduce any feelings that the respondent may have to maintain politeness or avoid offending the interviewer,⁴⁶ compared to a face to face conversation. Tucker points out that even if the race of interviewer effects are

42. Anderson, Silver, and Abramson, "The Effects of the Race of the Interviewer," 290.

43. *Ibid.*, 319.

44. *Ibid.*

45. *Ibid.*, 320.

46. Patrick R. Cotter, Jeffrey Cohen, and Philip B. Coulter, "Race-of-Interviewer Effects in Telephone-Interviews," *Public Opinion Quarterly* 46 (1982), 279.

small, they are compounded much quicker and to a greater extent with the larger caseloads of telephone interviewers, as compared to door to door interviewers. "[T]he cumulative effects of intra-interviewer correlations can be substantial."⁴⁷ Where these effects occur should be examined in greater detail.

Studies paying careful attention to the race of the interviewer in telephone surveys have similar results to those studying interviewer effects in direct personal interviews. Cotter argues that the race of the interviewer does not have an effect (in his study) on almost all non-racially involved questions. The one exception is a question on gun-control. In addition, white respondents tend to give answers that are more pro-black than when they are interviewed by a black interviewer. Cotter advocates that it would appear that white respondents are generally more likely to be affected by race of interviewer effects than are black respondents. Perhaps whites in his study were still affected by the need for the maintenance of polite conversations, whereas black respondents behaved the same for interviewers of both races. One important note in this study is that Cotter uses all women interviewers for the black interviewers, where the white interviewers are both men and women. These effects may therefore be caused by the sex of the interviewer, than by the interviewer's race.⁴⁸ The effects of the sex of the interviewer will be examined next.

At this point however, it may be plausible to say that respondents were more polite to blacks because they were women, than only because they were black. Nonetheless, some of the race of interviewer effects may not be spurious. Correcting them may be problematic. Since the race of the respondent is unknown at the time of the initial call, if the researcher wishes to have respondents and interviewers matched by race, it may be a game of chance.⁴⁹ Prior contact may be a method by which to match respondents and

47. Tucker, "Interviewer Effects in Telephone Interviews," 94.

48. Cotter, Cohen, and Coulter, "Race-of-Interviewer Effects," 280-81.

interviewers. Still, this can add on to the cost of the study considerably, as well as creating problems such as re-interviewing sensitivity. It is difficult therefore, for researchers to eliminate, or in this case, even to establish definitively, race of interviewer effects in telephone surveys.

Interviewer Gender

Gender effects in telephone interviews have also been the subject of several statistical studies. Some effects of the gender of the interviewer are thought by some to impact significantly on the survey estimates. One area of importance in nonsampling error is that of refusals to participate. Gender may be an important factor in this sphere. For example, Groves argues that previously the fear of crime has been used to explain nonresponse in surveys. An unknown male voice, he continues, may be perceived as a greater threat to a respondent's safety than a stranger who is female. In fact, previous studies have demonstrated that men garner higher rates of refusal in general than do women interviewers. Specifically, men have refusal rates that are consistently five percentage points higher than women interviewers. While novices of both sexes tend to have higher refusal rates, this does not explain entirely the sex differential.⁵⁰

Nealon in 1983 found that male interviewers had significantly shorter interviews. In addition, women interviewers may be given more information by respondents than men interviewers.⁵¹ Respondents in these instances clearly feel less threatened and are more willing to open up to a female interviewer than a male interviewer. Or, women may value more conversational interaction than men, and therefore their sessions will be prolonged in

49. Cotter, Cohen, and Coulter, "Race-of-Interviewer Effects," 284.

50. Richard Groves and Nancy H. Fultz, "Gender Effects Among Telephone Interviewers in a Survey of Economic Attitudes," *Sociological Methods & Research* 14 (1985), 36.

51. Groves, *Survey Errors and Survey Costs*, 32.

comparison. In either case, this may have interesting consequences for different areas of questioning.

One particularly interesting finding shows that when respondents are interviewed by a man, they will express greater optimism toward varying economic indicators. In fact, Groves finds that the responses to male interviewers are three to five percent more positive than those to female interviewers.⁵² This may be caused by several factors. First, more male interviewers were employed toward the end of his study, when the public was admittedly more optimistic about the economic conditions of the country. On the other hand, these results may show that if optimistic economic attitudes are perceived as a demonstration of strength, male respondents may wish to give this portrayal to an interviewer in order to appear more successful. The gender of the respondent therefore is important in comparison to the gender of the interviewer.

Alternatively, respondents may assume that a male asking economic questions for the University of Michigan is a professional economist, whereas a female asking similar questions is a clerical person. Either interpretation could lead respondents to believe that the male interviewers would be more optimistic about their own economic future.⁵³

There could therefore exist some underlying social and psychological reasons why the sex of the interviewer has an effect on the economic optimism of the respondent. Or, as Groves points out, this study may merely reflect the age of the interviewer. Male interviewers in this study are generally younger than the female interviewer. Younger interviewers could be eliciting answers that are more optimistic as a function of their age, outlook, or because they simply reach a different type of respondent.⁵⁴ These effects

52. Groves, *Survey Errors and Survey Costs*, 48.

53. *Ibid.*, 49.

54. *Ibid.*, 45.

thought to be caused by sex may also therefore be problematic to study, or to apply to further research.

As in race matching in telephone interviews, it would be difficult to arrange for respondents and interviewers to be matched by gender prior to the interview. In addition, the predominance of female interviewers may threaten the applicability of these findings to the general population. As well, these studies, as with some of the studies encompassing the issue of race, may fail to control for other variables that could explain a gender effect.⁵⁵ Therefore, while these effects may be significant in some studies, the results may be problematic to apply to all studies. Moreover, attempts to correct these problems are not forthcoming because of the difficulty, and the cost of doing so. Perhaps if some of these studies paid more careful attention to other variables, then the flaws of application could be avoided. Otherwise, while these studies show some degree of interviewer effects, their results may be construed as controversial at best, and meaningless at worst.

Conclusion

Interviewer error can impact on survey results in many different aspects of the research project. It may distort the conclusions that researcher make of survey results. This can have dangerous potential consequences for the ways that people are believed to be associated with particular policy options. Researchers should not conclude without empirical evidence that a certain type of researcher will produce the best results. The acceptance of a certain type of interviewer as producing less effects without further study can ultimately result in skewed results and flawed conclusions.

Interviewers need better training, and should learn that consistency, above all else,

55. Groves, *Survey Errors and Survey Costs*, 33.

is important for the study of interviewer effects. They must be careful to present questions in the same manner to all respondents, in vocalization, question wording, and in enumerating all response alternatives. In cases where interviewers must record respondent answers verbatim, they must be taught good methods of recording, so that information is not lost at this stage. Interviewers may also influence respondents in ways other than in their interaction with respondents. Clearly the socio-economic and demographic characteristics of the interviewer need to be examined carefully in relation to respondent answers. The interviewer's age, race, or sex could have the effect of either inflating or decreasing the survey estimates, depending on the respondent's reaction to these characteristics, the respondent's need to remain polite, and the need to give socially acceptable answers. In addition, I did not uncover any studies of this in a Canadian context, such as where the effect of a French or English accent in interviewing, or where other ethnic groups were studied as interviewers. The idea of producing a survey with English Canadian interviewers in Quebec close to the time of the referendum, for example, could be a large source of interviewer effects. This area clearly merits further investigation, as it could potentially affect survey estimates as much, if not more, than the areas already discussed.

One problem that needs to be solved is that in research that examines interviewer effects, only one variable should be manipulated at a time. Otherwise, it becomes difficult to determine which factor, or which combination of factors clearly influenced the survey estimates. Interviewer error, and the skewed results it produces, could cause survey results to be misinterpreted or misapplied to many different areas of society, such as government funding programs, or to policy approval plebiscites. This could have potentially disastrous consequences for large segments of the population. Surveys are clearly an important part of the information age, and as long as interviewers are involved, their effects need to be

continually examined.

Chapter 2

Question Related Influence and Survey Error

Problems connected with the survey instrument also comprise an important part of the discussion of nonsampling survey error. Researchers often avoid discussion of this type of error. Errors related to questions are difficult to pin-point because of the complexities of the survey, hard to prove in the extent of their impact, and problematic to solve. This type of error overlaps somewhat with the discussion of interviewer influence on survey results. It is clearly related on some levels, though it can become more complex than the influence of the interviewer. While traits or behaviours of interviewers may be obvious, and at times clearly influencing responses, the problems surrounding the questions themselves may be more ambiguous. Problems associated with questioning may be further contrasted with the extensive study of sampling error. The comparative surprising lack of knowledge about questions and questioning can have dire consequences for the potential usefulness of survey data.¹ Information gleaned from a survey may be limited in its applications if it is not a true representation of the concepts it attempts to measure, or limited in its usefulness, at least for those who commissioned the work. While the survey may accurately reflect some underlying concept, this may not be the one that the researcher intended to measure, which can be problematic if the researcher applies the survey results to other unrelated areas nonetheless.

Others claim that nonsampling survey error is the product of deliberate malevolent manipulation. Cynics argue that researchers can readily produce the survey results desired by a careful change in questioning. It is debated, however, whether this would be easy to accomplish on demand. Measuring one clear concept is difficult enough on its own

1. Hans-J. Hippler and Norman Schwarz, "Response Effects in Surveys," in *Social Processing and Survey Methodology*, eds. Norbert Schwarz and Seymour Sudman, (New York: Springer-Verlag, 1987), 102.

attempt, but attempting to skew the results, and therefore measure another underlying concept, would be even more confounding. Other critics may be equally wrong in taking a fundamental view of survey results, or marginals, as wholly representative of public opinion.² Since concepts are difficult to measure, surveys may not completely and definitively represent public opinion. Opinions are in fact multifaceted and contradictory, and may not be well represented by survey questions. Nonsampling error does present the possibility of swaying marginal results a great deal. This can have practical implications if surveys are treated as referenda for public issues of concern.³ Nonsampling error, in general, and question influence in particular, must therefore be taken very seriously as a central part of the discussion of nonsampling survey error. Some of the more specific areas that should be examined are question design, context, wording, scales — measurement and options, order and length. Without such an examination, survey results might be otherwise misunderstood, or worse, could be seriously misused in their application in government and market research, as well as in the understanding of public opinion.

Questionnaire Design, Type and Effects

The overall design of the questionnaire is an important place to begin the examination of the impact of questions on nonsampling error. It is often overlooked as a source of potential effects, as well as in the increase of error in the survey.⁴ Sanchez points out that interviewer knowledge and experience will not balance out the effects of problems

2. Howard Schuman, "Ordinary Questions, Survey Questions, and Policy Questions," *Public Opinion Quarterly* 50 (1986), 441.

3. *Idem*, "Context Effects: State of the Past/State of the Art," in *Context Effects in Social and Psychological Research*, eds. Norman Schwarz and Seymour Sudman, (New York: Springer-Verlag, 1992), 12.

4. Maria Elena Sanchez, "Effects of Questionnaire Design on the Quality of Survey Data," *Public Opinion Quarterly* 56 (1992), 206-7.

in survey design. Training and supervision can not overcome the problems inherent in poorly designed and worded questions that leave the respondents and the interviewers confused. The survey instrument must be clear, and the intentions of the researcher must be understood by the interviewers, so that directions are strictly followed and interviewer influence is minimized. A survey must aim to be consistent in meaning to all respondents and interviewers. Reducing overall problems of survey design error will ultimately be beneficial in ways other than simply improving the quality of responses. Costs may be reduced in interviewer training and supervision, as well as in data retrieval.⁵ All of these improvements will only serve to improve continually the responses to the surveys, as well as the usefulness of the information for later studies.

One important factor in discussing overall survey design is the type of survey used. Mail and telephone surveys may differ in important ways with respect to effects on respondents. While some of these effects will be discussed in greater detail later, the differences between mail and telephone surveys should be outlined clearly first. One obvious difference, first discussed by Schuman and Presser in 1981, deals with the appearance of questions and order on the respondent. In mail surveys, respondents are able to review the entire questionnaire before answering any questions. In telephone surveys however, respondents are limited by their own powers of recall of previous questions, in order for this information to impact on later questions.⁶ Early questions may indeed influence what respondents reply if they recall earlier questions that infer what the researcher is attempting to measure. Mail and self-administered surveys may allow the

5. Sanchez, "Effects of Questionnaire Design," 216-7.

6. John Tarnai and Don A. Dillman, "Questionnaire Context as a Source of Response Differences in Mail and Telephone Surveys," in *Context Effects in Social and Psychological Research*, eds. Norman Schwarz and Seymour Sudman, (New York: Springer-Verlag, 1992), 115.

respondent to put all of the questions in a larger context, which at times may be beneficial to both the respondents, who are able to synthesize better the survey information, and the researchers, who may receive more complete information.

Clearly there exist both advantages and disadvantages to the two types of surveys. Schwarz argues that without the visual cues provided by a self-administered survey, respondents will be less likely to pick up comparison informational questions. Response scales of comparative questions may therefore be more accurate in telephone surveys if respondents are forced to answer more truthfully, and without the influence of social desirability factors because they cannot recall their previous responses or response scales of the questions.⁷ More truthful and complete reporting would clearly result in better survey information.

Tarnai however, argues that mail surveys may have distinct advantages over telephone surveys. His study shows that "telephone respondents are more likely to select the extreme positive end of the response continuum when making a response than are mail survey respondents, and this is true primarily for questions in a series using vague quantifier response categories."⁸ Telephone respondents, limited in their power to recall all of the categories that the interviewer lists, may simply select the one that is most memorable: usually one of the end categories. Ayidiya agrees that question effects may be minimized in mail surveys.⁹ Mail surveys may minimize effects because the respondent is

7. Norbert Schwarz and Hans-J. Hippler, "What Response Scales May Tell Your Respondents: Informative Functions of Response Alternatives," in *Social Information Processing and Methodology*, eds. Norbert Schwarz and Seymour Sudman, (New York: Springer-Verlag, 1987), 176.

8. Tarnai and Dillman, "Questionnaire Context as a Source of Response Differences," 120.

9. Stephen Ayidiya and McKee J. McClendon, "Response Effects in Mail Surveys," *Public Opinion Quarterly* 54 (1990), 230.

not limited in time or recall to answer the survey questions.

In addition, the mail survey may be better in terms of social influences of interviewers on respondent behaviour. Interviewers may affect responses in that their presence will result in respondents wishing to give answers that appear consistent with their previous answers, but are otherwise untrue. As well, respondents may desire to appear socially respectable, and to conform to the norms of society, instead of giving the most honest answers possible. One study shows that telephone respondents tend to report more socially desirable behaviour than those answering mail surveys.¹⁰ These effects, Ayidiya continues, should disappear in the case of mail surveys where an interviewer is not present. Nonetheless, Ayidiya's studies prove that context effects still occur in mail surveys. This, he argues, must therefore be caused by internalized norms of the respondents, rather than just interviewer effects.¹¹ These effects can not therefore just be avoided by changing survey methods. In addition, the magnitude of these effects may vary depending on the level of interest of the respondent in the survey. A more interested respondent may report with greater consistency in a self-administered questionnaire. Therefore, with all of these factors in mind, the issue is more complex than whether or not an interviewer, or what types of questions, appear before the respondent.

The context of the question in both telephone and mail surveys remains important. One study demonstrates that different attitudes can be "retrieved about the same brand depending on the function that the attitude might serve. For example, Perrier water could serve as a social function (status) or a utilitarian function (taste).¹² In either mail or telephone surveys, the context of the question indicates to the respondent what the

10. Tarnai and Dillman, "Questionnaire Context as a Source of Response Differences," 119.

11. Ayidiya and McClendon, "Response Effects in Mail Surveys," 234-5.

researcher would like to know. Context can certainly clear up ambiguity, and may even cause the respondent to recall more instances of occurrences of what the researcher is attempting to measure. All of these items are dependent on the question being clear and concise, and well understood by the respondent.

Question Comprehension

Question comprehension is extremely important in all types of surveys. Clear questions mean that respondents at least understand what they are asked, regardless of the amount of truth in their answers. Questioning is not as easy as it seems. Asking the right question in survey research is paramount, unlike general conversational situations where a person can continue to question, even repetitively, to uncover information.

A fundamental paradox of survey research is that we start from the purpose and use the method of ordinary questioning employed in daily life, yet our results are less satisfactory for this purpose than for almost any other.¹³

Yet one important point is that questioning in surveys is very different from regular conversation. First of all, standardizing questions in surveys produces complexities that do not necessarily arise in normal conversation, where more give and take occurs between the respondent and the questioner. The normal rules of conversation are inherently violated in such a situation.¹⁴ The survey atmosphere itself is artificial, and a dictated conversation does not make the situation more natural.

12. Barbara A. Bickart, "Question Order Effects and Brand Evaluations: The Moderating Rule of Consumer Knowledge," in *Context Effects in Social and Psychological Research*, eds. Norman Schwarz and Seymour Sudman, (New York: Springer-Verlag, 1992), 66.

13. Schuman, "Ordinary Questions," 432.

14. Norman M. Bradburn and Seymour Sudman, "The Current Status of Questionnaire Design," in *Measurement Error in Surveys*, eds. Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz and Seymour Sudman, (New York: John Wiley and Sons, 1991), 36.

Problematic questions and a rigid format are not the only difficulty in the survey interview. Respondents may be an important part of the question understanding equation. Fee finds that among the best questions, only half of the respondents interpret them as intended by the researcher. Some questions are understood by less than one quarter of the respondents. Fee continues to argue that respondents often mentally alter the question so as to make it more applicable to their lives, and therefore easier to answer.¹⁵ Respondents may misunderstand a question because of the complexity of the concepts used, or because too many concepts are used in one question. Respondents may not even be fully aware that they do not understand, nor are they always willing to admit confusion.¹⁶ Social desirability can therefore play a part in the case of respondent who does not want to admit confusion. A great number of problems can occur therefore once the question is asked to the respondent. Even clear concepts may not be understood, and the artificiality of the survey interview may prevent the respondent from asking for clarification.

Problems are not always obvious, and even an experienced interviewer can fail to ascertain that respondents have difficulty with question comprehension. "Clearly an apparently adequate answer can mask misunderstanding: interaction coding cannot in all such cases be used to identify who has and who has not understood the question or the response task."¹⁷ This may be especially likely to occur in cases where the respondent is given a list of alternative responses to select. Or, respondents may select a "no opinion" or "don't know" response, if they are unclear about the concept or the issues at hand. These can be problematic for interpretation of public opinion on a given measure.

15. Joan Flynn Fee, "Symbols in Survey Questions: Solving the Problems of Multiple Word Meanings," *Political Methodology* 7 (1981), 72.

16. Jean Morton-Williams and Wendy Sykes, "The Use of Interaction Coding and Follow-up Interviews to Investigate Comprehension of Survey Questions," *Journal of the Market Research Society* 26, no. 2 (1984), 109.

17. *Ibid.*, 114-5.

A related issue is that respondents may interpret the same questions differently. Fee argues that certain concepts are clearly related to certain demographic attributes, and therefore occur nonrandomly in populations. "The tendency of definable populations to perceive a concept within a survey question differently from other definable populations makes it crucial for a researcher to measure meaning in order to understand the biases implicit within the question."¹⁸ Clear questions will therefore not always give the best responses if the researcher fails to take into account the biases inherent in certain contexts or concepts used in question wording. One example could be education, where the same question will be understood differently by respondents with different levels of education. Clearly those with higher levels of education will be able to grasp more difficult concepts that may elude a less educated respondent. Bishop believes that attitudes as revealed in surveys may be more a reaction by a psychological factor to response effects, than real attitudinal variation.¹⁹ The target population must therefore be kept in mind when designing and implementing questions for survey research. Clearly respondents are important to consider in the quest for clearly understood questions and an improved survey design.

Question Context

As mentioned briefly in the preceding section, the context in which survey questions are posed is of great importance for respondents' understanding and responses. One aspect of this arises when a series of questions is asked about a particular activity. Such a sequence of questioning may suggest to the respondents that if they have answered a question about an issue previously, then they should indicate in future questions that they

18. Fee, "Symbols in Survey Questions," 89.

19. George F. Bishop, "Issue Involvement and Response Effects in Public Opinion Surveys," *Public Opinion Quarterly* 54 (1990), 210.

have thought about such an issue, are interested in the issue, or that it is important to them.²⁰ At this point, the desire of the respondents to appear consistent throughout the survey interview becomes apparent. The series of questions makes the issue of consistency even more paramount in their minds, as they are more likely to recall the preceding related questions. Yet even repeated measures have difficulty in remaining consistent in maintaining the issue and context captured by the successive questions.

Context may be problematic to control over time. The exact environment of the survey, the interaction of the interviewer and respondent, and even the reading of the questions may vary between survey projects, or even between interviews. "[I]t is difficult to hold context constant from one survey to another: questions are almost always repeated on a selective basis, and it is rare for an earlier question order to be maintained perfectly."²¹ Even if the context of the interview is strictly adhered to, it may not be possible to make broad conclusions about trends across quantitative studies.²² Generalizations about public opinion therefore may be premature, or even misleading. Schuman concurs that changes in marginals may just as much be a function of differing contexts and question reading across surveys than actual shifts in public opinion. Constancy may not be maintained across time or across variables, so surveys may be measuring differing concepts in different time frames.²³ This can make comparison of marginals problematic.

One particularly controversial area of the study of question context is that of

20. George F. Bishop, "Issue Involvement and Response Effects in Public Opinion Surveys," *Public Opinion Quarterly* 54 (1990), 216.

21. Howard Schuman, Stanley Presser, and Jacob Ladwig, "Context Effects on Survey Responses to Questions About Abortion," *Public Opinion Quarterly* 45 (1981), 216.

22. Schuman, "Context Effects," 12.

23. *Idem*, "Ordinary Questions," 433.

question ordering, and how marginals are affected. This issue becomes especially clear in the case of surveys dealing with abortion. In Schuman's study, a general approval of abortion question receives substantially more support if it is asked first in a series of questions about abortion. If placed after a specific question dealing with the issue of abortion and a defect in the foetus, then support dwindles. Interestingly enough, the question on abortion and foetal defects appears not to be affected by its placement in the questionnaire. An explanation may be in order here. Schuman argues that when the specific question is asked first, it suggests to the respondent that the general question that follows does not include the specific case as mentioned in the previous question. Therefore, people who feel that a defective foetus is only one of few adequate reasons for a pregnancy to be terminated will be less likely to support a general approval of abortion question if the defective foetus reason is not included.²⁴ This context effect could appear because of the assumption surrounding the placement of the two questions. The support for abortion may actually be constant, but the marginals will definitely change given different question placement. Schuman concludes that the "context effect on the general abortion item seems to be both large and highly reliable."²⁵ Abortion questions therefore prove that context effects do indeed exist, and survey researchers should be careful to be aware of the underlying assumptions that arise because of question order.

Context effects may be diminished by interposing neutral items between questions of a sensitive nature.²⁶ Since the context of a question may only exist briefly in the short term memory of the respondent, such neutral items may cause respondents to forget their previous answers, and therefore give answers to future questions without the need to

24. Schuman, Presser, and Ludwig, "Context Effects on Survey Responses," 220-221.

25. *Ibid.*, 218-219.

26. Howard Schuman, Graham Kalton, and Jacob Ludwig, "Context and Contiguity in Survey Questionnaires," *Public Opinion Quarterly* 47 (1983), 113.

remain consistent, or to think about the context of the previous questions. Yet dispersing unrelated items throughout a questionnaire may not eliminate all context effects.²⁷ The importance of a researcher's awareness of context effects should therefore not be devalued. When different researchers arrive at differing marginals, explanations such as context, question order and question wording often arise. Still, some of the differences are clearly due to changes in the concepts being measured, and not the methods by which researchers attain such figures.

Schuman cautions that question content is generally responsible for large shifts in the marginals, not minor variations in wording. He argues that since many major issues are discussed at a general level, rather than a specific one, with only two sides presented in the debate, that question content must be responsible for changes in public opinion. Schuman advocates that content refers to the overall discussion of the issue, and that questions change meaning depending on the wording used. Minor variations in wording however, should not affect the overall content of the understanding of the issues in a particular question. In the case where public opinion is argued to have been manipulated because of a minor change in question wording, Schuman believes that there is simply not enough change in the question wording, and therefore a shift in question content, in order for such large shifts to occur. In this way, content refers to the issue, and not to the particular question wording.

Marginals must be understood as to exist at the moment they are taken, and not to be extrapolated to different time periods or contexts. This, Schuman argues, is important, because such marginals do not take into account important public issues or leadership that can have a one-time impact on them.²⁸ The background that exists while the survey is

27. Howard Schuman, Graham Kalton, and Jacob Ludwig, "Context and Contiguity in Survey Questionnaires," *Public Opinion Quarterly* 47 (1983), 114.

being taken should therefore also be considered when interpreting marginals. In addition, context effects on marginals may be especially likely when researchers attempt to simplify complex issues with a single survey item. This further links context with question wording.

Question Wording: Abstraction and Frequency

Question wording broadly encompasses the issues that arise in the phrasing of survey questions. Though some dismiss the issue of wording in favour of the issue of context, others maintain that it remains vital to the understanding of nonsampling error. Bradburn and Sudman argue that from the beginning of public opinion polling in 1935, researchers recognized the impact of question wording on respondent answers. George Gallup himself strongly advocated the use of split-half tests to judge question wording effects.²⁹ Yet despite the recognition of the problematic nature of question wording, other than the use of two questionnaires, researchers felt that there was little one could do to remedy this problem.

In fact, Hippler argues that researchers emphasize their interest in general relationships between variables — which ought to be measurable and understandable given *any* question wording. Researchers do not seek, he continues, to examine the relationships discovered by particular questions. Hippler explains that Stouffer and De Vinney (1949) call this “form resistant correlations”,³⁰ further emphasizing the importance of the general relationship and not the particular question wording. Yet this issue is not as simple as providing a disclaimer about larger variable relationships. More

28. Schuman, “Ordinary Questions,” 436.

29. Bradburn and Sudman, “The Current Status of Questionnaire Design,” 29.

30. Hippler and Schwarz, “Response Effects in Surveys,” 103.

recent studies clearly prove that question wording does have an impact on marginal results, and question wording is clearly interpreted as a snapshot of public opinion, often without other circumstances or context placed in the explanation of survey results.

Closely related to context, the use of particular words may change over time in survey questions. Repeated measurements may therefore be rendered noncomparable.³¹ Therefore, while the wording of a question may remain the same, the results may change because of the passage of time, and the changing nature of words and concepts in society. In addition, even minute changes in wording may have a great impact in the understanding of a question. "Consider the difference between the words *a* and *the*. Such articles in English tend to be treated as unimportant words, yet it is really no surprise that substitution of one for the other can have large effects."³² It can make a sentence more specific, or even change the meaning entirely. If this switch occurs without the awareness of the researcher, such as inadvertently in the reading of a question, the marginals may be greatly affected.

In fact, as mentioned in the discussion of question context, the meaning of certain words may differ greatly among different respondents. Fee finds that how respondents understand certain concepts is not only related to background characteristics, but may also be a predictor of political and policy choices.³³ This may be stimulated in some cases by the wording of the question. Hippler argues that negative forms of questions are more likely to be supported than the affirmative form of the same question.³⁴ How respondents interpret the words of the question, and how that question is placed in context with their

31. Hippler and Schwarz, "Response Effects in Surveys," 105.

32. Schuman, "Ordinary Questions," 435.

33. Fee, "Symbols in Survey Questions," 89-90.

34. Hippler and Schwarz, "Response Effects in Surveys," 105.

particular background characteristics, will trigger differing responses given different versions of the same question.

Deciding on the best question wording may prove to be a tricky task. Too simple a wording may insult the respondents, yet too complicated may be easily misunderstood.³⁵ Leading questions or questions using too sensitive wording that will derive an emotive response should be avoided. Yet if these concepts or words are used, then attempting to find equivalent wordings for split-half tests could prove to be difficult. In addition, DeLamater argues that all of the possible response alternatives to a question should be included.³⁶ Sensitive or threatening topics are often perceived as less likely to receive valid answers from respondents. Including all of the alternatives to a particular topic indicates to the respondent that the interviewer is not just looking for socially acceptable responses, and that the respondent should feel free to offer all types of information.

DeLamater explains that Bradburn and Sudman prove that respondents will answer almost all survey questions because the norms of the interview environment will suggest to them a requirement of doing so. In fact, DeLamater argues that more than half of all survey interview refusals occur before the interviewer reads the introduction to the survey. The respondent therefore has little, if any, information about the potentially sensitive nature of the topic, and therefore is refusing on the general principle of not wanting to participate in survey research, and not based on the topic itself. A decision to answer a particular question, in contrast, is based on specific information about the topic at hand.³⁷ Therefore, survey refusals that are given before the interview begins can not be affected by

35. John DeLamater, "Response-Effects of Question Content," in *Response Behaviour in the Survey Interview*, eds. W. Dijkstra and J. VanderZouwen, (Toronto: Academic Press, 1982), 22.

36. *Ibid.*, p. 25.

37. DeLamater, "Response-Effects in the Survey Interview," 17.

question wording, whereas refusal to answer a specific question can be affected by changes in the question wording.

Problems occur in question wording experiments that often make finding the root of the problem difficult to identify. In some experiments, for example, more than one wording or variable in the question is altered at the same time. When other facets of the question are not held constant, it becomes impossible to determine which wording manipulation really produce the change in result.³⁸ Comparing wordings that are substantially different may also be problematic. Abramson finds that comparing the prefaces to questions posed by Gallup and SRC lead a researcher to make very different predictions about the stability of partisanship.³⁹ Therefore, in order for effective comparisons to be made between versions of questions, there must exist a fair amount of consistency, with only one real variable being manipulated in each case.

The addition or subtraction of words from a question is another way to vary the question, its underlying concepts, and its interpretation by the respondents. This may be just as effective in producing change in the marginals as the substitution of certain words. Hippler argues that if the researcher adds nonredundant words to the question, the complexity of the question will be increased, and therefore will in turn increase the variation of respondent answers. Respondents may interpret the question differently, and react to differing facets of the same question. Hippler argues that it is therefore more important for questions to be kept simple, than to be kept short.⁴⁰ Clarity can be found in

38. Nico Molenaar "Response-Effects of 'Formal' Characteristics of Questions," in *Response Behaviour in the Survey Interview*, eds. W. Dijkstra and J. VanderZouwen, (Toronto: Academic Press, 1982), 84.

39. Paul R. Abramson and Charles W. Ostrom, "Question Wording and Partisanship: Change and Continuity in Party Loyalties During the 1992 election Campaign," *Public Opinion Quarterly* 58 (1994), 43.

40. Hippler and Schwarz, "Response Effects in Surveys," 107.

brevity, though this is no longer considered to be the rule in designing survey questions.

Ambiguity in survey questions should be attempted to be eliminated by the researcher. Since respondents may be hesitant to ask for clarification in the case of an ambiguous word or question, they may give information different from what the question really asks.⁴¹ Their answer will depend on how they interpret the ambiguous word in question, often based on the context of the question in the questionnaire, the topic of the survey, and interviewer signals.⁴² Clearly problems may develop here—at least insofar as inconsistency develops between surveys, or even within the same survey, when given at different time periods. Words may gain new meaning for the respondent, and understanding of a particular concept may change the reporting of a person's views.

Words that do not appear difficult to understand may in fact be ambiguous with respect to particular questions. They become ambiguous in that they change meaning in different contexts. For example, comparative adjectives like "richer" and "poorer" may be problematic. Given such adjectives, respondents may feel compelled to give a frame of reference for explanation as to why they assigned certain responses such descriptions. "The concepts used in this question were thus seen as complex, imprecise and confusing; many respondents clearly were not able to handle them at the level of broad abstraction perhaps hoped for by the researcher."⁴³ The words, and the concepts they represent are not difficult to understand. What makes this task problematic is for the respondent to be able to put these concepts in context, and to assign them meaning in terms of response

41. Bradburn and Sudman, "The Current Status of Questionnaire Design," 35.

42. Fritz Strack, "'Order Effects' in Survey Research: Activation and Information Functions of Preceding Questions," in *Context Effects in Social and Psychological Research*, eds. Norman Schwarz and Seymour Sudman. (New York: Springer Verlag, 1992), 127.

43. Morton-Williams, and Sykes, "The Use of Interaction Coding," 116-7.

categories.

Some problems do arise because of the nature of the particular word used. Certain words may bring about a negative or positive label to an issue. This may have an effect on public support for certain policies. "Smith (1987) has argued that different labels for the welfare issue may bring to mind different associations, actually changing the stimuli to which respondents are reacting."⁴⁴ Smith's results prove that substantial differences do occur when the word "welfare" is substituted with "assistance for the poor" and "caring for the poor." People are much more likely to support public aid if "welfare" is not used. Smith describes "welfare" as arousing negative feelings associated with minorities, wastefulness, dependency, unproductiveness, and abuse.⁴⁵ Yet these explanations of the lack of support when "welfare" is used may not be the entire truth. Differing and even conflicting explanations may exist that explain why support for public assistance drops. Tourangeau finds that "some people oppose welfare because they think that it is mismanaged, whereas others oppose it because they think that it is misconceived."⁴⁶ The distinction between these two explanations must be carefully examined.

Some words may gain a more positive association with other concepts underlying a survey question. For example, Smith advocates that unemployment may gain more acceptance if it is associated with workfare programs, and not with the word "welfare." These and other examples show the importance of examining the words and the concepts

44. Kenneth A. Rasinski, "The Effect of Question Wording on Public Support for Government Spending," *Public Opinion Quarterly* 53 (1989), 392.

45. Tom Smith, "That Which We Call Welfare By Any Other Name Would Smell Sweeter: An Analysis of the Impact of Question Wording on Response Patterns," *Public Opinion Quarterly* 51 (1987), 79.

46. Roger Tourangeau, "Context Effects on Responses to Attitude Questions: Attitudes as Memory Structures," in *Context Effects in Social and Psychological Research*, eds. Norman Schwarz and Seymour Sudman, (New York: Springer-Verlag, 1992), 35.

understood by respondents in survey questions. If policy makers are not aware of the original question wording, they could be led to believe that the public has a callous attitude toward the poor, where in fact, respondents were reacting to the wording of the question.⁴⁷ Social programs are complex issues, and clearly may be misjudged by survey measures. The problem at hand is then not one of unrecognizable words, but of the imprecision surrounding difficult concepts in varying contexts.⁴⁸ These issues must be clarified not only for the respondent, but also for the researcher, and for the social scientist or policy maker who will ultimately make use of the survey results.

Quantifiers are meant to clarify survey responses, and aim to add precision to a survey measure. Abramson argues that when "usually" is not included in a question that is meant to judge long term meaning, respondents treat it more as a snapshot of opinion only at that particular point in time.⁴⁹ Quantifiers can in fact make survey responses more complex.

While we assume that the quantifiers involved are ordered in their intensity, we know little about the characteristics of such quantifying adverbs. For example, do these imprecise quantifiers have some common meaning that is more precise than is apparent on the surface? ... One man's 'rarely' is another man's 'hardly ever.'⁵⁰

Without precise definitions, such as numerical examples, quantifiers may distort respondent answers. Using such quantifiers may not actually improve the distinction of the frequency of respondent behaviour or participation in an activity.⁵¹ Respondents have a great deal of difficulty defining the meaning of quantifiers. "In general, interviewers

47. Smith, "That Which We Call Welfare," 82-83.

48. Morton-Williams and Sykes, "The Use of Interaction Coding," p. 126.

49. Abramson and Ostrom, "Question Wording and Partisanship," 24.

50. Norman M. Bradburn and Carrie Miles, "Vague Quantifiers," *Public Opinion Quarterly* 43 (1979), 92-3.

reported that giving a more precise meaning to degrees of "often" was difficult for respondents and that they were reluctant to give responses."⁵² The researcher should try to make the question as clear as possible, so that the respondent will give the most accurate response, and not a widely varied estimation as understood by a quantifier. While the quantifier may imply a certain frequency given the context of the question, it may still vary greatly between questions.

Context is in fact extremely important in the use of quantifiers, even given the differential between questions. The placement of the adjectives in the group of response categories may give the respondent a sense of "standard" in order to answer the question.

Thus the mean response for 'sometimes' in the shooting in Hollywood westerns context is higher than the mean response for 'very often' in the context of earthquakes. In short, "'often' for an improbable event is less 'often' for a highly probable event."⁵³

In another example, context can also be proven important to the understanding of quantifiers, even for the same topic. Schwarz discusses how respondents will gain insight from both the question and the possible responses from context. If respondents are asked how often they are irritated, they must understand what is meant by irritation before they respond to the question. If the possible responses are "several times a year" to "less than once every three months," then the respondents will correctly assume that the researcher intends to measure more severe types of irritations, rather than day to day annoyances.⁵⁴ Therefore, the same question with different quantifiers as responses can definitively affect

51. Susan Pepper and Lubomir S. Prytulak, "Sometimes Frequently Means Seldom: Context Effects in the Interpretation of Quantitative Expressions," *Journal of Research in Personality* 8 (1974), 100.

52. Bradburn and Miles, "Vague Quantifiers," 95.

53. *Ibid.*, 94-5.

54. Schwarz and Hippler, "What Response Scales May Tell Your Respondents," 170.

respondent answers. In addition, quantifiers can have very different meanings in different questions that concern different topics. The probability of occurrence appears to affect how respondents judge its frequency on scale items.

Another way by which respondents may be affected on this type of scale is whether they favour the occurrence, or if they participate in such an activity themselves. Bradburn argues that a person who favours less, or participates less, in an activity is more likely to rate it as occurring more often. In contrast, those who engage in the activity or highly approve of it will be likely to use lower level quantifiers.⁵⁵ Pepper asserts that lower frequency expressions such as "never" are fairly precise.⁵⁶ These expressions are therefore more likely to be able to be compared among different measures and different questions. Nonetheless, context remains an important factor in how quantifiers are understood by respondents.

Researchers add in quantifiers to survey instrument responses to improve the explanatory power of respondent answers. To some degree, this is helpful for both the respondent and the researcher. Bradburn asserts that the variance would be improved if individual estimates were included in surveys instead of the use of quantifiers. This however, he understands to be not always feasible in survey research. Respondents have difficulty with providing exact estimates,⁵⁷ and this could possibly result in an increase in non-responses in the data set, which would weaken the applicability of the results. Nonetheless, researchers continue to believe in the importance of questions that involve figures in almost all survey research, such as in the "how often" type questions. While respondents may have difficulty providing exact estimates, this does not negate the

55. Bradburn and Miles, "Vague Quantifiers," 94-5.

56. Pepper and Prytulak, "Sometimes Frequently Means Seldom," 95-6.

57. Bradburn and Miles, "Vague Quantifiers," 100.

necessity of the inclusion of these questions. Therefore, researchers continue to use quantifiers, and hope that the clarification of the quantifier by context will aid the respondent in understanding the researcher's scales.

The perception of the respondent attitudes or opinions as judged by scale measures is affected by the context of the question within the survey instrument. Bishop discusses an interesting study that measures self-perceptions of respondent knowledge of government officials' records and interest in politics. Bishop finds for example, that if respondents are asked difficult questions about their congressional representatives' record before being asked about their interest in politics, they are much less likely to respond that they follow politics much of the time.⁵⁸ Clearly respondents infer from the preceding political knowledge questions that these are the measures by which the interviewer gauges respondent political interest, and so therefore they are not that knowledgeable (and or interested) in politics. One way to change the process by which respondents infer information, and place themselves in a perceived context of a knowledge scale, may be weakened by dispersing the questions throughout the questionnaire.

This particular context effect relates not just to the question, but to its placement in the survey instrument. Bishop asserts that it may not be the spacing of the questions that affects respondents so much as the distinction in content of questions that are placed between key variable questions. Bishop continues to explain that context effects will inevitably occur throughout the interview, despite any buffers that the researcher inserts in the questionnaire.⁵⁹ Therefore, it may be better to gain a more clear understanding of

58. George F. Bishop, "Context Effects on Self-Perceptions of Interest in Government and Public Affairs," in *Social Information Processing and Survey Methodology*, eds. Hans-J. Hippler, Norbert Schwarz and Seymour Sudman, (New York: Springer-Verlag, 1987), 180.

59. Bishop, "Context Effects on Self-Perceptions of Interest," 182.

context effects so that they can be factored into the analysis of the data, instead of attempting to use ineffective measures to eliminate them altogether.

Respondents react differently to quantifiers and the context in which they are used. Morton-Williams finds that respondents will provide quantifying adjectives when presented with a numerical scale.⁶⁰ They react better to the quantifiers than to the presentation of absolute numbers, as mentioned previously. Respondents may not even regard a numerical scale as the best evidence of how they relate to a distribution. For example, Schwarz finds that respondents will use their biographical knowledge of how much they and others watch television, in order to locate themselves on a scale of television viewing. Respondents may not even consider how many actual hours of television they watch in evaluating their own viewing habits.⁶¹ Thus two people with very different patterns of television viewing may both consider themselves "average viewers." Therefore, respondents may have very different images of quantifiers, even if exact numerical examples are provided. This can make potential inter-subjective comparison difficult, even within the same survey.

Context and quantifiers can provide additional information to both the respondents and researchers. It can also however, create confusion and obfuscation of the issues being studied. Differences in populations can occur even in the same survey because of the use of scales where people interpret the estimates differently. Since scales play such an important role in survey responses and the use of quantifiers, they should be further examined in this discussion of nonsampling error.

60. Morton-Williams and Sykes, "The Use of Interaction Coding," 119.

61. Schwarz and Hippler, "What Response Scales May Tell Your Respondents," 165.

Scales — Measurement and Options

Researchers turn to scales to measure attitudinal items that may not be captured by a single question or by too few response options. While they have generally accepted the idea that question wording has the ability to impact on marginal distributions, they still seek to explore the general relationships between broad ideological issues. To minimize the problems that arise because of question wording, researchers turn to multi-item scales.⁶² Or, in a single question, too few responses may not provide enough differentiation for respondents to use such a type of scale effectively. Alwin argues that in this case respondent choices may be random, since respondent are unable to map effectively their attitudes according to the available options. Alwin continues to explain that while this supports the idea of increasing the number of on one response scale items for increasingly reliable measures, more than seven points will only increase the ambiguity of survey responses, and will no longer aid in the respondent's task.⁶³ Matell states that a delicate balance exists between the number of categories which produce a coarse scale and a scale so finely tuned it reaches beyond the respondents' discriminatory power of understanding the differences between responses.⁶⁴ Researchers must therefore attempt to balance explanation with clarity for the respondent and the researcher's needs. They make an effort to gauge the success of their scales at gleaning such information, especially in relation to the reliability and validity of the survey instrument and the data collected.

Some authors argue that the reliability of the survey instrument does not increase

62. Fee, "Symbols in Survey Questions," 71.

63. Alwin and Krosnick, "The Reliability of Survey Attitude Measurement," 148-9.

64. Michael S. Matell and Jacob Jacoby, "Is There an Optimal Number of Alternatives for Likert Scale Items? Study I: Reliability and Validity," *Educational and Psychological Measurement* 31 (1971), 657.

with a larger number of response categories in a Likert-type scale. Matell asserts that what is most affected by such scales is the directional component, and only in very small part, the intensity component.⁶⁵ In other words, the direction of the relationship, whether positive or negative, is most affected by scales. Respondents will likely choose the same intensity on a scale, be it positive or negative. This intensity, as often represented by items such as "agree slightly, agree a great deal" will remain the same, even if the direction changes. Scale components, according to Matell, do not affect the intensity of the result, as they had been intended to indicate, as they explain only the direction of the relationship between variables.

Furthermore, Matell argues that converting a large number of scale items does not necessarily increase the reliability and validity of the survey measure. Therefore, if reliability and validity are not affected, they must exist independent of the number of scale points used by the researcher for Likert-type items.⁶⁶ This view contrasts with the arguments of the researchers previously cited, who generally argue that reliability and validity increase with the number of response categories within a single question, up to a maximum of seven points. It is agreed within survey research that reliability will increase when the number of questions referring to a single issue is increased. This argument however, is in reference to the greatest number of response categories that should be included for maximum reliability in a single question. Which view of the maximum number of these categories is correct? Part of the problem could be the overlap that respondents attribute to scale items with vague quantifiers, and have a difficult time interpreting. The large variance in these variables could be weakening the statistical results, and therefore the results would appear to strengthen when condensed into fewer

65. Matell and Jacoby, "Is There an Optimal Number of Alternatives," 659.

66. *Ibid.*, 666.

categories. Yet even with fewer categories, it remains important to study each individual category, and the order in which they are presented.

Middle Responses

One particularly interesting part of the study of survey response order is that which examines the middle response. The order of the responses can affect which response the respondent is more likely to select. The middle response alternative seems particularly vulnerable to response order effects. Schuman and Presser (1981) find that respondents are much more likely to select a middle response when it is explicitly offered, than if they are forced to volunteer spontaneously such a response.⁶⁷ This may clearly affect the outcome of the distribution of responses. Bishop adds that in that same study, if the middle response is deleted from both the offered and omitted questions, no significant difference in distributions exists for respondents selecting the extreme ends of possible response alternatives.⁶⁸ Nonetheless, the initial findings are significant. They indicate that even the inadvertent suggestion of a middle response will change the outcome of the distributions.

Bishop further argues the point started by Schuman and Presser. His later study finds that when interviewers inadvertently mention the possibility of a middle response in a question, but do not repeat it in later response possibilities, it is still more likely to be selected by respondents.⁶⁹ Ayidiya concurs with Bishop on the issue of increased agreement with a middle response when offered.⁷⁰ This still makes survey comparison problematic, if researchers do not specify whether the middle response was offered, or if it

67. George F. Bishop, "Experiments with the Middle Response Alternative in Survey Questions," *Public Opinion Quarterly* 51 (1987), 220.

68. *Ibid.*, 221.

69. *Ibid.*, 226.

70. Ayidiya and McClendon, "Response Effects in Mail Surveys," 241-2.

is deleted from one survey but not the other. Researchers should look at responses also not only in terms of "middle of the road", but also the actual order of the choices.

The order of the responses may affect whether or not respondents choose them. Bishop argues that respondents will be more likely to choose a middle response when it is offered in the last position because it is easier to remember.⁷¹ Ayidiya disagrees at this point, and instead asserts that in his study, the middle response is more frequently selected when it is offered in the middle, rather than in the last position.⁷² The response order does appear, at least minimally to affect respondent choices. The direction of change and the position of most impact, however, are disputed among researchers. Abramson argues that excluding the middle response will increase variation. He states that a Gallup measure that excludes the middle response tends to become more volatile.⁷³ Measures clearly need to reflect best the actual opinions and behaviours of the respondents. A measure that distorts this picture needs to be minimized in its impact. Which is better? Should a middle response be offered?

Kalton presents the position that not offering a middle response is a better choice because respondents do inherently have an opinion on one side or the other of neutral, and the lack of a middle response will force respondents to make a decision. On the other hand, Kalton continues, if respondents truly fall in a middle position, and one is not explicitly offered, they will be forced to choose a false response, which will thereby distort the data.⁷⁴ Which is the lesser evil? Perhaps it would be better to include the response, so as to represent fully possible respondent opinions. Or, the interviewer could present the

71. Bishop, "Experiments with the Middle Response," 221.

72. Ayidiya and McClendon, "Response Effects in Mail Surveys," 238.

73. Abramson and Ostrom, "Question Wording and Partisanship," 44.

74. Graham Kalton, Martin Collins and Lindsay Brook, "Experiments in Wording Opinion Questions," *Applied Statistics* 27, no. 2 (1978), 150-1.

option in the question, but not repeat it in the list of possible responses. In relation to placement, middle responses may be best offered where they are numerically correct, in the centre of the distribution of responses. Then, if later researchers choose to compare this survey with other surveys which do not have a middle response, it could be deleted from the first survey, as occurred in Schuman and Presser's work, with no adverse affects.

Don't know - No Opinion

While not truly a middle response, the "don't know" or "no opinion" response choices face a similar debate over inclusion in survey measures. Noelle-Neuman argues that it is better not to present these options if the researcher would like to have the maximum amount of information from respondents.⁷⁵ This assumes that the offering of such a response will have a definite impact on the response patterns of respondents. In fact, these distributions can differ between the types of surveys administered. Ayidiya argues that in interviewer administered surveys, interviewers can discourage "don't know" responses and probe for further information. This is not possible on a self-administered or mail survey, where only the question and response alternatives can impact on respondent choices. Ayidiya finds that offering a "don't know" or "no opinion" response substantially increases the likelihood that these responses will be selected by the general population.⁷⁶ There must exist reasons why people select such answers, and this should be further examined.

Morton-Williams explains that many respondents select the "don't know" response because they do not feel qualified to give an answer otherwise. When people are not

75. Elisabeth Noelle-Neuman, "Wanted: Rules for Wording Structured Questions," *Public Opinion Quarterly* 34 (1970/71), 200.

76. Ayidiya and McClendon, "Response Effects in Mail Surveys," 240.

offered the possibility of answering "don't know", and are told to tell the researcher whatever they think, even if they are unsure, they are more likely to offer an opinion than to volunteer a "don't know" response.⁷⁷ People therefore can be convinced to offer an opinion by an interviewer. They appear less likely to hesitate in giving an answer when further probing by the interviewer occurs. This, while not cost effective, this does present an effective solution to minimizing the occurrence of "don't know" responses. It also shows that many people do in fact have opinions, but are simply shy to express them if they feel unqualified to do so. Or, respondents will feel they must participate in the "conversation" of the survey interview if forced to do so by the interviewer. The survey interview therefore benefits from increasing the comfort quotient of the respondent, probing more deeply, and reassuring respondents about the importance of their answers. Clearly better data will be a result of these better inquiring practices in survey interviews.

Closed ended and Open-ended Questions

The decision to offer open-ended or closed-ended questions naturally relates to this discussion of "don't know" and "no opinion" alternatives, as well as to the idea of a middle response and scaling. Closed-ended questions may be more efficient at obtaining a variety of responses.⁷⁸ When people are given a list of responses to choose from, they may recall more of their own behaviour than if they had just been asked an open-ended question. Sometimes respondents may go blank when asked an open-ended question. Closed-ended questions will jog their memory, as well as providing examples of acceptable responses, or the idea of what issues the researcher would like to discuss. Fee argues that open-ended questions are more expensive to gather in terms of administrative

77. Morton-Williaras and Sykes, "The Use of Interaction Coding," 116-7.

78. *Ibid.*, 123-5.

interviewer costs, and may not provide enough time for the interviewer to obtain information on several different areas of research.

In addition, coding open-ended questions may be problematic, and more prone to error than would be pre-coded, closed-ended response alternatives.⁷⁹ In this way, closed-ended questions may be more efficient. They still however, face the problems of scaling, response order and middle alternatives, as previously mentioned. Schuman argues that therefore "at least some open-ended questions that allow investigators and readers to appreciate, so far as possible, the frames of reference and alternatives that come most spontaneously to respondents."⁸⁰ As well, in coded answers, opposing sides should be enlisted to create balanced response alternatives that take into account different frames of reference.⁸¹ These measures could help to improve the quality of the survey questions and responses. This could in turn create more realistic, balanced distributions with marginals acceptable to both sides of the debate.

Question Order and Response Category Order

Even if the dilemma of open or closed ended questioning is solved, the ordering of questions in a survey instrument still presents a challenging issue to many researchers. Bradburn advocates that it appears in many cases that question ordering is guided by intuition and tradition than by any real factual knowledge or empirical testing. He further argues that the complexity and length of survey interviewing does not make it practical to vary questions as much as would be desirable to gauge adequately order effects. In addition, even if such effects are measurable, they may only apply to a particular

79. Fee, "Symbols in Survey Questions," 73.

80. Schuman, "Ordinary Questions," 438.

81. *Ibid.*

population or survey, and therefore could not be applied out of such a context. Bradburn's only hope is that a base can be built up over time which indicates the sensitivity of certain types of questions to order effects.⁸² Other researchers concur with these conclusions.

These researchers argue that while it may be understood that a question may affect later responses, the size and direction of the effects are not well understood or quantified. Bickart submits that a carryover or backfire effect occurs because of an order effect. When respondents give answers consistent with previous responses, carryover occurs. On the other hand, backfire is present when respondent replies are inconsistent with previous responses. When these effects are present, Bickart argues, then researchers may inappropriately conclude that opinions have changed, when in fact they have remained stable.⁸³ This analysis by Bickart clearly indicates that respondents are consciously aware of their previous responses, and of the need to be either consistent or inconsistent with them caused by a desire to render information that fulfills what they assume to be the researcher's objectives. It does not however, recognize that some behaviour may simply be random behaviour on the part of the respondent, and is not part of a larger theory of respondent behaviour.

Effects may still occur if respondents do not remember the previous questions in a survey. In fact, order effects can occur for responses because respondents are incapable of remembering a large number of possibilities. Ayidiya presents two concepts: recency and primacy effects. Recency effects refer to the idea that respondents will select the last response category offered, while primacy effects indicate the tendency of the respondent

82. Norman M. Bradburn and William M. Mason, "The Effect of Question Order on Responses," *Journal of Marketing Research* 1 (1964), 61.

83. Barbara A. Bickart, "Carryover and Backfire Effects in Marketing Research," *Journal of Marketing Research* 30 (1990), 54.

to select the first alternative offered. Ayidiya's findings are consistent with that of Schuman and Presser, who state that recency effects occur more frequently than primacy effects in oral interviews. In contrast to the oral interviews, both researchers maintain that primacy effects predominate in mail surveys or interview surveys that use written demonstrations. "Krosnick and Alwin's (1987) cognitive theory of response-order effects proposes that this occurs because the first alternative receives deeper cognitive processing in visual presentations and the last alternative receives deeper processing in oral presentations."⁸⁴ It would appear that respondents have difficulty remembering all of the alternatives when they are read, while the first alternative in a written series will dominate their thoughts. Other interesting findings further support these suggestions.

In particular, Kalton discusses how Matthews (1927) finds that when children are presented with a list of response alternatives, they choose the response that is located above the other choices thirty-four percent more frequently. In addition, they choose the alternative printed to the left of the alternative three percent more frequently. Kalton also explains that Payne (1951) concurs with these findings, where in his study, ideas at the top of the list were favoured six percent more than the middle, and two percent more than the last alternative.⁸⁵ These findings clearly provide the base on which the later research of Schuman and Presser and Ayidiya are founded. While the absolute numbers vary between the studies, clearly the placement of the alternatives in the survey question has an effect on respondent choices.

In addition to the placement of response alternatives, the placement of the actual question within the survey may have an effect on respondent behaviour. Kalton asserts that as an interview progresses, the interviewer develops a greater rapport with the

84. Ayidiya and McClendon, "Response Effects in Mail Surveys," 237.

85. Kalton, Collins and Brook, "Experiments in Wording Opinion Questions," 154.

respondents, and therefore the respondents will become encouraged or motivated to be more open.⁸⁶ This may be less the influence of the question order, than the increasing ease of the respondent. Still, it may help to realize that better and more complete information may be attained toward the end of the survey, and therefore those final questions may be the most reliable and valid.

Nonetheless, other work does exist that focuses on the influence of question order. Strack divides question influence into two categories. The first, assimilation, indicates that the responses to a preceding and subsequent question follow the same direction. The second category, contrast effects, means that the direction (of opinion) between the two questions is opposite.⁸⁷ This may occur regardless of whether respondents are aware of the preceding questions and their responses. Strack believes that the conscious memory of answering a question is not necessary for the question to have exerted influence. Merely having answered the question will affect future responses. If the respondent does recall the previous question, "then they can decide *if* they want to use the information and *how* they want to use it."⁸⁸ This however, does not explain the type of effect that is produced by the question order.

Benton argues that two types of effects are produced by the question order. The first effect interacts with the respondent's willingness to give an opinion, whether positive or negative. The second effect relates to the direction of the opinion, or the balance between positive or negative evaluations.⁸⁹ Part of the explanation for these effects may lie

86. Bradburn and Sudman, "The Current Status of Questionnaire Design." 39.

87. Fritz Strack and Leonard L. Martin, "Thinking, Judging, and Communicating: A Process Account of Context Effect in Attitude Surveys," in *Social Information Processing and Survey Methodology*, eds. Hans-J. Hippler, Norbert Schwarz and Seymour Sudman, (New York: Springer-Verlag, 1987), 24.

88. *Ibid.*, 28.

in the amount of information present in the question. This relates to the idea that the order of the question, such as specific-general order, will be interpreted by the respondent as meaning that the second question does not include the premise of the first question,⁹⁰ as discussed in the case of abortion earlier in this discussion. Other effects occur because of the perceived self-informativeness of the respondent, or when subsequent questions are influenced by the opinions or perceptions of the preceding questions' content.

Alspach and Bishop for example, find that in their study the approval ratings of the governor of Ohio range much higher when the "approval of the governor" question was asked before, rather than after the presidential approval question. Alspach and Bishop conclude that "these results have significant implications for the ordering of questions on sociopolitical objects or figures within a survey interview."⁹¹ In this case, the order effect occurs because of the status of the individuals, how they compare, and how popular they are perceived to be. Alspach and Bishop believe that because of the effects of order on popular approval questions, these questions should be placed in numerous places within the questionnaire, and an average should then be computed. In the least, they say, questions that are subject to order effects should be placed away from questions or concepts that will elicit strong responses in respondents.⁹² The problems of question order are clearly identified, and yet solutions do not appear forthcoming that can be applied to all surveys. Researchers use many methods: multiple measures, buffers, or split-half tests. Each case may require its own particular analysis, whereafter it may need a new and

89. J. Edwin Benton and John L. Daly, "Question Order Effect in a Local Government Survey," *Public Opinion Quarterly* 55 (1991), 640-1.

90. Strack and Martin, "Thinking, Judging, and Communicating," 131.

91. Steven E. Alspach and George F. Bishop, "Question Order Effects of Presidential Approval Ratings on Gubernatorial Approval Ratings: A Research Note," *Social Forces* 69, no. 4 (1991), 1241.

92. *Ibid.*, 1246.

creative way to counteract the effects of question order.

Question Length

Question Length is the last section to be examined in the effects of questions on nonsampling error. The length of questions can potentially impact as much on respondents as the effects of question order and scaling. Bradburn and Sudman advocate that longer questions will stimulate respondent memory, and will allow them more time to contemplate their responses.⁹³ Responses will be more complete, and less likely to be random thoughts that respondents feel obligated to provide if they have only a short period of time. Traditionally, researchers are instructed to pose questions that are brief, and to the point. Yet Laurent asserts in contrast that responses are a function of the question length. When the interviewer speaks for a greater length of time, so does the respondent. Laurent states that longer responses should contain more information, and so therefore may be valuable in survey research.⁹⁴ Longer questions do have value in survey research, and brevity is no longer the key issue.

Laurent further argues that when only half of the questions in a survey interview are lengthened, all of the questions receive responses containing more information. He believes that this occurs because respondents interpret longer questions as an indicator for the need to present more complete and accurate reporting. In addition, longer questions may be seen by the respondent as meaning that the interviewer does not want to rush the respondent, and does not have the need to move on quickly. In that case, the perception of time restraints by the respondent are alleviated, which could have been detrimental to

93. Bradburn and Sudman, "The Current Status of Questionnaire Design," 36.

94. Andre Laurent, "Effects of Question Length on Reporting Behavior in the Survey Interview," *Journal of the American Statistical Association* 67 (1972), 298.

adequate search and information retrieval activities.⁹⁵ Longer answers by the respondents make them feel that they are participating more in the "conversation" of the survey interview. As well, the additional time provided by longer questions may mean that the respondent will recall more information. Longer questions, while more time consuming, could be quite beneficial for survey research. If even adding a few longer questions will result in better, more complete information overall in the survey, it may be important for every researcher to consider their use in surveys. One negative effect of doing so may be that if the questions are too long, fatigue effects will occur, and the respondent will lose interest and provide less complete information. Yet if moderation is used, the addition of some length to some questions could outweigh any potential harmful effects.

Conclusion

This chapter has attempted to discuss some of the problems relating to the effects of questions in surveys on nonsampling error. These problems are often difficult to locate, and problematic to analyze. The context of questions may cause respondents to interpret questions differently, and therefore provide substantially different responses. This can be problematic within a survey, if questions are not answered as intended by the researcher, as well as problematic for comparing the results from the same survey taken at two different time periods. Question wording can clearly affect the marginal distributions. People may see questions with slightly different wordings as focussing on different concepts. Or, the negative-positive nature of the question may invoke different responses, and therefore change the interpretation of public opinion, even though no such change may have occurred. The scaling of the responses, the order of the scales, the offering of a "don't know" or "no opinion" option also affect the survey results. Respondents react to

95. Andre Laurent, "Effects of Question Length on Reporting Behavior in the Survey Interview," 305.

the stimuli provided by the differences in questions, which can result in a greater number of non-responses, and less generalizable data. The debate over open-ended or closed-ended questions is still raging among researchers, as each justifies the use of one as providing more information, and accurate responses. Each of these subsections presents new challenges to survey research, in minimizing the impact of the questionnaire on survey responses. Yet the overall impact of these problems may still remain a mystery, at least insofar as one may generalize across surveys.

Do researchers purposefully manipulate questions so as to produce specific results? This certainly occurs, though one hopes that it is only to gauge the extent of such impact, and not to be used for convincing others of the merits of a particular viewpoint. Given that researchers do not typically provide enough information about their surveys (in non-academic forums) to judge adequately potential sources of influence, we may never know the exact facets and shifts of public opinion. Certainly all survey research should be examined with a cautious eye. This task could be made easier if researchers agreed to provide consistent and uniform information about their question content and analysis. One clear problem with this provision is that consistency between surveys is difficult to achieve, especially across companies and with the passage of time. Nonsampling error must continue to be examined in all survey research. It clearly plays a role in survey results, and the public, government, and researchers, must be aware of its potential impact so that survey results are used wisely and interpreted correctly.

Chapter 3

Statistical Analysis of Nonsampling Error

Statistical Analysis remains vital to the examination of problems relating to nonsampling error. Without hard evidence, theories of influence on surveys will remain in the realm of myth. Strict scrutiny must be paid to the examination of this error, not only to establish its existence, but so that the impact of potential solutions can be judged. While some attention has been paid to measuring the impact of interviewers, little statistical effort has been aimed at the problems relating to question order and wording.

Understanding questions and question order is crucial to the production of valid and reliable empirical data. While a survey may indeed measure a concept, questions the researcher asks may in fact reflect a different concept other than the one intended to be studied. As well, distortion may occur not only because of the intentions of the researcher, but depending on how respondents subconsciously react to, or understand questions. If swaying of survey results occurs, even unintentionally, the overall picture of the survey and of the understanding of the issue may fall prey to the manipulations of interpreters in the media, or to powerful information provider agencies. Therefore, it remains extremely important to examine carefully the impact of questions in an empirical format so as to contain the potential impact of problematic questions on the entire field of survey research.

The data used for this specific examination of survey error were originally made into a computer readable data set by David A. Northrup and Anne E. Oram for the Institute for Social Research in October, 1994.¹ The data relate to both the 1993 Canadian Federal Election, and the 1992 Referendum on the Charlottetown Accord. Since several different

1. David A. Northrup and Anne E. Oram, *The 1993 Canadian Election Study, Incorporating the 1992 Referendum Survey on the Charlottetown Accord*, (North York: Institute for Social Research, York University 1994).

data sets using differing respondents are combined in this data set, only certain variables can be used in combination with each other. For example, some data were collected before the 1993 federal election, after the election, or before the 1992 Charlottetown referendum. For this particular empirical examination of nonsampling error, I have selected variables that span three different sets of data contained in the Northrup and Oram study. It should be noted that the number of respondents decreases over the surveys, though at minimum, 2,530 respondents participated.² All of the data sets include respondents from all provinces. While limitations do exist in the data that can affect the representativeness of respondents in relation to the general population, such as not being able to contact those not at home during certain time periods, this data set is much larger, more complete, and representative of the general population than many other data studies of the same phenomenon. In fact, it is probably the largest data set that takes into account the potential problems of question related error in Canada. While some larger studies have taken place in the United States, they have not necessarily made an effort to explain more than one type of occurrence of nonsampling error, or attempted to isolate error over so many survey questions involving different issues.

This chapter will focus specifically on the statistical examination of data relating to nonsampling survey error. In particular, these data include variables relating to question wording, question order, and interviewer gender. While the data set does focus more on the problems relating to questions, interviewer gender is an important issue in relation to nonsampling survey error. Though the researchers did not gather other information about the interviewers, such as foreign accents and perceived nationality of the interviewer, the gender of the interviewer should be factored in as a potential source of influence in telephone surveys. Unfortunately, the researchers were not consistent in including

2. Northrup and Oram, p. 15.

interviewer gender as a variable in all of the data sets, and one key variable, which will be examined later, would be better understood if the interviewer's gender were known. In addition to interviewer gender and variables relating to question wording and order, the socioeconomic and demographic background of the respondents, as well as their political affiliations, will be taken into account. These factors will help to explain what is causing variation in respondent choices, and if these socio-economic and demographic factors are more influential than nonsampling errors.

The Hypothesis

While this research will encompass many different variables, a general hypothesis can be stated that will apply to differing variable situations. This can occur because though the variables relate to differing topics, they all are meant to be judged against a standard reflecting the impact of question wording or order.

Research Hypothesis Diagram

This hypothesis clearly follows the theoretical underpinnings established in Chapters 1 and 2. The outcome of the survey, or variation of survey responses, will be affected by differing question wording, or if the questions are presented in a different order to respondents. Previous theories concur, stating that the last response in a telephone survey will receive more agreement because of memory effects or other psychological effects, as discussed in previous chapters. Wording may affect respondents by alluding to differing

concepts within a question, which the researcher may not be aiming to measure. Some questions may be affected more than others by this relationship between survey results and question wording or order. Indeed, the more sensitive a question, or the more ambiguous the wording of the question, the more likely it is for nonsampling error to occur.

I hypothesize that survey results will fluctuate because of nonsampling error in this empirical research examination. I also hypothesize that questions involving abortion, women, and health care will be more sensitive to nonsampling error than measures referring to the budget, or economic definitions. This may occur for several reasons. Policies affecting women, health care, or the controversial issue of abortion are very visible. As well, they often have an immediate and direct impact on the lives of many Canadians. Issues of the budget, such as in the case of sacrificing unemployment programs to fight the deficit, are less tangible, and directly affect fewer Canadians in the immediate future. This type of issue is also less tangible in that it does not tend to relate as clearly to specific personal characteristics or events for most Canadians. In addition, economic terms are often confusing, and people may not understand implications behind the survey questions, and not wish to ask for clarification. This is not to say that nonsampling error will disappear. In fact, it could be exacerbated in cases where respondents do not understand the question at all, and are merely repeating the last response. However, this is not likely to be the entire case. I believe that effects will occur for these less sensitive variables, but they will be less apparent and have weaker results in statistical analysis.

The Variables

The socio-economic and demographic variables included in this research are the respondent's gender, age, education, and income. In addition, the variable indicating the respondent's federal political party affiliation, at the time of the survey, is included in the

analysis. These variables are generally included in all social science research studies, and help to explain the background of survey respondents. They indicate if respondents are representative of the many segments of the general population, or if they are unusual, and therefore indicate that interpreters of the survey results should exercise caution in generalizing the results to a larger population. As well, such information can be very useful if it later shows certain segments of the population to be very sensitive to certain measures. For example, those with low incomes may be particularly affected by the perceived socio-economic status of interviewers in a face-to-face interview. If this is known, and substantiated, then future research can be geared to be sensitive to this problem, and correct for any irregularities in the data. In regards to this particular survey, the researchers give no cause for users of their data to be concerned about the respondents or the data. Even if a few anomalies exist in the data, the large number of respondents should offset any irregularities and normalize the sampling distribution.

The Dependent Variables

Among the data available in this grouping of surveys amalgamated by Northrup and Oram, I will examine eight variables that use variations in wording or order to test for nonsampling error. Not all of the variables to be examined necessarily relate to political items. One variable makes reference to abortion, several to budget questions discussing unemployment, inflation, taxes and the deficit, others to health issues, or pensions, and one to senate representation. Of all the questions with order or wording alternatives, I believe the eight I selected to be most likely to show error effects. All of these issues are controversial, and the election and referendum highlight the importance and argumentative nature of the problems at hand. Not only are politicians interested in these issues, but ordinary Canadians have clearly taken an interest in these measures, especially where they have an immediate and direct impact on the individual lives of Canadians, such as in the

case of taxes and health care.

Some questions were closely duplicated between data sets, so the doubles have therefore not been included in this analysis. In addition, some variables relate to the popularity or perceived image of political party leaders, and I believe that in this case it to be more important to study variables less transient or affected by an election. The problems of abortion, taxes, the economy and the senate continue to plague Canadian politics even if an election is not immediately in the future. In addition, in order to make this project manageable, it is necessary to limit the total number of dependent variables to be analyzed. These eight variables appeared particularly compelling, and overall, a manageable workload.

Each of the dependent variables to be examined in this research, as well as the transformations necessary to make the variables usable for this type of analysis, need to be explained. The first variable relates to abortion. Originally, different respondents were offered the question in three formats, with three varying response orders. Respondents were asked to indicate which category best represents their view. The first response read that abortion should never be permitted, the second that abortion should be permitted only after need had been established, and the third that abortion should be a matter of personal choice. I combined the three categories of this variable, so that a new variable, ABQUEST, ranges from most restrictive to least restrictive in approval of abortion services.

The second dependent variable asks the respondent if it is more important to control inflation or reduce unemployment. Again originally there are two versions of the questions where these options were asked in reverse order. For example, in the first, reduce unemployment is given as the first option, and in the second, control inflation is given as the first response category. These variables are again combined into a new variable,

ECQUEST, where the first category is "reduce unemployment", and the second, "control inflation". In addition, as a nominal variable in regression, this variable needs to be "dummied", meaning that only one of the categories is included in the analysis, and the computer considers the second category to be a control or reference group.

The next three dependent variables are very similar to the previous variable. Each consists of two possible alternative responses. The third dependent variable, this time relating to health care, also follows the same format. Its categories are "should not be provided if the person can pay for the services" and "same to everyone". It becomes the new variable, HLHQUEST. The fourth dependent variable relates back to the issues of money and jobs, where its responses indicate that it is more important that the federal government either "reduce unemployment" or "reduce the deficit". This newly combined variable is called EC2QUEST. Unlike the previous three variables, this dependent variable come from the data set created after the 1993 federal election. The previous variables are all located in the data set created before the election. The fifth dependent variable also derives from the post-election data set. It contains the categories, "pensions to all" and "no to above average", indicating to whom Canadians feel public pension money should be distributed. It is now called PENQUEST.

Finally, there are three dependent variables that take into account changing question wording. In the first, called DOCQUEST, respondents are asked about waste in the health care system. This question is modified in four ways by the interviewer adding on the words "for those with low income", "for those with some income", "for doctors" and "for hospitals". One of these four wording alternatives is read to the respondent by the interviewer according to a randomized computer program. In statistical analysis, this variable is treated as a dependent variable with four related wording variables, as order variables have been treated with the previous variables. In regression, these wording

variables must be dummied, and since the four are split into two separate groupings, only two categories are included in the analysis. That is, the first wording variable would be "doctor" and "not doctor" and the second wording variable would be "for those with low income" and "not for those with low income" once dummied. In this case, three non-redundant wording dummy variables could have been constructed, rather than two. This would have made the reference category slightly less broad. However, in the interest in consistency, and in making dummy variables for the other wording or order alternatives, I chose to create only two dummy variables. This is still a correct use of the statistical procedure.

In the second dependent variable involving wording variations, respondents are asked alternately whether Canada could get along without the GST, or without a tax "like the GST". This wording variable again must be dummied. Finally, the last dependent variable concerns whether people approve of a guarantee assuring women seats in the senate, or a guarantee of half the seats in the senate, or if aboriginal people should be guaranteed seats in the senate. These categories were combined to create the new variable SENQUEST. This variable contains the categories, "yes, agree", "no, disagree" and "abolish the senate". This variable is also used in conjunction with a separate wording variable, like the health care waste variable. Unlike all of the previous dependent variables, this question comes from the data set collected before the 1992 Charlottetown referendum. The tax variable originates in the pre-election data set.

Each dependent variable will be examined in combination with other variables from the appropriate data set. The abortion variable will be studied with its order variable, as well as the variables of interviewer gender, respondent gender, age, education, income, and political party affiliation. The unemployment/inflation variable, the health/universality variable, the GST variable, and the health care/waste variable will be examined in

conjunction with their respective order and wording variables as well as the same variables as the previous dependent variable, as they originate from the same data set as the abortion variable. The unemployment/deficit variable, coming from the data set gathered after the 1993 federal election, will be studied with the other variables of order, interviewer's gender, respondent's gender, age, and political party affiliation. The dependent pension/universality variable, with have these same independent variables, and in addition, a variable that indicates how the respondent feels about older Canadians. Finally, the last dependent variable discussing senate representation, will take into account its wording variable, as well as the variables of respondent gender, age, education, employment status, income, and political party affiliation.

The eight dependent variables used in this statistical research will be examined using the same five statistical procedures. Each will be thoroughly examined so as to determine if question wording, question order, interviewer gender, respondent gender or other socio-economic demographic factors are having an impact on the variable, or explain most of the variation within variable responses. The procedures of frequency analysis, crosstabs, multiple and logistic regression, and discriminant analysis appear to be best suited to the examination of these dependent variables. Each test is able to work with a data set of this size, and tell the researcher useful information about the variables. In this way, nonrandom error will begin to be exposed, explained, and possibly lead to solving the problem.

Statistical Analysis — Frequency Distribution Analysis

Before beginning the more complex statistical commands, each dependent variable should be run through a frequency analysis. This test tells the researcher whether a not a variable contains too many missing cases to be usable, is generally an abnormal

distribution, or appears to be useful for analysis. All of the variables I chose for use as dependent variables in this project appear to be acceptable. Some preliminary analysis can be gained from a comparison of variables across the frequency distributions. This occurs because in this procedure, the raw variables are included before I combine them, or dummy them, which must be done before more complex statistical commands can be run. In this case, comparison of the raw variables contains some interesting information.

In the variables relating to the issue of abortion, it appears that the last choice offered does gain respondent agreement for the categories of "never" and "permit = need". This clearly relates to the theories of recency effects, where it appears that the respondent agrees with the category listed last. Previous categories in this case may have been forgotten because of the complexity, wording, or that the respondent simply paid more attention to the question at the end. The category of "personal choice" does best in the middle, therefore not following the theories of recency or primacy effects. Perhaps those committed to personal choice see it as the best compromise (hence middle) solution. They may be eager to select the middle of the road response, and not want to appear at an extreme. This variable therefore initially does appear to have some question order effects.

In the next dependent series variable, dealing with unemployment and inflation, the "control inflation" category appears to suffer more effects than the category of "reduce unemployment". This may occur for several reasons. Inflation may not be fully understood as a concept, and so therefore, it may suffer order effects when it is suggested last, as more recent in the memory of respondents. In addition, unemployment is a chronic problem, and though important, a new or more recently suggested problem, such as inflation, may appear more significant to the respondent. This variable clearly does not appear to have effects as strong as the abortion variable, though it may have more minor fluctuations because of order. It will likely have much weaker statistical results in other

analyses, though it should not be discounted entirely.

The next variable, which concerns the universality of health care, suffers large effects on both categories of "not provide can pay" and "same to everybody". Health care and the rising cost of universality are very topical and well argued issues in current Canadian politics. Though each side of the debate may be crystallized, the opinions of ordinary Canadians may not be set in cement. They may be open to the suggestion of importance as indicated by response order. This question can have serious implications for the debate. If politicians rely on survey results to gauge public opinion, and the survey is flawed or influenced by order effects, then a real breach in national trust and approval can occur. This question therefore merits much more serious consideration in further tests.

The next variable concerns reducing the deficit, or reducing unemployment as a trade-off. Interestingly enough, the order effect appears largest for the variable in the first response suggested. That is, more people agree with that the first category should be dealt with than the second in both cases. It appears that the first response captures the attention of the respondent so much that the second response is discounted in both cases. This is important not only because it is the reverse of what has occurred thus far with the other variables, but also because the unemployment response does suffer effects, where it did not in the case of the second dependent variable analyzed. The results for this variable that discusses universality and health care, may therefore be stronger than those for the second variable that involves the issues of unemployment and inflation. It should be noted that this variable referring to health care comes from the post-election data set, and the second variables comes from the pre-election data set. Perhaps the Liberal emphasis on jobs pushed the issue of unemployment to the forefront, and therefore made it appear more important to Canadians.

Like the health care variables, the variables concerning pensions also suffer large order effects for both responses. Pensions, which are also affected by attacks against universality, are equally controversial at a time of budget cuts. In addition, the Liberals promised to scale back the federal M.P. pensions, which appeared far too generous against private industry pensions. This is also relates to the controversy over universality, and therefore may be subject to the power of order and suggestion in surveys. Canadians have become accustomed to universality, and though some wish to abolish it, the debate is difficult, and open to contradiction.

The tax issue involving the GST is a straightforward variable that will be better explained in Crosstabs. The next variable, dealing with health care and waste with four wording alternatives also shows significant effects for the different response categories. Clearly health care is not only an important issue, but potentially subject to nonsampling survey error. Finally, the last dependent variable concerning women, Aboriginal peoples and senate seats may suffer wording effects. More people disagree with allocating senate seats to women when "half of all senate seats" is specified. Fewer people disagree with allocating senate seats to Aboriginal peoples, though much fewer people agree overall. Perhaps people are willing to recognize the disparity in Aboriginal representation in government, though are also less willing to go out of their way to reconcile these problems. In addition, the idea of giving "half" of all senate seats to women may appear too extreme for many Canadians, though an unassigned number of guaranteed senate seats is not entirely objectionable. This clearly merits further examination in the study of question wording.

The initial frequency distribution analyses do yield some important results. They indicate that for some variables, order and wording effects are having an impact on this specific set of respondents. It remains to be seen whether this can be generalized to the

population at large. Nonetheless, more complex statistical results should be run to explore the problems that frequency distribution analysis has begun to uncover.

Statistical Analysis — Crosstabs

Since Frequency Distribution Analysis has already demonstrated movement between response groups, crosstabs will reveal the strength of the measure of association between the dependent variable, and the order or wording variable. This order or wording variable takes into account both versions of the question. When used in combination with the actual dependent variable, which in this case is an amalgamation of all of the different original variables with the response categories consistently grouped, crosstabs should show how strong the relationship is between order and the response categories. The frequencies distributions did not show the strength of the relationship; only preliminary analysis made it appear that relationships potentially exist because of the shift of respondents from one group to another as the order or wording of a question changes. The crosstabs should also indicate whether this relationship can be generalized to apply to the larger Canadian population with the Chi-Square Significance test. This is key to statistical analysis, as if the relationship does not apply to the general population, then any findings mean that this particular data set may be an anomaly, and cannot be reliably duplicated so as to solve problems of question wording or order in other survey research.

The list of all the dependent variables, ABQUEST, which relates to abortion, is an ordinal variable. Its categories range from most restrictive to least restrictive in providing abortion services to women. When used in conjunction with its order variable, it does not appear that a strong relationship exists. Cramer's V shows there to exist a weak relationship between the two variables. This relationship however, is significant, showing a Chi-square of less than 0.05, so it does apply to the general population. This is important

in that even if the impact of the order variable is small, it is important for all research in general on this topic containing an order variable. Other uses of this topic, and of ordered variable survey questions, may have an even greater impact on survey results, and therefore could pose greater potential nonsampling error problems.

The second crosstabs between the unemployment/inflation dependent variable and the order variable also proves to be significant. It is stronger than the measure of association for the previous variable combination, though still very weak at 0.057. This is not entirely unexpected because the frequency distribution did not show a large variation between responses when the order was reversed. Significance can be achieved in some part in virtually any relationship when a large enough data set is used. Nonetheless, given the importance of examining this particular problem, the significance, and the weak but evident measure of association, may be cause for continued study.

The next crosstabs, involving the dependent variable relating to health care and universality displays a significant Chi-Square and a stronger measure of association than the last variable combination when paired with its order variable. Cramer's V indicates a strength of 0.14876 on a scale of zero to one. While still weak according to most standards of measures of associations, this result is the strongest yet among the different variable combinations. True to the frequency distribution, this variable has demonstrated large movement between responses and therefore, a stronger measure of association result in a crosstabs test is not completely unexpected.

The third crosstabs involves the dependent variable relating to unemployment and the deficit and its order variable. This variable originates from the data collected after the election, in contrast to the previous dependent variables that derive from data gathered before the election. This combination of variables does not prove to be significant. As

2 of/de 2

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

TABLE 1. Crosstabs Results: Measures of Association and Significance for Dependent Variables and Order/Wording Variables

Variable	Cramer's V	Chi Square Significance
Abortion Variable by Order Variable	0.05919	0.00088 ^a
UI/Inflation Variable by Order Variable	0.05707	0.00644*
Health/Pay Variable by Order Variable	0.14876	0.00000*
UI/Deficit Variable by Order Variable	0.03393	0.27856
Pension Variable by Order Variable	0.11884	0.00000*
Health Care Waste/Pay by Low Income/Some Income (Order Variable); by Doctor/Hospital (Order Variable)	0.01283 0.01690	0.819153 0.78224
GST by Question Wording	0.08520	0.00000*
Women/Aboriginal Senate Variable by Question Wording	0.13345	0.00000*

a. An * denotes significance below the 0.05 level.

well, Cramer's V produces fairly weak results. This insignificant relationship could be reflective on non-normal behaviour in the population. These results can not therefore apply to the general population, and should be interpreted to be meaningful only in the context of this survey.

The fifth crosstabs, derived from the same post-election data set, refers to pensions and universality. Unlike the previous crosstabs, this relationship proves to be significant and somewhat stronger than most of the other crosstabs Cramer's V results at 0.11884. In an analysis of the frequency distributions of this variable, strong changes appear to occur between categories with a change in question order. Therefore, like the health variable, this should not be completely unexpected. These questions seem to be vulnerable to the problems of recency effects, appearing through the question order test. Either these issues are more volatile and vulnerable, or the overreaching issue of universality is driving the instability in these two variables.

The sixth and seventh crosstabs are related in that in each case the dependent variable consists of the "less waste" and "not get health care" response options. In the first case, this variable, DOCQUEST is paired with the wording variable that contains the options "doctor" and "hospital". The second wording variable contains the options "low income" and "some income". This means that when the computer indicated for the interviewer to read the first word order, she added on the words, "for low income individuals", for example. Using this method, the question was slightly altered in four different ways. For both of these crosstabs, the results were completely insignificant, and Cramer's V produces very weak results. In this case therefore, the results can not apply to the general population and are not very strong even in comparison to other results from this data set. Question wording has not in this test made a very great impact on the dependent variable.

The next crosstabs measures the strength of the relationship between the question referring to the GST and its alternate wording. Respondents in this case were alternately given the question, "Could Canada get along without the GST?", and "Could Canada get along without a tax like the GST?". These two wording alternatives combined with the agreement variable produce a significant but weak result. This significant result contrasts with the insignificant result of the crosstabs which involved a word experiment dependent variable. The result is slightly stronger than this last crosstabs test as well, though still weak according to a general measure of association standard.

The last crosstabs involves the variables that experiments with question wording concerning women, aboriginals, and senate representation. This variable had clearly shown changes in response patterns in the frequency distributions. In crosstabs, this dependent variable is combined with its three category wording variable. The Pearson Chi-Square indicates a significant result, and Cramer's V shows a stronger result in comparison to other results from variable combinations in this research, with a figure of 0.18873. This data was collected before the 1992 Charlottetown referendum, when the senate and issues of representation were at the forefront of Canadian politics. Considering the importance of the issue, and the possibility of misconstruing the results of surveys on this issue, the question of wording should definitely be further investigated in this case.

Overall, these crosstabs have generally shown significant but weak results. Though significance can occur in part because of the large size of the data set, I believe it still to be important in the search for the impact of question wording and order on responses in surveys. Without significance, these results could be discounted as not applicable to the general population. Nonsampling error could therefore be deemed not to exist or impact on other survey results, when it really can be a large factor in the interpretation of survey results. When issues play a major role in policy formation, it becomes of paramount

importance to understand true Canadian public opinion without the distortion of nonsampling survey error. Therefore, I believe these crosstabs to be the beginning of the investigation of nonsampling error. This research proves that even if small, nonsampling error does exist in survey research. Weaker results are not a downfall in this investigation, but instead the beginning of the recognition and correction of nonsampling error.

Statistical Analysis — Multiple Regression

Multiple Regression is the third statistical test that I use to investigate the problem of nonsampling error in this research. In order to make the dichotomous dependent variables usable for multiple regression, it is necessary to "dummy" them. This procedure involves creating separate variables for each category in a nominal variable. For example, in a variable where the responses are "yes" and "no", there are now two variables; one representing "yes" and the absence of "yes" (or all other categories in the variable), and a variable representing "no" and the absence of "no". Then, only one of these variables is included in the analysis as a dependent variable, because it would be redundant to analyze both of the dummied variables. In this way, while some of the assumptions surrounding multiple regression are violated by using nominal and dichotomous variables, the statistical output is still useful as an explanatory tool, and collinearity is avoided.

Once again, I performed a multiple regression for each dependent variable, for a total of eight. For the first regression that involves the abortion issue, the dependent variable is not manipulated. For the following seven regressions, the dependent variable is dummied. The last variable posed a bit of a dilemma. It consists of the categories of "yes", "no", and "abolish the senate". In one sense, this can be seen as a range of agreement to disagreement, with "abolish the senate" just a very extreme form of "no". Or, this variable can be seen as a nominal variable, where the categories have no absolute ordering; in this

way, "abolish the senate" is not exactly a form of "no", but instead a distinct category that can not be placed in an ordering scheme. I decided that since I had dummied all of the other variables with a "yes/no" format, I would be consistent and dummy this variable as well for regression analysis. The computer will therefore use only two categories as dependent variables, and use the third as a control group.

Regression can be thought of as an equation where the dependent variable is being explained by the addition of independent variables and statistical error. Using standardized regression coefficients, the regression equation can be stated as:

$$Y = B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + B_5X_5 + B_6X_6 + B_7X_7 + \dots + e$$

where Y = the dependent variable

B_1X_1 = order/wording variable (dummied)

B_2X_2 = interviewer's gender

B_3X_3 = respondent's gender

B_4X_4 = respondent's age

B_5X_5 = respondent's income

B_6X_6 = respondent's highest level of education attained

B_7X_7 = political party affiliation (each party is dummied, so therefore, this variable could represent "liberal" for example. Then each new party would have a new Beta Weight.)

e = error

The variables that represent the different wording or question order are dummied. In addition, nominal variables such as political party affiliation, and interviewer's gender,

must be dummied. This ensures that the variables will be treated correctly by the computer. Though using dichotomous or nominal variables does violate some of the assumptions surrounding multiple regression, the results will still be meaningful in a statistical research sense.

The regression equation changes depending on the number of variables included. The sign before the Beta Weights will be positive or negative depending on the standardized Beta Weight itself. As the actual weight or strength of the coefficient changes, so does its influence on the dependent variable. Finally, a term representing error must always be included in a regression equation. Regression will help to explain the error present in the analysis. Unfortunately, the explanatory power of regression in terms of error may be somewhat limited because of the inclusion of dichotomous, nominal variables. Since the inclusion of these variables violates some of the assumptions of regression, it may interfere with the error related statistical tests. This remains to be seen as the regression analyses are examined.

Each regression analysis produces several statistical tests. The first part of the regression shows a correlation matrix, similar in nature to the measures of association from crosstabs. I will highlight the stronger relationships shown by the correlation matrix for each regression analysis. This is more than what crosstabs produces, because it combines all possible variables in the matrix. Therefore, combinations are examined that were not produced in crosstabs. Though the most important relationship in the correlation matrix is that between the dependent variable(s) and the order or wording variable(s), the relationships between the independent variables will also be discussed. While this is generally viewed as a test for collinearity, I believe that they should be mentioned, especially in cases where there does not exist any strong relationships between the dependent variable and order/wording variables, or between these variables and other

independent variables.

The next step is the entering of the independent variables into the equations. Generally, there has been a three step entering process of variables; on the first, only the word/order variable is entered, on the second step, the interviewer's gender has been entered, and on the third step, all of the other variables are entered. For each step, an R-square and an F-test or significance test are produced. The R-square measure indicates the percentage of variation in the dependent variable explained by the independent variables entered on that step. In addition, the actual standardized Beta weight coefficients are calculated and displayed, each with its own significance calculation. In the tables included in this analysis, some correlation matrix results, as well as the standardized regression coefficients and R-square results from the last step have been enumerated. Next in the regression analysis output is a normal probability plot, which gauges the normality of the error distribution for the equation against a perfect example, and finally, a standardized scatterplot, which provides insight into the error included in the regression. Each of these statistical tests will be reported and explained for the eight regression analyses.

The first regression analysis uses the abortion variable as the dependent variable. Also in the analysis are the order variable, the respondent's gender, the interviewer's gender, the respondent's age, income and education, and political party affiliation, which has been dummied, so that Liberal, PC, Reform, and NDP are included and Bloc has been left out as a control group. This dummied of nominal variables, such as political party affiliation, occurs in each regression analysis. While the important relationships in the correlation matrix are those between the dependent and independent variables, relationships between independent variables will also be discussed. This is a test for collinearity. In some cases, the relationships between the independent variables are the only strong and significant relationships in the regression. The correlation matrix for the

first regression reveals strong and significant results between two of the order variables, Liberal and PC, income and the respondent's gender, and age and education. The order variables are strongly related because of the nature of both variables which take into account the order in which the questions were asked. Neither of the order variables is linked strongly or significantly with the dependent variable. The Liberals and PCs may be related because of the established nature of these two political parties, and that they both allow their members to choose freely their opinions on the abortion debate. They appear related as well because this is a function of the party choices available. It comes as no great surprise that income and gender are related. Women still earn less than men, and the demographics of this survey may reflect that inequity. Age and education are naturally related because of the educational system in Canada, and the advancement of people through the system. All of these relationships are significant, and can therefore be generalized to the larger Canadian population. Unfortunately, as in crosstabs, no strong relationship is evident between the dependent variable and the order variable, or in this case, the other variables included in the equation.

The next test confirms that the dependent variable is not best explained by the independent variable selected in this instance. The R-square measure yields a result of only 0.0002 when the order variables are in the equation. When the interviewer's gender variable is added, the R-square rises to only 0.00003. In both of these measures, the F-test does not show significant results. When all of the other variables, including income, education, age, the respondent's gender, and political party affiliation, are included, the R-square rises to 0.01969. This indicates that about two percent of variation of the dependent variable expressing the level of abortion support, is explained by all of these variables. This is significant, with an F-test result of 0.0000. Of the variables included, the strongest standardized Beta Weights that are also significant are income, age, Reform (dummyed

political party variable), education, and Progressive Conservative. All of these however, are fairly weak, below the 0.1 range. The normal probability plot shows that these cases do not conform well to the normal distribution. In addition, the standardized scatterplot shows strong vertical clusters of cases. Error problems may exist according to this test. This could be a result of links between the independent variables. The scatterplot could also be the product of having included nominal variables in the regression. This scatterplot indicates a non-normal error distribution because of socio-economic and political dichotomies.

These tests appear to indicate that the control variables are having the greatest effect on the dependent variable. The order variable and the interviewer's gender variable do not seem to make much of a difference for the dependent variable. Therefore, in this regression analysis, it can be determined that the abortion variable is not highly affected by question order, or by the interviewer's gender. Instead, socio-demographic and political background characteristics of the respondents are more influential in determining respondent choices. Nonetheless, even these variables are quite weak. Yet having very weak influence by the order variable contradicts some of the indicators of previous tests, which suggest that order may be having an impact on this dependent variable, even if it is very slight. The next few statistical tests may reveal more information about this problem.

The second regression considers the variable measuring the issues of unemployment and inflation. In its correlation matrix, stronger relationships are revealed for the variables, PC by Liberal, the respondent's gender by income, Reform by Liberal and age by education. Again, no strong relationships appear for the dependent variable, or for the order variable, which are the most important among the variables included in this regression. All of these results are significant below the 0.05 level. Clearly, as in the previous regression, political parties appear to have links between them. Indeed, the

results for Party relationships are stronger than all other results. Age and education again are linked as they were in the last regression. As well, this regression appears to reflect the same income disparity by sex. Considering that this uses the same data set, it does not come as a complete surprise that these relationships emerge in both regressions.

The order variable explains slightly more than half of one percent of the dependent variable, with an R-square of 0.00620. This calculation is significant. The Beta weight for the order variable is also significant at this stage. When the interviewer's gender is factored in, the R-square rises, but only to 0.00648. The significance also rises from 0.0008 to 0.0028. This is still however, statistically significant and far below the 0.05 tolerance level for significance. When all of the variables are entered on the next step, the R-square result makes a huge leap to 0.02407, and the significance drops to 0.00000. This appears to indicate that the interviewer's gender is having little effect compared to the rest of the independent variables, which seem to have the greatest impact on the dependent variable. In addition to the order variable, two political parties, PC and NDP, and the respondent's gender are significant. None of these Beta weights though, are very strong. The normal probability plot does stray quite far from the normal line, with the farthest outlier at 1.83257 points away from the diagonal. The standardized scatterplot also shows a non-normal distribution of error. Once again, this could be caused by the violation of regression assumptions with the inclusion of nominal data, or it could occur because of links between the independent variables. Considering the strength of the independent variables in the data set in explaining the dependent variable, and strong relations among each other in the regression analysis of some of the dependent variable examinations, this could be true.

Order appears to be having a minimal but significant impact on the ECQUEST dependent variable, as compared to the abortion dependent variable regression, where it

had no significant impact. This is especially interesting because in the previous regression involving the abortion variable, and in the frequency distribution analysis, it appeared that order made a large impact on the dependent variable. Nonetheless, in the regression output for the last two variables, the order variable showed little impact on the dependent variable. In this case, order does not appear to be having a great impact on the dependent variable in either regression test. The order variable has a significant impact on one variable, but not on the other. It also does not have a huge impact on either dependent variable in terms of the absolute strength of the relationship. Therefore, the first test may or may not be a good indication of the presence of nonsampling error. While in both cases the percentages and absolute figures for Beta Weights are very small, the comparison, and the significance of the results may have further implications for conclusions in this research.

In the third regression, the dependent variable concerns health care and universality. Interestingly, the dependent variable and the order variable appear to have a significant, though moderately weak relationship. Still, considering the weak figures that have been produced throughout this research, this figure of -0.132 is fairly high. Other strong relationships appear between gender and income, Liberal and PC, Liberal and Reform, and age and education. This is consistent with the results of the last two regressions, whose variables also originate from the same data set. The order variable produces the highest R-square result for a single variable yet, at 0.01750 . Almost two percent of the dependent variable has been explained by the order variable. This even competes with the last regression, which only attained an R-square result of this magnitude with all of the variables added in to the equation. This result is significant. The inclusion of the interviewer's gender raises the R-square to 0.01889 , and the significance remains the same at 0.00000 . When all of the variables are taken into account, the R-

square rises above two percent, to 0.02396, with a perfect significance.

Only two variables have strong, significant Beta weight coefficients. The order variable and NDP, have the largest impact on the dependent variable. It is interesting to note that none of the independent variables of socio-economic characteristics are significant or extremely strong in this regression, even though this regression uses the same data set as the previous regressions. The normal probability plot and the standardized scatterplot suffer the same problems as in the other regression outputs. While this may indicate problems, the lack of strength in the independent variables in this regression may mean that the error is non-normal because of the error assumption violations, and not because of more serious problems. This regression remains interesting because it does not engage the strength of independent variables, and the order variable is immediately an important presence in comparison both to the other independent variables, and the order variables of the other regressions. This variable combination also produces interesting results in the previous tests, and remains consistent in producing strong results in comparison to the other variables' statistical tests.

The regression involving the issues of unemployment and the deficit is derived from variables generated from a different data set than the previous regressions. While stronger and significant relationships appear between some of the same independent variables, such as age and education, gender and income, and between political parties, the dependent variable appears to be linked with other variables, in this case, gender and NDP. These results in the correlation matrix are moderately strong in comparison to other results. This correlation matrix may show that fewer women indicate increasing support for the NDP. Women may have increasing support for the Liberals, which would explain

TABLE 2. Multiple Regression: Correlation Matrix Results

Regression (Dependent Variable)	Variables	Correlation	Significance
ABQUEST (Abortion)	Order Variable (1) by Order Variable (2)	- 0.474	0.000* ^d
	PC by Liberal	0.295	0.000*
	Income by Respondent's Gender	0.168	0.000*
	Age by Education	0.289	0.000*
ECQUEST (Unemployment/ Inflation)	PC by Liberal	- 0.302	0.000*
	Respondent's Gender by Income	- 0.155	0.000*
	Reform by Liberal	- 0.151	0.000*
	Age by Education	0.170	0.000*
HLHQUEST (Health/ Universality)	Dependent Variable by Order Variable	- 0.132	0.000*
	PC by Liberal	- 0.297	0.000*
	Liberal by NDP	- 0.166	0.000*
	Respondent's Gender by Income	0.173	0.000*
	Age by Education	0.288	0.000*
EC2QUEST (Unemployment/ Deficit)	Dependent Variable by Respondent's Gender	- 0.139	0.010*
	Dependent Variable by NDP	0.153	0.005*
	Interviewer's Gender by NDP	- 0.106	0.038*
	Respondent's Gender by Liberal	0.119	0.023*
	Liberal by PC	- 0.358	0.000*
	Liberal by NDP	- 0.374	0.000*
	Liberal by Reform	- 0.389	0.000*
PENQUEST (Pensions/ Universality)	Dependent Variable by Order Variable	0.190	0.001*

TABLE 2. Multiple Regression: Correlation Matrix Results (Continued)

Regression (Dependent Variable)	Variables	Correlation	Significance
PENQUEST (Continued)	Dependent Variable by Reform	- 0.141	0.008*
	Liberal by PC	- 0.376	0.000*
	Liberal by NDP	- 0.376	0.000*
	Liberal by Reform	- 0.386	0.000*
TAX (GST question)	Dependent Variable by Wording Variable	0.107	0.000*
	Dependent Variable by PC	- 0.167	0.000*
	Liberal by PC	- 0.300	0.000*
	Respondent's Gender by Income	- 0.169	0.000*
	Age by Education	0.239	0.000*
	Age by Income	0.107	0.000*
WASQUEST (Waste/ Health Care)	Dependent Variable by PC	0.103	0.000*
	Liberal by PC	- 0.299	0.000*
	Liberal by NDP	- 0.166	0.000*
	Liberal by Reform	- 0.149	0.000*
	Dependent Variable by Reform	0.106	0.000*
	Respondent's Gender by Income	- 0.167	0.000*
	Age by Education	0.314	0.000*
SENQUEST (Senate Representation)	Wording Variable (1) by Wording Variable (2)	- 0.968	0.000*
	Dependent Variable by Wording Variable	- 0.150	0.000*
	Dependent Variable by Respondent's Gender	- 0.162	0.000*
	Liberal by PC	- 0.263	0.000*
	Liberal by NDP	- 0.194	0.000*
	Respondent's Gender by Employment Status	- 0.227	0.000*

TABLE 2. Multiple Regression: Correlation Matrix Results (Continued)

Regression (Dependent Variable)	Variables	Correlation	Significance
SENQUEST (Continued)	Respondent's Gender by Income	- 0.104	0.000*

a. An * denotes significance below the 0.05 level.

TABLE 3. Multiple Regression: Standardized Beta Weights

Regression (Dependent Variable)	Variable	Beta Weight	Significance	R square	Signif F
ABQUEST (Abortion)	Order Variable (1)	- 0.002954	0.9054	0.01969	0.0000**a
	Order Variable (2)	- 0.009895	0.6911		
	Income	- 0.058912	0.0086*		
	Education	- 0.043641	0.0014*		
	Respondent's Gender	0.031554	0.0289*		
	Reform	- 0.043641	0.0554		
	PC	0.052311	0.0289*		
	Age	0.052332	0.0229*		
	Interviewer's Gender	- 0.001314	0.9522		
	NDP	0.033695	0.1404		
	Liberal	- 0.32759	0.1757		
ECQUEST (Unemploy- ment/Inflation)	Order Variable	- 0.079839	0.0006*	0.02407	0.0000*
	PC	0.065173	0.0107*		
	Respondent's Gender	0.075560	0.0015*		
	NDP	0.044595	0.0680*		
	Interviewer's Gender	0.021991	0.34464		
	Age	- 0.036443	0.1245		

TABLE 3. Multiple Regression: Standardized Beta Weights (Continued)

Regression (Dependent Variable)	Variable	Beta Weight	Significance	R square	Signif F
ECQUEST (Continued)	Reform	- 0.024426	0.3138		
	Education	0.013443	0.5701		
	Income	- 1.009614	0.6855		
	Liberal	0.027264	0.2446		
HLHQUEST (Health/ Universality)	Order Variable	- 0.132747	0.0000*	0.02396	0.0000*
	NDP	- 0.050807	0.0283*		
	Education	0.004300	0.8501		
	Interviewer's Gender	0.034764	0.1175		
	Income	- 0.033533	0.1401		
	PC	0.014191	0.5581		
	Reform	0.023097	0.3168		
	Respondent's Gender	- 0.013859	0.5397		
	Age	- 0.004521	0.8432		
	Liberal	- 0.018324	0.4548		
EC2QUEST (Defici/ Unemploy- ment)	Order Variable	0.025758	0.6618	0.06761	0.0143*
	Respondent's Gender	- 0.149769	0.0122*		
	NDP	0.179575	0.0398*		
	Interviewer's Gender	0.078007	0.1954		
	Age	- 0.020696	0.7263		
	PC	- 0.060530	0.4744		
	Reform	- 0.011284	0.8983		
	Liberal	0.105750	0.2998		
PENQUEST (Pensions/ Universality)	Order Variable	- 0.172194	0.0035*	0.06956	0.0154*
	Interviewer's Gender	0.088096	0.1371		

TABLE 3. Multiple Regression: Standardized Beta Weights (Continued)

Regression (Dependent Variable)	Variable	Beta Weight	Significance	R square	Signif F
PENQUEST (Continued)	Feel About Old People	- 0.015687	0.7908	0.05156	0.0000*
	Respondent's Gender	- 0.011360	0.8455		
	PC	- 0.004362	0.9596		
	Age	- 0.040396	0.4909		
	NDP	0.058504	0.5049		
	Reform	- 0.077565	0.3778		
	Liberal	0.10061	0.3328		
TAX (GST Question)	Wording Variable	0.109649	0.0000*	0.04410	0.0000*
	Age	- 0.051154	0.0241*		
	PC	- 0.188500	0.0000*		
	Respondent's Gender	- 0.054452	0.0152*		
	Liberal	- 0.068864	0.0047*		
	Interviewer's Gender	0.029518	0.1807		
	Income	0.015497	0.4911		
	Education	- 0.051154	0.0241*		
	Reform	- 0.039104	0.0883		
	NDP	9.022E-04	0.9688		
WASQUEST (Waste/ Health Care)	Wording Variable (1)	0.021745	0.2696	0.04410	0.0000*
	Wording Variable (2)	0.021721	0.3101		
	Interviewer's Gender	0.023425	0.0081*		
	PC	0.105787	0.0000*		
	Income	- 0.050795	0.0238*		
	NDP	- 0.068300	0.0028*		
	Reform	0.108723	0.0000*		

TABLE 3. Multiple Regression: Standardized Beta Weights (Continued)

Regression (Dependent Variable)	Variable	Beta Weight	Significance	R square	Signif F
WASQUEST (Continued)	Respondent's Gender	0.073094	0.0011*		
	Education	- 0.001508	0.9479		
	Respondent's Gender	0.073094	0.0011*		
	Age	- 0.004607	0.8426		
	Liberal	0.016825	0.4876		
SENQUEST (Senate Representation)	Wording Variable (1)	- 0.091084	0.0002	0.08556	0.0000*
	Wording Variable (2)	- 0.204216	0.0000*		
	Education	0.085751	0.0000*		
	Respondent's Gender	0.129291	0.0000*		
	PC	0.127605	0.0000*		
	Reform	- 0.078717	0.0001*		
	Income	0.019537	0.3313		
	NDP	- 0.032222	0.1183		
	Age	- 0.034222	0.0903		
	Employment Status	0.063786	0.0019*		
	Liberal	- 0.047217	0.0271*		

a. An * denotes significance below the 0.05 level.

the shift away from the NDP. In reference to the relationship between the dependent variable and the NDP, the NDP have argued that unemployment must be tackled, and only incorporated concerns about the deficit in the province of Ontario, once they were elected. Women too may be concerned more about unemployment, as they earn less, and face underemployment more in Canada. This may be related to the fact that there exists a high concentration of women in public sector jobs that have become targets of deficit reduction.

The R-square result for just the order variable in this case is insignificant and poor at 0.00068. It rises considerably when the gender of the interviewer is taken into account, to 0.00404, and though significance lowers from 0.6651 to 0.5720, it still does not meet the upper limit of significance tests. When all of the other variables are added into the equation, including age, gender of the respondent, and political party affiliation, the R-square rises dramatically to indicate that over six and one half percent of variation in the dependent variable is explained by these independent variables. In addition, this figure is significant at 0.0143. Only two variables have strong Beta weights, and are significant. Again, the respondent's gender and NDP take the lead. These variables are therefore stronger in this test, and confirm the correlation matrix findings. The normal probability plot does not conform to normal, with the furthest outlier 2.02237 points away from the diagonal. The standardized scatterplot contains diagonal clumpings, also indicating non-normality in this test, which may be reflective of the same problems of assumption violation, or inter-variable relationships. In all, this variable, appears to suffer some of the same problems that occur with the abortion variable in regression. In this case, the independent variables other than order appear to explain more of the variation in the dependent variable.

The next regression engages the last dependent variable that involves a variation in question order. It is derived from the same data set, gathered after the 1993 election, as the

most recent regression explained. It involves the issues of pensions and universality, and like that of health and universality, indicated a large impact on the variable by order variation of responses in the frequency distribution analysis. Like the health variable regression, this dependent variable also appears linked with the order variable in the correlation matrix. In the correlation matrix it is also linked with the Reform party dummied variable. The other variables indicating stronger and significant relationships in the matrix include the political party variables. As in the last regression using this data, the socio-economic independent variables do not appear to be strongly linked together, or to be largely influential in the regression thus far.

The order variable in this regression explains more of the dependent variable than ever occurred in the previous regressions. The R-Square result is significant, and very strong at 0.03599. The addition of the gender of the interviewer variable causes the R-square to rise to 0.04439, and it remains significant. In this regression, I included an additional step addition, where I added in the variable "rating> how feel about old people". I believe that this variable could have an impact since the issue of pensions clearly relates to older Canadians. This variable however, does not appear to explain much more of the variation in the dependent variable, as the R-square result rises only to 0.04580. This does remain significant. Finally, when the remaining variables of the respondent's gender, age and political party affiliation are included, the R-square rises astonishingly to 0.06956. This result is also significant at 0.0154. Clearly the independent variables are making a difference of three percent in explaining the dependent variable. Nonetheless, this is not as important as the fact that the order variable initially explains over three percent of the dependent variable's variation on its own; a dramatic increase over all other order variable results.

Unfortunately, only one variable's Beta weight is significant in this regression. The

order variable has a weight of -0.172194, and is significant at 0.0035. Though other cases have had more variables significant, and comparatively stronger, since this is the order variable, this is still an important finding. The normal probability plot and standardized scatterplot suffer the same problems as in the other regressions. Nonetheless, despite any potential problems, clearly in the case of pensions and universality, the importance of question order can not be dismissed according to this regression, or the other statistical tests previously performed. This variable should definitely be further scrutinized with more complex statistical tests.

The next regression involves the variable that deals with the changing in wording concerning the GST. This variable comes from the data set created before the 1993 election, like the first regressions in this series. The correlation matrix shows there to exist a significant relationship between the dependent variable and the wording variable. In addition, the dependent variable appears to be related to the political party PC variable. As in the other cases involving this data set, income and gender, age and education are related, as well as the political parties.

The wording variable, when included on its own step in the regression, explains just over one percent of the dependent variable, as indicated by the R-square measure. This result is significant. When the interviewer's gender is added in to the regression, the R-square rises only one tenth of one percent, though it is still significant. As in the other cases from this data set, the interviewer's gender does not appear to play an important role in explaining the dependent variable. When the other variables of age, income, education, gender, and political party are included, the R-square rises dramatically to a significant 0.05156. Clearly the other independent variables are important factors in this regression. Significant and stronger Beta weights include the wording variable, age, the respondent's gender, Liberal and PC. The normal probability plot indicates non-normality, as with all of

the previous regressions. What is important about this case is that in addition to the socio-economic independent variables, the wording variable is playing an important explanatory role. In other cases, this has been an "either/or" situation, where only one of the order variables or the independent variables alone appear to be the strong influence in the regression equation. Therefore, while most of the understanding of the dependent variable is achieved through the other independent variables, the wording variable should still be closely examined.

The next regression's dependent variable relates to waste and the health care system. It is part of another experiment in question wording. In fact, in this case, there are four alternate question wordings, which have been dummied so as to be included in this regression. Despite all of these new wording variables, the strong significant relationships emerge between the dependent variable and PC, the dependent variable and Reform, Liberal by each of the other political parties, gender by income and age by education. The wording variables do not appear in the correlation matrix as vital to either the dependent variable, or to other independent variables.

When only the wording variables are included in the regression, only one tenth of one percent of the dependent variable is explained. This result is not significant. When the interviewer's gender is added into the equation, this R-square figure rises to almost one half of one percent, and becomes significant at 0.0216. Finally, when all of the independent variables are included, over four percent of the dependent variable is explained. This result is again significant. In this regression, strong and significant Beta Weights emerge for the variables of the interviewer's gender, the respondent's gender, income, and the political parties of PC, NDP and Reform. Once again, the normal probability plot and the standardized scatterplot indicate non-normal error distributions. Yet, this case is important because it provides the first instance of the interviewer's gender

having an impact on the dependent variable. This is even more remarkable in that it is derived from the data set gathered before the 1993 election, where the interviewer's gender does not appear to make an impact on the dependent variable in all of the other regression analyses. While the wording variable does not have the strength as seen in other examples, because of the new found importance of the interviewer's gender variable, this case surely merits further investigation.

The last regression involves the only case where variables attained from the pre-referendum data set are implemented. In this case, the wording variables have been dummied, and only two of the three newly dummied variables are included. The third is left out for the computer to use as a control or reference group. The correlation matrix reveals the two alternate wording variables to be related. In addition, the dependent variable is related to the first wording variable, as well as being related to the respondent's gender. This particular relationship is interesting because the variable concerns the guarantee of women and aboriginals senate seats. The respondent's gender is therefore a vital part of the approval of this measure. Also producing stronger and significant relationships are gender and employment status, gender and income, and Liberal by PC and NDP. This data set distinctly reflects the inequity suffered by women in Canadian society, with such apparent relationships between income, employment status and gender. It is also therefore consistent with the other data sets produced by these researchers. In an issue as sensitive as government representation, I believe it to be a good indicator of the data set that it is both consistent and reflective of Canadian society.

The wording variables produce a fairly large and significant R-square result of 0.02848. This is one of the higher results for just the wording or order variables, when entered alone on the first step of the regression analysis. Unfortunately, the gender of the interviewer was not included as a part of this data set. I find this to be a particularly large

oversight, especially considering the nature of the issue. As well, since the respondent's gender appears to have an impact on the dependent variable, it would not be unreasonable to assume that a great potential exists for the interviewer's gender to explain variation in the dependent variable. It is possible that this survey was a self-administered test, and therefore an interviewer could not impact on the respondent. Nonetheless, it does not appear that this survey was part of a mailed, self-administered survey, but instead part of the telephone surveys. Therefore, the interviewer's gender may be problematic, but not able to be explained in this analysis. When the other variables of age, education, income, employment status, gender and political party affiliation are added in on the next regression step, the R-square rises to an astoundingly significant 0.08556. This is the largest R-square measure of all the regressions performed in this research. In addition, almost all of the Beta weights are significant. Most are weak, though the stronger weights of the wording variable, education, gender, and PC, are noticeable.

Though the normal probability plot and standardized scatterplot again show non-normal distributions, the results of this regression are too strong to discount. Clearly the issue of senate representation for women and aboriginal peoples is sensitive to wording and gender differences. In addition, a third data set has confirmed the findings of links between gender and income. Clearly the disparities between the sexes are established, and these results show how generalizing to a larger population is possible and understandable. Though it is disappointing that the interviewer's gender can not be included in this analysis, these results definitely merit further examination on their own.

Altogether, these regression analyses appear to be an improvement over the crosstabs results. They clearly expand on the independent variables as having explanatory powers over variation in the dependent variables. As well, the regression outputs have confirmed that while small, the order and wording variation have impacted on the original

variables in some cases. This is very important to the study of nonsampling error because it establishes that such error does indeed exist, and plays a role in survey results. Also, the importance of interviewer gender does emerge in at least one case. Though socioeconomic and demographic variables do appear to account for much variation in the independent variables, they still do not explain the majority of variation in dependent variables. Other factors may be at work. Further statistical tests will show investigate the wording and order variables, and possibly explain more of their interaction with the dependent variable.

Though the standardized scatterplot in each of the regressions has indicated a non-normal error distribution, this may not be a great problem. It may reflect the inclusion of nominal variables in the equation. Since this occurred in all of the cases, this would be a probable explanation. Logistic regression will help to eliminate this problem, and hopefully confirm the positive findings of multiple regression without error distortion. I believe that despite some poor results, and error and case distributional problems, multiple regression has overall been a positive step for this research. It has helped to verify some suspected variable relationships, and definitively establish a strong foundation for further statistical research of nonsampling error.

Statistical Analysis — Logistic Regression

Logistic regression is a similar procedure to multiple regression, though it is better suited to dichotomous variables, such as the series of dependent dichotomous variables in seven of my statistical examinations. Logistic regression, in fact, requires that the dependent variable have no more than two response categories. The first dependent variable, which involved the issue of abortion, is not a nominal variable, and therefore its categories have not been dummied. It can not therefore be included in the statistical test of

logistic regression. It will be included in the discriminant analysis following this section. This procedure should not be affected in terms of error problems caused by assumption violations like multiple regression. Therefore, a more accurate understanding of these variables may emerge in using this procedure. Since one variable can not be included, there will therefore only be seven logistic regressions outputs explained in this section. The output produced by logistic regression is very similar to that of multiple regression. This examination will indicate the significance of the procedure, the computer classification of the dependent variable, and a comparison of logistic regression coefficients and significance to those of multiple regression.

The first logistic regression turns to the variable representing the issues of unemployment and inflation. As in the multiple regression procedure, the variables of interviewer gender, respondent gender age, income, education and political party affiliation are included in the logistic regression procedure. This produces a significant Model chi-square result, with ten degrees of freedom. In attempting to classify the dependent variable, the computer attains an overall accuracy rate of 67.24 percent. This regression procedure also creates logistic regression coefficients. In this logistic regression, the order variable, the respondent's gender, and PC are the larger coefficients that are also significant. With the exception of NDP, which just passes over the significance threshold, this exactly duplicates the results of the multiple regression, though the actual standardized regression coefficients in this case are much higher. This therefore appears to be a stronger regression result, which is positive in that the multiple regression output did not seem very promising in explaining the dependent variable. In this case, the order variable does appear to make a significant impact on the dependent variable, as it did in multiple regression.

TABLE 4. Logistic Regression: Logistic Regression Coefficients

Regression (Dependent Variable)	Variable	Standardized Logistic Regression Coefficients	Signifi- cance	Model Chi- Square Signifi- cance	Overall% Classified Correctly
ECQUEST (unemployment /inflation)	Order Variable	- 0.3482	0.0006**	0.0000*	67.24
	PC	- 0.3365	0.0127*		
	Respondent's Gender	- 0.3313	0.0014*		
	Interviewer's Gender	0.1025	0.3497		
	Liberal	0.1536	0.2329		
	NDP	0.4045	0.0665		
	Reform	- 0.2166	0.3225		
	Age	- 0.0001	0.0126*		
	Education	0.0069	0.5850		
	Income	- 5.1E-05	0.6885		
HLHQUEST (health/ universality)	Order Variable	- 0.5415	0.0000*	0.0000*	56.90
	NDP	-0.4061	0.0289*		
	Interviewer's Gender	0.1549	0.1174		
	Respondent's Gender	- 0.0569	0.5409		
	Liberal	- 0.0846	0.4553	0.0116*	65.95
	PC	0.0747	0.5553		
	Reform	0.2115	0.3142		
	Age	- 1.4E-05	0.8415		
	Education	0.0017	0.8523		
	Income	- 0.0002	0.1401		
EC2QUEST (inflation/ unemployment)	Order Variable	0.1116	0.6694	0.0116*	65.95
	Respondent's Gender	- 0.6648	0.0121*		
	NDP	1.2049	0.0324*		
	Interviewer's Gender	0.3693	0.1961		

TABLE 4. Logistic Regression: Logistic Regression Coefficients (Continued)

Regression (Dependent Variable)	Variable	Standardized Logistic Regression Coefficients	Signifi- cance	Model Chi- Square Signifi- cance	Overall % Classified Correctly
EC2QUEST (Continued)	Liberal	0.4548	0.2986		
	PC	- 0.3462	0.4932		
	Reform	- 0.0590	0.9064		
	Age	- 0.0030	0.7527		
PENQUEST (pensions/ universality)	Order Variable	- 0.7291	0.0036*	0.0133*	62.41
	Interviewer's Gender	0.3982	0.1370		
	Respondent's Gender	- 0.0565	0.8214		
	Liberal	0.4226	0.3404		
	PC	- 0.0343	0.9468		
	NDP	0.3357	0.5150		
	Reform	- 0.5083	0.3390		
	Age	- 0.0060	0.4930		
	Feel About Old People	- 0.0003	0.7561		
TAX (GST question)	Wording Variable	0.4776	0.0000*	0.0000*	66.40
	Reform	- 0.3724	0.0791*		
	PC	- 0.9852	0.0000*		
	Respondent's Gender	- 0.2384	0.0154*		
	Liberal	- 0.3465	0.0042*		
	Interviewer's Gender	0.1413	0.1800		
	NDP	0.0066	0.9744		
	Age	4.53E-05	0.5557		
	Education	- 0.0236	0.0362*		
WASQUEST	Income	8.30E-05	0.4906	0.000*	59.52
	Wording Variable (1)	0.1034	0.2625		

TABLE 4. Logistic Regression: Logistic Regression Coefficients (Continued)

Regression (Dependent Variable)	Variable	Standardized Logistic Regression Coefficients	Signifi- cance	Model Chi- Square Signifi- cance	Overall% Classified Correctly
WASQUEST (continued) (waste/ health care)	Wording Variable (2)	0.0941	0.3076	0.0000*	65.56
	Interviewer's Gender	0.2672	0.0077*		
	PC	0.5645	0.0000*		
	Income	- 0.0003	0.0239*		
	NDP	- 0.5496	0.0038*		
	Reform	1.0919	0.0000*		
	Respondent's Gender	0.3067	0.0011*		
	Liberal	0.0773	0.4915		
	Age	- 1.3E-05	0.8514		
	Education	- 0.0007	0.9380		
SENQUEST (Senate Representation)	Wording Variable (1)	- 0.4429	0.0002*	0.0000*	65.56
	Wording Variable (2)	- 0.9373	0.0000*		
	Respondent's Gender	- 0.5643	0.0000*		
	Liberal	- 0.2341	0.0295*		
	PC	- 0.7460	0.0000*		
	NDP	- 0.2502	0.1106		
	Reform	- 0.9461	0.0001*		
	Age	- 0.0005	0.0523*		
	Employment Status	- 0.0573	0.0011*		
	Education	0.0611	0.0038*		
	Income	9.43E-05	0.3710		

a. An * denotes significance below the 0.05 level.

The health care/universality variable is the next dependent variable to be used in logistic regression. In this case, a significant Model chi-square result is also produced. The computer does not accurately predict the dependent variable as well though; only 56.90 percent of cases are accurately classified. Strong and significant logistic regression coefficients emerge for the order variable and the NDP variable as well in this instance, as had occurred in multiple regression. As in the previous logistic regression, the weights appear much stronger than in multiple regression, and therefore may have a greater impact on the dependent variable than multiple regression predicted. Multiple regression had indicated that the order variable was important in this analysis, and logistic regression confirms this observation.

The next logistic regression relies on the deficit/unemployment variable as the dependent variable. In this test, eight degrees of freedom are indicated, and a significant Model chi-square result of 0.0116 is produced. The computer is able to correctly classify 65.95 percent of all cases of the dependent variable accurately. This is better than the last logistic regression, though still not as good as the first. This dependent variable comes from the data set arrived at after the 1993 federal election. Therefore, this is comparable in consistency to the first logistic regression, which also contained the issue of unemployment, and had used variables derived from the data set gathered before the election. In this case, respondent gender and NDP also emerge as the strong and significant logistic regression coefficients, as they had in multiple regression. This result therefore confirms and strengthens the findings of the previous statistical test since the regression coefficients are again stronger in logistic regression.

The next logistic regression's dependent variable is also derived from the data set produced after the election. It involves the issues of pensions and universality. Nine degrees of freedom are indicated, and a significant Model chi-square result is produced at

0.0133. As occurred in multiple regression, the order variable is the only strong and significant logistic regression coefficient, at - 0.7291. Nonetheless, other stronger logistic regression coefficients are produced that are not significant. For example, the interviewer's gender variable, as well as the political party variables of PC, Liberal, NDP, and Reform, all have stronger logistic regression coefficients. Nonetheless, the strong order variable is important for several reasons. First, it confirms that the order variable is having an impact on the dependent variable, as indicated in previous tests using this dependent variable. Second, the significance of this regression coefficient allows it to apply beyond this data set to the general population. This indicates that it is not an anomaly of this data set, and this is important in the understanding and correction of nonsampling error in surveys. If order error only applies to one data set, then it would be difficult to establish that it applies to all data sets, that it can be replicated, and that it should be dealt with as a serious problem overall.

The next logistic regression involves variables that revert back to the data set accumulated before the election. This is the first of the wording experiments. In this case, the dependent variable involves the Goods and Services Tax. A perfectly significant Model chi-square result is produced, with ten degrees of freedom indicated by the computer. In the classification table, 66.40 percent of the dependent variable is able to be classified by the computer. The wording variable again produces a strong and significant regression coefficient, as occurred in multiple regression. In addition, PC, Liberal, and the respondent's gender variables are also indicated as strong and significant coefficients in logistic regression as well. Age, however, is not strong or significant as it had been in multiple regression. Instead, Reform, which had not been either strong or significant in multiple regression becomes both in logistic regression. As in the previous logistic regression, what is important in this case is that the wording variable continues to be

influential on the dependent variable.

The dependent variable for the next logistic regression is also an experiment in question wording, this time involving the issues of health care and waste. The computer does not classify the dependent variable as well in this case, with only 59.52 percent classified correctly overall. The logistic regression however, is perfectly significant, and eleven degrees of freedom are indicated. Unfortunately, as occurred in multiple regression, the wording variables are not strong or significant regression coefficients in this logistic regression. Income loses both its strength and significance in logistic regression as well. Fortunately, the interesting variable of this examination, the interviewer's gender variable, gains strength and remains significant. This confirms the findings of multiple regression, and again stresses another area of nonsampling error: that of interviewer influence, as discussed in chapter 1. Therefore, while the wording variable does not stand up to the hypothesis expectation of influence and affect the dependent variable, the findings of the impact of the interviewer's gender may be equally important for the study of nonsampling error.

The last logistic regression involves the issue of senate representation for women and aboriginal peoples, and uses an experiment in question wording. In this case, as occurred in multiple regression, the wording alternatives have been dummied, and only two have been included to avoid redundancy. Like the previous logistic regression, eleven degrees of freedom are indicated by the computer, and the Model chi-square result is perfectly significant at 0.0000. In this examination, the computer is able to classify 65.56 percent of the cases of the dependent variable. Both wording variables in this analysis produce extremely strong and significant logistic regression coefficients. Education loses its strength, though the respondent's gender and PC retain both their strength and significance. In addition, Reform becomes very a very strong and statistically significant

coefficient. As in multiple regression, almost all of the regression coefficients are significant, though fewer weights than in multiple regression are very strong. This consistency between types of regression, as well as the dramatic increase in the importance of the wording variable confirms the presence of the nonsampling error in this variable. This variable clearly should be further examined for error, and its impact on the dependent variable.

Logistic regression has overall been a useful tool in this analysis. For the most part it has confirmed the findings of multiple regression. It has also shown the key variables to be having a greater impact on the dependent variable than multiple regression indicated by way of logistic regression coefficient strength. I believe that this demonstrates that the problems of multiple regression are therefore the product of the violation of multiple regression assumptions by involving nominal dichotomous variables. These results are therefore not affected by more problematic error distribution problems. Logistic regression has continued to confirm the suspicion that nonsampling error exists in these variables, as initially indicated by frequency distribution analysis. Though crosstabs did not confirm that diagnosis, these two types of regression appear not only to confirm the presence of error, but to encourage further investigation.

Statistical Analysis — Discriminant Analysis

Discriminant analysis can be performed on all types of dependent variables. Therefore, I will examine eight discriminant analyses, or one for each dependent variable. It should be noted that the dependent variables do not have to be dummied in order to be used in this analysis, and they are therefore used in their original form. Discriminant analysis produces several statistics that will be discussed in this examination. First, the squared Canonical correlation is very similar to the R-square result from multiple

TABLE 5. Discriminant Analysis: Results

Variable	Canonical Correlation	Box's M Significance	Percentage of Grouped Cases Classified Correctly
ABQUEST (Abortion)	0.1531005	0.00088* ^a	38.00
ECQUEST (Unemployment/ Inflation)	0.1551476	0.0000*	57.71
HLHQUEST (Health/ Universality)	0.1547769	0.0010*	57.30
EC2QUEST (Deficit/ Unemployment)	0.2600186	0.2204	59.86
PENQUEST (Pensions/ Universality)	0.2637379	0.0001*	59.66
TAX (GST Question)	0.2270677	0.0000*	61.34
WASQUEST (Waste/ Health Care)	0.2099984	0.0000*	58.03
SENQUEST (Senate/ Representation)	0.2945581	0.0000*	50.36

a. An * denotes significance. This indicates that the relations between predictors did differ among groups being analyzed, and therefore the assumptions of discriminant analysis have been violated.

regression. These will be compared from the original multiple regression to the new, discriminant analysis of the same variables. In addition, functions created by discriminant analysis will show which variables are strong within the statistical test. Box's M will yield a result that indicates whether the relations between the predictors differ among the groups being analyzed. If they do differ, then some of the assumptions behind discriminant analysis are violated. Finally, the computer will attempt to classify the cases into groups. This result should be better than what would occur solely on the basis of chance.

The first discriminant analysis, involving abortion as the dependent variable issue, shows the variables of Liberal, PC, age, education and income to be significant. These variables have consistently surfaced in other tests as significant, so this confirms their importance in this analysis. The result of the Canonical correlation, when squared, is similar in meaning to the R-square measure in multiple regression. In this case, when squared, the result is very close, at 2% compared to 1.9% in multiple regression. Two functions are produced in this discriminant analysis. The first function shows the order variable and age to be the strongest, while the second function shows the order variable and education to be the strongest. Age and education may both be important predictors in attitudes towards abortion regulation. As one grows older, one gains more knowledge of the issue, and possibly changes opinion based on new information available.

The Box's M statistic proved to be significant at 0.0000, indicating that the relation between the predictors did differ among the groups being analyzed by this discriminant analysis. This does violate some of the underlying assumptions of discriminant analysis. Finally, in this analysis, out of three possible groups, 38 percent of all cases are classified correctly by the computer. This is only slightly better than would occur if this were left to chance. These findings appear to confirm the overall explanatory power of the variables involved in the analysis. While they may not be extremely influential, these variables do

play a role in the understanding of the dependent abortion variable.

The second discriminant analysis for the dependent variable of unemployment and inflation finds the order variable, the respondent's gender, and the political party variables to be significant. The squared Canonical correlation results in an identical figure to the R-square measure from multiple regression at 2.4%. Only one function is produced, and the strongest variables in it are the order variable, the respondent's gender, and PC. Box's M is again significant, indicating that some assumptions of discriminant analysis have been violated. Finally, the computer is able to correctly classify 57.71% of all cases out of two groups. Again, this is only slightly better than chance. These findings again confirm the minimal impact of these variables, and their ability to explain only a small amount of variation in the dependent variable. Therefore, in this case, it would appear that the smaller findings of multiple regression, as compared to the larger coefficients present in logistic regression, may be closer to a realistic picture of these variables.

The third discriminant analysis involves the dependent variable relating to health care and universality. Significant variables in this analysis include the order variable, the interviewer's gender variable, and the NDP political party variable. The squared result of the Canonical correlation again yields the identical result of the final R-squared from multiple regression, at 2.39%. This figure is significant. In the only function produced, the order variable appears as the sole strong factor. Box's M is again significant, as has occurred in the previous discriminant analyses. Of the two possible groups, the computer is able to classify 57.30% of cases accurately. This figure is very similar to the previous discriminant analysis. It is interesting that this occurs, because fewer of the variables in this analysis appear as important as in the last discriminant analysis, and yet the results are very similar. This could be the work of a stronger order variable, which allows the

TABLE 6. Discriminant Analysis: Significance and Standardized Canonical Discriminant Function Coefficients

Dependent Variable	Variable in Function	Significance	Discriminant Function Coefficient Result for Function 1	Discriminant Function Coefficient Result for Function 2 (where applicable)
ABQUEST (Abortion)	Order Variable (1)	0.9838	0.18365	0.33619
	Order Variable (2)	0.0005* ^a	0.45531	0.79103
	Interviewer's Gender	0.8257	0.05813	0.10008
	Respondent's Gender	0.1250	- 0.12146	0.27311
	Liberal	0.0406*	0.32569	0.13595
	PC	0.0050*	- 0.33748	0.16551
	NDP	0.18869	- 0.11350	0.32578
	Reform	0.1617	0.29736	- 0.10611
	Age	0.0057*	0.49843	0.17210
	Education	0.0187*	- 0.39043	0.40079
	Income	0.0086*	0.38094	- 0.18483
ECQUEST (Unemployment/ Inflation)	Order Variable	0.0008*	0.51929	
	Interviewer's Gender	0.4939	- 0.14346	
	Respondent's Gender	0.0005*	0.49134	
	Liberal	0.0191*	- 0.19603	
	PC	0.0004*	0.42376	
	NDP	0.0400*	- 0.29062	
	Reform	0.2025	0.15930	

TABLE 6. Discriminant Analysis: Significance and Standardized Canonical Discriminant Function Coefficients (Continued)

Dependent Variable	Variable in Function	Significance	Discriminant Function Coefficient Result for Function 1	Discriminant Function Coefficient Result for Function 2 (where applicable)
ECQUEST (continued)	Age	0.1276	0.23762	
	Education	0.7824	- 0.08771	
	Income	0.7721	0.06273	
HLHQUEST (Health/ Universality)	Order Variable	0.0000*	0.8650	
	Interviewer's Gender	0.0936	- 0.22718	
	Respondent's Gender	0.9026	0.09064	
	Liberal	0.4188	0.11981	
	PC	0.3218	- 0.09278	
	NDP	0.0285*	0.33187	
	Reform	0.1324	- 0.15096	
	Age	0.9758	0.02957	
	Education	0.9755	- 0.02812	
	Income	0.1254	0.21917	
EC2QUEST (Deficit/ Unemployment)	Order Variable	0.6651	0.10256	
	Interviewer's Gender	0.3373	0.31018	
	Respondent's Gender	0.0198*	- 0.59069	
	Liberal	0.3978	0.42064	
	PC	0.0438*	- 0.23932	
	NDP	0.0103*	0.70676	
	Reform	0.2388	- 0.04483	

TABLE 6. Discriminant Analysis: Significance and Standardized Canonical Discriminant Function Coefficients (Continued)

Dependent Variable	Variable in Function	Significance	Discriminant Function Coefficient Result for Function 1	Discriminant Function Coefficient Result for Function 2 (where applicable)
EC2QUEST (continued)	Age	0.6721	- 0.08240	
PENQUEST (Pensions/ Universality)	Order Variable	0.0012*	0.66457	
	Interviewer's Gender	0.1016	- 0.34468	
	Respondent's Gender	0.8506	0.04465	
	Liberal	0.0511*	- 0.39288	
	PC	0.4070	0.01713	
	NDP	0.4108	- 0.22970	
	Reform	0.0166*	0.30187	
	Age	0.3718	0.15857	
	Feel About Old People	0.3222	0.06156	
TAX (GST Question)	Order Variable	0.0000*	- 0.49299	
	Interviewer's Gender	0.1276	- 0.13341	
	Respondent's Gender	0.0034*	0.24570	
	Liberal	0.7572	0.31140	
	PC	0.0000*	0.84045	
	NDP	0.0866	- 0.00408	
	Reform	0.8550	0.17683	
	Age	0.7994	- 0.05945	
	Education	0.0342*	0.23106	
	Income	0.1069	- 0.07003	

TABLE 6. Discriminant Analysis: Significance and Standardized Canonical Discriminant Function Coefficients (Continued)

Dependent Variable	Variable in Function	Significance	Discriminant Function Coefficient Result for Function 1	Discriminant Function Coefficient Result for Function 2 (where applicable)
WASQUEST (Waste/ Health Care)	Order Variable (1)	0.2771	0.11793	
	Order Variable (2)	0.3238	0.10823	
	Interviewer's Gender	0.0065*	0.28269	
	Respondent's Gender	0.0000*	0.35431	
	Liberal	0.2599	0.08192	
	PC	0.0000*	0.51251	
	NDP	0.0000*	- 0.33114	
	Reform	0.0000*	0.52653	
	Age	0.5683	- 0.02244	
	Education	0.6764	- 0.00734	
	Income	0.0009*	- 0.24672	
SENQUEST (Senate/ Representation)	Order Variable (1)	0.4166	0.03603	0.20772
	Order Variable (2)	0.0000*	0.73634	0.28803
	Respondent's Gender	0.0000*	0.43850	- 0.37997
	Liberal	0.0686	0.19502	0.54522
	PC	0.0000*	0.45774	0.07741
	NDP	0.3729	0.13415	0.39258
	Reform	0.0008*	0.29816	0.35692
	Age	0.0057*	0.14097	0.38991
	Education	0.0005*	0.30709	0.02010

TABLE 6. Discriminant Analysis: Significance and Standardized Canonical Discriminant Function Coefficients (Continued)

Dependent Variable	Variable in Function	Significance	Discriminant Function Coefficient Result for Function 1	Discriminant Function Coefficient Result for Function 2 (where applicable)
SENQUEST (continued)	Employment Status	0.0000*	- 0.20681	0.42240
	Income	0.0365*	- 0.06864	0.02567

a. An * denotes significance below the 0.05 level.

computer to predict the same as it can with a weaker, but a larger number of variables. Also, the order variable has appeared fairly strong in previous tests, and this appears to be confirmed in discriminant analysis for the health variable.

The next discriminant analysis involves the variable representing unemployment and the deficit, as attained from the data set gathered after the 1993 election. The respondent's gender, PC, and NDP appear as the only significant variables in this analysis. The Canonical correlation is again consistent with the R-square measure from multiple regression, at 6.7%. This is also significant. In the function created by the computer, the strong variables are the respondent's gender, and the political party variables of Liberal and NDP. Box's M in this analysis is not significant. Therefore, the assumptions of discriminant analysis are not violated, and the predictors do not differ among the groups being analyzed. The computer is able in this case to predict slightly more cases than previously, at 59.86% accuracy. These results are also consistent with the other tests conducted with this dependent variable. They indicate that the order variable is not as important as gender and political party affiliation. This may be explained by the fact that the deficit and unemployment were very key issues in the election, and with this fresh in respondent's minds, it impacts more on their responses than a change in question order.

Pensions and universality are at the heart of the dependent variable in the next discriminant analysis. The order variable, Liberal and Reform are the significant variables in this analysis. The Canonical correlation is again significant, and identical to the R-square measure from multiple regression at 6.95%. The strongest variable among the discriminant function coefficients is the order variable. In this case Box's M is significant, and 59.66% of all cases are classified correctly into two groups. The order variable also proved to be very strong and influential in terms of the dependent variable in other statistical tests with this dependent variable. This discriminant analysis is therefore

consistent with these findings.

The dependent variables in the next three discriminant analyses will involve challenges in question wording. The first relates to the question representing changes in the question involving the GST. The significant variables in this analysis are the wording variable, the respondent's gender variable, PC, and age. Clearly the Conservative party is related to the GST, because it implemented the GST when in power federally in Canada. The squared product of the Canonical correlation is again the same as the R-squared result at 5.15%. This figure is significant. The wording variable and the PC political party variable are strongest within the discriminant function. Box's M is significant, and therefore some of the assumptions of discriminant analysis are violated. Finally, the computer does better than in previous discriminant analyses, and accurately classifies 61.34% of cases into two groups. These results appear to enhance the importance of the question wording variable, as compared to previous results. As well, it becomes apparent that this issue is indeed political, rather than more economic in nature in the minds of Canadians, and is affected by political party affiliation.

The next discriminant analysis involves the dependent variable that relates to the issues of health care and waste. The significant variables in this analysis are the interviewer's gender, the respondent's gender, all political party affiliations, and income. Income should clearly play a role in a variable that relates to the waste of public tax money. Political parties should also be related because of their role in this important government function. The Canonical correlation again proves consistent, with a result of 4%. This is again significant. The strongest variables in the discriminant function are PC and Reform. This may be caused by the stress of these political parties against wasteful practices in government. Box's M is significant, and only 58.03% of cases are predicted accurately. This analysis does not appear as strong as some of the previous discriminant

analyses. Nonetheless, it does remain consistent with some of its own results in other statistical tests, such as in multiple regression.

The last discriminant analysis involves the dependent variable that relates to the issue of guaranteed senate representation for women and aboriginal peoples. As in previous tests, almost all of the variables are significant, with the exception of NDP, and one of the two order variables. The squared Canonical correlation is significant, and very high as it had been in multiple regression, at 8.6%. This discriminant analysis produces two functions. In the first, the stronger variables are one of the order variables, the respondent's gender variable, and the PC political party variable. In the second, Liberal and employment status emerge as the stronger variables. Box's M is significant, and a high number of cases are accurately classified by the computer, at 50.36% accuracy. This is much better than would occur on the basis of chance. These results, like other statistical results for this variable, appear quite strong. This further confirms the analysis that the wording variable is having an impact on the dependent variable.

Discriminant analysis has for the most part confirmed the statistical results gained in other tests. It has indicated that some of the larger results should be tempered, and that a more cautious judgement should be made of some analyses. Nonetheless, this test demonstrates that order and wording variables are having an impact on some of the variables. Therefore, even if smaller than indicated by other tests, such as logistic regression, nonsampling error does appear to impact on the dependent variables in this statistical study.

Conclusion

This statistical research has been an interesting and fruitful examination of the data made into a computer readable set by Northrup and Oram. They have done a great service

to the statistical research community by providing data that can be scrutinized and subjected to new statistical tests. Indeed, by allowing the examination of their data by other researchers, the secrecy often surrounding nonsampling error can be eliminated, and the understanding of such error can increase, while problems are corrected.

Some conclusions of the results of order and wording effects on the dependent variables of this empirical examination can be drawn. I have used frequency distribution, crosstabs, multiple and logistic regression and discriminant analysis to examine eight dependent variables. In the first, it is unclear whether the approval of abortion services dependent variable suffers order effects. In initial examination, it appears that this variable, or at least two of its response categories are affected by response order. However, further statistical tests do not clearly confirm this finding. In fact, while an order effect may exist, it appears quite small, and not always applicable to the general population. The problem may be that the response categories are not all consistently affected. In this case, for whatever reason, the overall statistical results are not strong enough to accept unflinchingly the hypothesis. The null hypothesis for this case must therefore be accepted. Nonetheless, since abortion is a very sensitive issue, it would be useful to conduct further investigation into this issue.

The variable concerning unemployment and inflation also does not appear initially to be strongly affected by response order. Further probing however, does indicate that while this effect is minimal, it is significant, and applicable to the larger Canadian population. While a definitive acceptance can not occur in this case, the null hypothesis can not be fully accepted either, since a significant order effect did appear in statistical tests. The second dependent variable involving unemployment, as well as the deficit, also produces fairly weak results. Again however, these results are generally significant, and somewhat stronger than the first economic question results. In addition, the dependent

variable in this case becomes more involved in other variable relationships than it had in the case of the unemployment/inflation variable. In this case, as in the previous two cases, it appears that the independent variables explain much more of the dependent variable than the order variable. Since the order variable does not exercise a noticeable influence here, the null hypothesis must again be accepted.

The health/universality variable statistical tests are the second statistical example where the order variable appears to influence the dependent variable. Though weak, the order variable does appear to have a significant impact on the dependent variable. The pension/universality variable also has positive results in this area. In both cases, the order variable, without the other independent variables, appear to explain at least a minimal part of the dependent variables. In these cases, I would tentatively accept the hypothesis, that the order variables do affect the dependent variables. This however, must be done with the full knowledge that according to these statistical results, the impact is small, and should be further tested, since not all tests clearly confirm this conclusion.

The waste/health care dependent variable fares the worst among the wording variation variables. Its four alternative wording variables have very little impact on the dependent variable, and are hardly linked to the dependent variable in any of the statistical tests performed in this statistical examination. The hypothesis in this case must therefore be soundly rejected. The case of the GST variable is somewhat more unclear. The wording variable, as in the case of the health/universality order variable, does indeed seem to have an impact, though small, on the dependent variable. Of course, small in this research is a matter of comparison. Against the other results, the impact of the wording variable on the dependent variable is quite large. Nonetheless, the hypothesis here is only tentatively accepted, since not all tests clearly indicate the strong influence of this variable on the dependent variable. Further tests are certainly warranted in this case to establish the

impact of the wording variable.

The last variable, involving the issue of guaranteed senate seats for women and Aboriginal peoples clearly indicates the strong presence of wording effects in of the all statistical tests. In this case, I unhesitatingly accept the hypothesis. Nonetheless, I believe that this should be further examined, especially in further research where the impact of the interviewer's gender can be included in the analysis. The impact of the interviewer's gender does appear with other dependent variables, and it should be included in all statistical research, but especially in combination with an issue so clearly affected by gender issues.

Tenuous thought they may be, variable links do exist in many of these variable combinations. Socio-economic and demographic characteristics do help to explain respondent behaviour. Nonetheless, in some cases, other factors are at work that contribute to variation in respondent choices. Theoretical research has pointed to the issues of question wording, question order, and interviewer influence, even when unintended, such as in the case of the gender of the interviewer. I believe that rather than emphasize the minimal impact of order and wording on statistical results, this research should confirm the existence of nonsampling error, and provide a foundation for further investigation into this area. Nonsampling error must be accepted as a part of survey research, so it can begin to be corrected for, and so that statistical examination will not suffer interpretation problems and lingering doubt over researchers' conclusions.

Conclusion

The study of nonsampling error is vital to the complete understanding and correct usage of survey research. Though this type of error often remains hidden and undiscussed by survey researchers, it does not lessen its potential for misdirecting or weakening survey results. The distortion of survey results does not necessarily occur intentionally; therefore it should become even more important to understand the nuances of error. Researchers are often reluctant to discuss error in terms of their own work, so as to present a facade of strength and certainty in their statistical results. The inclusion of error, however, does not weaken statistical results. Instead, admitting that error exists, and attempting to correct it, should alleviate doubts about the quality of the research, and prove the researcher's decided determination in eliminating such error.

Nonsampling error, in comparison to definable sampling error, often appears vague in defining problems and enumerating solutions. This has caused it to be less discussed by both statisticians and less investigated by the research community. Nonetheless, this is slowly changing. Theoretical information has been written for some time that brings forward the debate over nonsampling error. In addition, some empirical studies have been produced on a larger scale that attempt to gauge nonsampling error. This particular research study has been an extension of those ideals in discussing the theoretical information, providing empirical data, and attempting to establish the existence and to determine the impact of nonsampling error on survey results. Specifically, this project has dealt with the areas of interviewer influence, and question and instrument influence on variation in survey responses.

Interviewers may cause respondents to change their responses because of several stimuli. Respondent's may be reacting in response to a desire to appear socially desirable in the eyes of the interviewer. They may not give responses that appear too extreme, so as

to normalize the presentation of their opinions. This urge to look respectable to the interviewer may be caused by the perceived socio-economic or demographic characteristics of the interviewer. For example, if the interviewer is of a different race or gender, the respondent may modify his responses so as not to offend the interviewer. This type of affect on survey responses is clearly not caused by overt behaviour of the interviewer. It is unavoidable in some cases, and studies have shown that it can be alleviated somewhat by race or gender matching.

Some problematic occurrences however, are caused by the interviewer, even if unintentionally. Interviewers may fail to ask questions or enumerate responses. They may record inconsistently or poorly. They emphasize different words when speaking with respondents, causing the respondents to believe that certain parts of a question or response options are more important, and should therefore be paid more attention in responding. Interviewers may respond differently to requests for help, and in turn receive a different quality of answer from respondents. These and other effects, as discussed in the first chapter, are often difficult to measure, repeat, and control systematically.

Interviewers must be offered consistent training and instruction in methods of interviewing and recording procedures. In addition, studies that attempt to measure interviewer influence must be careful to manipulate only one variable at a time. For example, if the researcher is measuring the impact of the interviewer's race, then the interviewers should be equally split among genders, so that not all of one race of interviewers is also all one gender. If not done correctly, this problem could affect survey estimates so that it would be difficult to determine which factor was actually having an effect on survey respondents. Since interviewers will always comprise an important part of survey research, it is important not only to maintain a high quality and consistency in interviewing, but to understand realistically the interviewer's impact on survey estimates,

so that this can be accounted for, and ultimately corrected in future survey analysis.

Problems surrounding questions in surveys are often more ambiguous than those related to interviewer influence. Though more difficult to establish, problems relating to questions are no less important in the field of nonsampling error. In fact, while sampling error is often accepted as a normal part of survey research, critics often charge that researchers deliberately use nonsampling error, in the form of question manipulation, in order to change survey estimates. Furthermore, this type of influence can occur in all types of survey research, while the direct impact of interviewers is certainly more limited in certain types of surveys, such as mail surveys.

Problems relating to questions are actually quite diverse. The response categories of questions may not capture all of the possible answers, or accurately reflect respondent attitudes or experiences. In addition, the use of broad, sweeping general categories may fail to distinguish between the nuances of respondent answers. Question comprehension is often targeted as a reason as to why questions fail to elicit accurate respondent answers. Questions that are too vague, or tax the cognitive abilities of respondents, may cause them to choose an answer that is not most correct, because they do not understand the question. In this case, respondents may choose answers to pass over the question, or choose the first available response, so as not to appear uneducated to the interviewer.

The context of the question is also an important factor. A question can change in meaning depending on its context within the survey, within the target population, or within a particular time frame. For example, a similar question asked twice in a survey may make the respondent answer the same both times because she wishes to appear consistent. Or, the previous question may elicit an attitude that will influence the next series of questions on a particularly sensitive topic. As well, questions asked at the time of an election may

yield different results than when those same questions are asked two years after an election. Context, or order variation, can clearly affect survey estimates. This area has been the subject of several investigations, especially with regard to the order of survey responses. The examination of the problematic nature of question order also was examined in the empirical work of this research.

Question wording is closely related to question order in nonsampling error. Subtle changes of question wording, including the addition or subtraction of certain words can drastically affect survey estimates. Particularly sensitive words, such as “welfare” may conjure up images in respondents’ minds which will cause them to react differently to questions than if a less sensitive word such as “social aid” had been listed instead. Ambiguous words or vague quantifiers are also problematic. Context may make these words slightly more clear, though absolute meaning in terms of occurrences distinctly differs among individuals. “Sometimes” may mean different things to different people, even if given in the same context. One important point in question wording experiments is that only one word or concept should be altered at a time. Otherwise, as in the interviewer influence test, it will be impossible to distinguish between the effects of different variables. Word experiments were also conducted in the empirical test section of this research.

Finally, other problems concerning questions include the difference in the quality and coding of open and closed ended questions. Open ended questions may elicit more information, but they may also be more open to variation in interviewer style, and subject to coding problems. In addition, the length of the question may mean a difference in the length of time that the respondent speaks. One study showed that the longer the question, the more information was provided by respondents. While this may result in more information, it is clearly limited by the resources of the researcher to be able to have

longer interviews, and may be subject to more interviewer influence. In sum, problems relating to questions can clearly affect the marginal distributions. Only empirical examination of this phenomenon will tell the extent of this impact, and only researchers can divulge the extent to which question changing tactics are employed in everyday survey research.

Statistical analysis of question order, question wording, and interviewer gender in Chapter 3 has been based on data grouped by David A. Northup and Anne E. Oram. In this section, eight dependent variables were examined so as to determine if they had been affected by changes in question wording, question order, by the interviewer's gender, or by other socio-economic or demographic factors, such as age, education, income, and political party affiliation. The tests of crosstabs, multiple and logistic regression and discriminant analysis were performed on these variables.

In general, the results from these tests were significant but weak. This means that the results can be applied to the general population, although the actual impact of order, wording and interviewer gender on survey response variation was fairly small. In some cases, such as in the sensitive issue of abortion, this empirical examination did not show any significant effect of order at all, in contrast to the study produced by Schuman, as discussed in the second chapter. Other variables did however show wording to impact on responses, such as in the case of the variable that related to the issues of guaranteed senate representation for women and Aboriginal peoples. Most variable results fell within these two extremes. In comparison to my initial assertions as to the sensitivity of certain measures, the actual survey results compare fairly well. Specifically, the health care variables, and the variable that concerns women and the senate, do appear to be more susceptible to nonsampling error than the other dependent variables. This however, does not prove to be true in the case of the abortion variable, which had also been predicted to

be more sensitive. Nonetheless, it does appear overall, that as researchers have previously asserted, marginal results that reflect sensitive issues do seem to be more affected by nonsampling error.

In about half of the cases, the hypothesis can be tentatively accepted, that order, wording variation as well as interviewer gender do indeed affect variation in the dependent variable. This acceptance is tentative, because clearly the results are not so strong as to completely eliminate all doubt as to what is causing most of the variation in survey results. To some extent, order, wording and gender do influence survey marginals, even if this influence is fairly minimal. This research does not therefore, disconfirm the existence and impact of nonsampling survey error. Instead, I would advocate that it proves error to exist, though its influence may be smaller than expected, at least for this example. This research should tie the previous theoretical research with a large scale empirical examination of nonsampling survey error, and provide a foundation for further research.

Opinions are complex, and survey research often attempts the impossible in trying to depict public opinion based on several simple survey questions. This in itself will cause error, and this error will continue to be exacerbated by poor interviewers, and flawed questions and survey instruments. Yet the cost of eliminating survey error may be prohibitively high, and the rewards for doing so not immediately forthcoming. Still, statistical firms who choose to present their findings as error free, and entirely representative of complex and multi-faceted issues do not do justice to the field of survey research.

Information about potential nonsampling survey error must be made public by survey research firms and analysts. As well, there must exist a greater forum for the discussion of error, reliability and validity that does not threaten researchers, or shed doubt

on their empirical findings. Instead, the examination of this error should confirm the efforts of researchers to produce the best possible unbiased survey results. Error will always be a part of survey research, and in this growing information age, surveys will forever be a part of life. In the research community, the least we can do is to expand the discussion of these issues, if not to help to eliminate the problems of nonsampling error altogether.

Appendix: Frequency Distributions for the Dependent Variables

For SunOS 4.0
Number 2129^c

SPSS -- Carleton University

License

This software is functional through August 31, 1999.

Try the new SPSS Release 4.0 features:

- * LOGISTIC REGRESSION procedure
- * EXAMINE procedure to explore data
- * FLIP to transpose data files analysis
- * MATRIX Transformations Language SPSS Graph
- * CATEGORIES Option:
 - * conjoint analysis
 - * correspondence
- * GRAPH interface to

See the new SPSS documentation for more information on these new features.

```
1 0 get file 'elec1993'.
```

File elec1993

Created: 15 FEB 95 06:31:00 - 872 variables

```
2 0 frequencies variables = cpsg6a cpsg6b cpsg6c
3 0 cpsrn10 cpsl9a cpsl9b cpsrn21 cpsl10a cpsl10b
4 0 cpsrn8 pese4a
5 0 pese4b pesrn2 pese5a pese5b pesrn3
6 0 cpsrn1 cpsg4 cpsl6b cpsrn6 cpsl11a cpsl11b cpsrn3
7 0 cpsrn4refe6 refe7 refe8 refrn14/statistics all.
```

There are 151,064 bytes of memory available.
The largest contiguous area has 150,821 bytes.

Memory allows a total of 5,386 values accumulated across all variables.
There may be up to 673 value labels for each variable

CPSG6A OPINION> 3 POSITIONS: ABORTION <CPSRN10>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Never Permitted	1	94	1.9	8.4	8.4
Permitted=Need	2	388	8.0	34.5	42.9
Peonal Choice	3	617	12.7	54.9	97.8
D.K.	8	13	.3	1.2	98.9
Refused	9	12	.2	1.1	100.0
	.	3747	76.9	Missing	
	Total	4871	100.0	100.0	
Mean	2.609	Std err	.033	Median	3.000
Mode	3.000	Std dev	1.096	Variance	1.202
Kurtosis	16.637	S E Kurt	.146	Skewness	3.200
S E Skew	.073	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	2933.000		
Valid cases	1124	Missing cases	3747		

CPSG6B OPINION> 3 POSITIONS: ABORTION <CPSRN10>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Permitted=Need	1	396	8.1	30.0	30.0
Personal Choice	2	755	15.5	57.2	87.1
Never Permitted	3	145	3.0	11.0	98.1
D.K.	8	11	.2	.8	98.9
Refused	9	14	.3	1.1	100.0
	.	3550	72.9	Missing	
Total		4871	100.0	100.0	
Mean	1.934	Std err	.030	Median	2.000
Mode	2.000	Std dev	1.107	Variance	1.225
Kurtosis	22.483	S E Kurt	.135	Skewness	4.054
S E Skew	.067	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	2555.000		

Valid cases 1321 Missing cases 3550

CPSG6C OPINION> 3 POSITIONS: ABORTION <CPSRN10>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Personal Choice	1	725	14.9	54.5	54.5
Never Permitted	2	89	1.8	6.7	61.2
Permitted=Need	3	493	10.1	37.1	98.3
D.K.	8	12	.2	.9	99.2
Refused	9	11	.2	.8	100.0
	.	3541	72.7	Missing	
Total		4871	100.0	100.0	
Mean	1.938	Std err	.035	Median	1.000
Mode	1.000	Std dev	1.282	Variance	1.643
Kurtosis	9.524	S E Kurt	.134	Skewness	2.341
S E Skew	.067	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	2577.000		

Valid cases 1330 Missing cases 3541

CPSRN10 QUESTION ORDERING <CPSG6A,CPSG6B,CPSG6C>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
-------------	-------	-----------	---------	---------------	-------------

		1	1124	23.1	29.8	29.8
		2	1321	27.1	35.0	64.8
		3	1330	27.3	35.2	100.0
		.	1096	22.5	Missing	
Total			4871	100.0	100.0	
Mean	2.055	Std err	.013	Median	2.000	
Mode	3.000	Std dev	.805	Variance	.647	
Kurtosis	-1.449	S E Kurt	.080	Skewness	-.099	
S E Skew	.040	Range	2.000	Minimum	1.000	
Maximum	3.000	Sum	7756.000			

Valid cases 3775 Missing cases 1096

CPSSL9A VIEWS: UNEMPLOYMENT/INFLATION <CPSRN21>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
ReduceUnemploymt	1	1110	22.8	58.5	58.5
ControlInflation	2	574	11.8	30.2	88.7
D.K.	8	183	3.8	9.6	98.3
Refused	9	32	.7	1.7	100.0
	.	2972	61.0	Missing	
Total		4871	100.0	100.0	
Mean	2.112	Std err	.051	Median	1.000
Mode	1.000	Std dev	2.207	Variance	4.870
Kurtosis	3.543	S E Kurt	.112	Skewness	2.272
S E Skew	.056	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	4010.000		

Valid cases 1899 Missing cases 2972

CPSSL9B VIEWS: UNEMPLOYMENT/INFLATION <CPSRN21>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
ControlInflation	1	484	9.9	25.8	25.8
ReduceUnemploymt	2	1134	23.3	60.4	86.2
D.K.	8	226	4.6	12.0	98.3
Refused	9	32	.7	1.7	100.0
	.	2995	61.5	Missing	
Total		4871	100.0	100.0	
Mean	2.584	Std err	.052	Median	2.000
Mode	2.000	Std dev	2.257	Variance	5.092

Kurtosis	2.191	S E Kurt	.113	Skewness	1.961
S E Skew	.057	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	4848.000		

Valid cases 1899 Missing cases 2972

CPSRN21 QUESTION ORDERING <CPSL9A,CPSL9B>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	1899	39.0	50.3	50.3
	2	1876	38.5	49.7	100.0
	.	1096	22.5	Missing	
	Total	4871	100.0	100.0	

Mean	1.497	Std err	.008	Median	1.000
Mode	1.000	Std dev	.500	Variance	.250
Kurtosis	-2.001	S E Kurt	.080	Skewness	.012
S E Skew	.040	Range	1.000	Minimum	1.000
Maximum	2.000	Sum	5651.000		

Valid cases 3775 Missing cases 1096

CPSL10A VIEWS ON GOVERNMENT SERVICES <CPSRN8>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
NotProvideCanPay	1	852	17.5	46.2	46.2
Same To Everyone	2	913	18.7	49.5	95.7
D.K.	8	71	1.5	3.9	99.6
Refused	9	8	.2	.4	100.0
	.	3027	62.1	Missing	
	Total	4871	100.0	100.0	

Mean	1.799	Std err	.033	Median	2.000
Mode	2.000	Std dev	1.422	Variance	2.022
Kurtosis	13.887	S E Kurt	.114	Skewness	3.682
S E Skew	.057	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	3318.000		

Valid cases 1844 Missing cases 3027

CPSL10B VIEWS ON GOVERNMENT SERVICES <CPSRN8>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Same To Everyone	1	683	14.0	35.4	35.4
NotProvideCanPay	2	1160	23.8	60.1	95.4
D.K.	8	68	1.4	3.5	99.0
Refused	9	20	.4	1.0	100.0
	.	2940	60.4	Missing	
Total		4871	100.0	100.0	
Mean	1.930	Std err	.033	Median	2.000
Mode	2.000	Std dev	1.458	Variance	2.125
Kurtosis	13.428	S E Kurt	.111	Skewness	3.637
S E Skew	.056	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	3727.000		

Valid cases 1931 Missing cases 2940

CPSRN8 QUESTION ORDERING <CPSF10A,CPSF10B>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	1844	37.9	48.8	48.8
	2	1931	39.6	51.2	100.0
	.	1096	22.5	Missing	
Total		4871	100.0	100.0	
Mean	1.512	Std err	.008	Median	2.000
Mode	2.000	Std dev	.500	Variance	.250
Kurtosis	-1.999	S E Kurt	.080	Skewness	-.046
S E Skew	.040	Range	1.000	Minimum	1.000
Maximum	2.000	Sum	5706.000		

Valid cases 3775 Missing cases 1096

PESE4A REDUCE UNEMPLOY/REDUCE DEFICIT <PESRN2>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
ReduceUnemploymt	1	950	19.5	57.1	57.1
Reduce Deficit	2	590	12.1	35.5	92.6
D.K.	8	108	2.2	6.5	99.1
Refused	9	15	.3	.9	100.0
	.	3208	65.9	Missing	
Total		4871	100.0	100.0	
Mean	1.882	Std err	.045	Median	1.000

Mode	1.000	Std dev	1.827	Variance	3.339
Kurtosis	7.309	S E Kurt	.120	Skewness	2.912
S E Skew	.060	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	3129.000		

Valid cases 1663 Missing cases 3208

PESE4B REDUCE DEFICIT/REDUCE UNEMPLOY <PESRN2>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Reduce Deficit	1	642	13.2	38.3	38.3
ReduceUnemploymt	2	907	18.6	54.1	92.4
D.K.	8	108	2.2	6.4	98.8
Refused	9	20	.4	1.2	100.0
	.	3194	65.6	Missing	
	Total	4871	100.0	100.0	

Mean	2.087	Std err	.044	Median	2.000
Mode	2.000	Std dev	1.811	Variance	3.280
Kurtosis	6.935	S E Kurt	.119	Skewness	2.827
S E Skew	.060	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	3500.000		

Valid cases 1677 Missing cases 1531

PESRN2 QUESTION ORDERING <PESE4A, PESE4B>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	1663	34.1	49.8	49.8
	2	1677	34.4	50.2	100.0
	.	1531	31.4	Missing	
	Total	4871	100.0	100.0	

Mean	1.502	Std err	.009	Median	2.000
Mode	2.000	Std dev	.500	Variance	.250
Kurtosis	-2.001	S E Kurt	.085	Skewness	-.008
S E Skew	.042	Range	1.000	Minimum	1.000
Maximum	2.000	Sum	5017.000		

Valid cases 3340 Missing cases 1531

PESE5A VIEWS ON PENSIONS TO ELDERLY <PESRN3>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Pensions To All	1	608	12.5	35.7	35.7
No To Above Avg	2	1055	21.7	62.0	97.8
D.K.	8	28	.6	1.6	99.4
Refused	9	10	.2	.6	100.0
	.	3170	65.1	Missing	
	Total	4871	100.0	100.0	
Mean	1.782	Std err	.026	Median	2.000
Mode	2.000	Std dev	1.092	Variance	1.191
Kurtosis	25.774	S E Kurt	.119	Skewness	4.623
S E Skew	.059	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	3032.000		
Missing cases	1701	Missing cases	3170		

PESE5B VIEWS ON PENSIONS TO ELDERLY <PESRN3>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
No To Above Avg	1	824	16.9	50.3	50.3
Pensions To All	2	764	15.7	46.6	96.9
D.K.	8	40	.8	2.4	99.3
Refused	9	11	.2	.7	100.0
	.	3232	66.4	Missing	
	Total	4871	100.0	100.0	
Mean	1.691	Std err	.031	Median	1.000
Mode	1.000	Std dev	1.271	Variance	1.615
Kurtosis	19.294	S E Kurt	.121	Skewness	4.195
S E Skew	.060	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	2771.000		
Valid cases	1639	Missing cases	3232		

PESRN3 QUESTION ORDERING <PESE5A, PESE5B>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	1701	34.9	50.9	50.9
	2	1639	33.6	49.1	100.0
	.	1531	31.4	Missing	
	Total	4871	100.0	100.0	
Mean	1.491	Std err	.009	Median	1.000

Mode	1.000	Std dev	.500	Variance	.250
Kurtosis	-2.000	S E Kurt	.085	Skewness	.037
S E Skew	.042	Range	1.000	Minimum	1.000
Maximum	2.000	Sum	4979.000		

Valid cases 3755 Missing cases 1096

CPSG4 THINK CANADA GET BY WITHOUT GST<CPSRN5>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Yes, Get By	1	2247	46.1	59.5	59.5
No, Not Get By	5	1304	26.8	34.5	94.1
D.K.	8	215	4.4	5.7	99.8
Refused	9	9	.2	.2	100.0
	.	1096	22.5	Missing	
	Total	4871	100.0	100.0	

Mean	2.799	Std err	.037	Median	1.000
Mode	1.000	Std dev	2.288	Variance	5.233
Kurtosis	-.825	S E Kurt	.080	Skewness	.735
S E Skew	.040	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	10568.000		

Valid cases 3775 Missing cases 1096

CPSL6B PAY HIGHER TAXES MAINTAIN PGMS <CPSRN6>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Yes, Pay More	1	1365	28.0	36.2	36.2
No, Not Pay More	5	2222	45.6	58.9	95.0
Depends	7	110	2.3	2.9	97.9
D.K.	8	71	1.5	1.9	99.8
Refused	9	7	.1	.2	100.0
	.	1096	22.5	Missing	
	Total	4871	100.0	100.0	

Mean	3.676	Std err	.034	Median	5.000
Mode	5.000	Std dev	2.085	Variance	4.348
Kurtosis	-1.309	S E Kurt	.080	Skewness	-.297
S E Skew	.040	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	13876.000		

Valid cases 3775 Missing cases 1096

CPSRN6 QUESTION WORDING <CPSL6B>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	1242	25.5	32.9	32.9
Such As HlthCare	2	1285	26.4	34.0	66.9
Such As Welfare	3	1248	25.6	33.1	100.0
	.	1096	22.5	Missing	
	Total	4871	100.0	100.0	
Mean	2.002	Std err	.013	Median	2.000
Mode	2.000	Std dev	.812	Variance	.660
Kurtosis	-1.484	S E Kurt	.080	Skewness	.003
S E Skew	.040	Range	2.000	Minimum	1.000
Maximum	3.000	Sum	7556.000		
Valid cases	3775	Missing cases	1096		

CPSL11A VIEWS:HEALTH CARE SYSTEM<CPSRN3,RN4,RN9>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Less Waste	1	842	17.3	45.3	45.3
Not Get HlthCare	2	925	19.0	49.8	95.2
D.K.	8	80	1.6	4.3	99.5
Refused	9	10	.2	.5	100.0
	.	3014	61.9	Missing	
	Total	4871	100.0	100.0	
Mean	1.843	Std err	.035	Median	2.000
Mode	2.000	Std dev	1.498	Variance	2.245
Kurtosis	12.192	S E Kurt	.114	Skewness	3.503
S E Skew	.057	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	3422.000		
Valid cases	1857	Missing cases	3014		

CPSL11B VIEWS:HEALTH CARE SYSTEM<CPSRN3,RN4,RN9>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Not Get HlthCare	1	597	12.3	31.1	31.1
Less Waste	2	1228	25.2	64.0	95.2
D.K.	8	79	1.6	4.1	99.3
Refused	9	14	.3	.7	100.0
	.	2953	60.6	Missing	

		Total	4871	100.0	100.0
Mean	1.987	Std err	.034	Median	2.000
Mode	2.000	Std dev	1.467	Variance	2.153
Kurtosis	12.546	S E Kurt	.112	Skewness	3.543
S E Skew	.056	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	3811.000		
Valid cases	1918	Missing cases	2953		

CPSRN3 QUESTION WORDING <CPSL11A,CPSL11B>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Doctor	1	1878	38.6	49.7	49.7
Hospital	2	1897	38.9	50.3	100.0
	.	1096	22.5	Missing	
Total		4871	100.0	100.0	
Mean	1.503	Std err	.008	Median	2.000
Mode	2.000	Std dev	.500	Variance	.250
Kurtosis	-2.001	S E Kurt	.080	Skewness	-.010
S E Skew	.040	Range	1.000	Minimum	1.000
Maximum	2.000	Sum	5672.000		
Valid cases	3775	Missing cases	1096		

CPSRN4 QUESTION WORDING <CPSL11A,CPSL11B>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Low Income	1	1809	37.1	47.9	47.9
Some	2	1966	40.4	52.1	100.0
	.	1096	22.5	Missing	
Total		4871	100.0	100.0	
Mean	1.521	Std err	.008	Median	2.000
Mode	2.000	Std dev	.500	Variance	.250
Kurtosis	-1.994	S E Kurt	.080	Skewness	-.083
S E Skew	.040	Range	1.000	Minimum	1.000
Maximum	2.000	Sum	5741.000		
Valid cases	3775	Missing cases	1096		

REFE6 SHLD WOMEN BE GUARANTEED SEATS IN SENATE

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Yes	1	632	13.0	60.5	60.5
No	5	339	7.0	32.5	93.0
Abolish Senate	7	12	.2	1.1	94.2
D.K.	8	59	1.2	5.7	99.8
Refused	9	2	.0	.2	100.0
	.	3827	78.6	Missing	
Total		4871	100.0	100.0	
Mean	2.779	Std err	.072	Median	1.000
Mode	1.000	Std dev	2.311	Variance	5.342
Kurtosis	-.804	S E Kurt	.151	Skewness	.773
S E Skew	.076	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	2901.000		

Valid cases 1044 Missing cases 3827

REFE7 WOMEN BE GUARANTEED HALF SEATS IN SENATE

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Yes	1	405	8.3	45.9	45.9
No	5	421	8.6	47.7	93.7
Abolish Senate	7	12	.2	1.4	95.0
D.K.	8	41	.8	4.6	99.7
Refused	9	3	.1	.3	100.0
	.	3989	81.9	Missing	
Total		4871	100.0	100.0	
Mean	3.344	Std err	.076	Median	5.000
Mode	5.000	Std dev	2.266	Variance	5.136
Kurtosis	-1.283	S E Kurt	.164	Skewness	.193
S E Skew	.082	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	2949.000		

Valid cases 882 Missing cases 3989

REFE8 ABORIGINAL BE GUARANTEED SEATS IN SENATE

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Yes	1	396	8.1	65.6	65.6

No	5	151	3.1	25.0	90.6
Abolish Senate	7	13	.3	2.2	92.7
D.K.	8	40	.8	6.6	99.3
Refused	9	4	.1	.7	100.0
.		4267	87.6	Missing	
Total		4871	100.0	100.0	

Mean	2.646	Std err	.098	Median	1.000
Mode	1.000	Std dev	2.397	Variance	5.745
Kurtosis	-.396	S E Kurt	.199	Skewness	1.020
S E Skew	.099	Range	8.000	Minimum	1.000
Maximum	9.000	Sum	1598.000		

Valid cases 604 Missing cases 4267

REFRN14 QUESTION WORDING <REFE6,REFE7,REFE8>

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	900	18.5	35.6	35.6
	2	961	19.7	38.0	73.6
	3	669	13.7	26.4	100.0
	.	2341	48.1	Missing	
Total		4871	100.0	100.0	

Mean	1.909	Std err	.016	Median	2.000
Mode	2.000	Std dev	.782	Variance	.612
Kurtosis	-1.350	S E Kurt	.097	Skewness	.161
S E Skew	.049	Range	2.000	Minimum	1.000
Maximum	3.000	Sum	4829.000		

Valid cases 2530 Missing cases 2341

Preceding task required 9.58 seconds CPU time; 18.64 seconds elapsed.

8 0 finish.

8 command lines read.
0 errors detected.
0 warnings issued.
10 seconds CPU time.
21 seconds elapsed time.
End of job.

Bibliography

- Abramson, Paul R. and Charles W. Ostrom. "Question Wording and Partisanship: Change and Continuity in Party Loyalties During the 1992 Election Campaign." *Public Opinion Quarterly* 58 (1994): 21-48.**
- Alspach, Steven E. and George F. Bishop. "Question-Order Effects of Presidential Approval Ratings on Gubernatorial Approval Ratings: A Research Note." *Social Forces* 69, no. 4 (1991): 1241 - 1248.**
- Alwin, Duane F. and Jon A. Krosnick. "The Reliability of Survey Attitude Measurement." *Sociological Methods and Research* 20 (1991): 139 - 175.**
- Anderson, Barbara, Brian D. Silver, and Paul R. Abramson. "The Effects of the Race of the Interviewer on Race-Related Attitudes of Black Respondents in SRC/CPS National Election Studies." *Public Opinion Quarterly* 52 (1988): 289 - 321.**
- Andrews, Frank M. "Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach." 48 *Public Opinion Quarterly* 48 (1984): 409 - 442.**
- Asher, Herbert B. "Some Consequences of Measurement Error in Survey Data." *American Journal of Political Science* 18 (1974): 469 - 485.**
- Ayidiya, Stephen A. and McKee J. McClendon. "Response Effects in Mail Surveys." *Public Opinion Quarterly* 54 (1990): 229-247.**
- Bagozzi, Richard P. and Claes Fornell, "Theoretical Concepts, Measurements, and Meaning." In *A Second Generation of Multivariate Analysis, Volume 2: Measurement and Evaluation*. Edited by Claes Fornell. New York: Praeger Publishers, 1982.**
- Bailar, Barbara, Leroy Bailey, and Joyce Stevens. "Measures of Interviewer Bias and Variance." *Journal of Marketing Research* 14 (1977): 337 - 343.**

- Bailar, Barbara, and Paul P. Biemer. "Some Methods for Evaluating Nonsampling Error in Household Censuses and Surveys." In *W.G. Cochran's Impact on Statistics*. Edited by Poduri S.R.S. Rao and Joseph Sedransk. New York: John Wiley & Sons, 1984.
- Barath, Arpad and Charles F. Cannell. "Effect of Interviewer's Voice Intonation." *Public Opinion Quarterly* 40 (1976): 370 - 373.
- Back, Kurt W. and Kenneth J. Gergen. "Idea Orientation and Ingratiation in the Interview: A Dynamic Model of Response Bias." *American Statistical Association: Proceedings of the Social Statistics Section* (1963): 284 - 288.
- Bennett, Scott. *The Quantitative Approach in Political Science: An Introduction*. Queenston: The Edwin Mellen Press, 1986.
- Benton, J. Edwin and John L. Daly. "Question Order Effect in a Local Government Survey." *Public Opinion Quarterly* 55 (1991): 640 - 642.
- Bickart, Barbara A. "Carryover and Backfire Effects in Marketing Research." *Journal of Marketing Research* 30 (1993): 52-62.
- Bickart, Barbara A. "Question-Order Effects and Brand Evaluations: The Moderating Rule of Consumer Knowledge." In *Context Effects in Social and Psychological Research*. Edited by Norman Schwarz and Seymour Sudman. New York: Springer-Verlag, 1992.
- Biemer, Paul P., Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman. *Measurement Errors in Surveys*. New York: John Wiley & Sons, 1991.
- Bishop, George F. "Context Effects on Self-Perceptions of Interest in Government and Public Affairs." In *Social Information Processing and Survey Methodology*. Edited by Hans-J. Hippler, Norbert Schwarz and Seymour Sudman. New York: Springer-Verlag, 1987.

- Bishop, George F. "Experiments with the Middle Response Alternative in Survey Questions." *Public Opinion Quarterly* 51 (1987): 220-232.
- Bishop, George, F. "Issue Involvement and Response Effects in Public Opinion Surveys." *Public Opinion Quarterly* 54 (1990): 209 - 218.
- Bradburn, Norman M. and Carrie Miles. "Vague Quantifiers." *Public Opinion Quarterly* 43 (1979): 92-101.
- Bradburn, Norman M. and William M. Mason. "The Effect of Question Order on Responses." *Journal of Marketing Research* 1 (1964): 57 - 61.
- Bradburn, Norman M. and Seymour Sudman. "The Current Status of Questionnaire Design." In *Measurement Errors in Surveys*. Edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz and Seymour Sudman. New York: John Wiley & Sons, 1991.
- Cotter, Patrick R., Jeffrey Cohen, and Philip B. Coulter. "Race-of-Interviewer Effects in Telephone Interviews." *Public Opinion Quarterly* 46 (1982): 278 -284.
- DeLamater, John. "Response-Effects of Question Content." In *Response Behaviour in the Survey Interview*. Edited by W. Dijkstra and J. VanderZouwen. Toronto: Academic Press, 1982.
- Fee, Joan Flynn. "Symbols in Survey questions: Solving the Problems of Multiple Word Meanings." *Political Methodology* 7 (1981):71 - 91.
- Feldman, J.J., Herbert Hyman, and C.W. Hart. "A Field Study of Interviewer Effects on the Quality of Survey Data." *Public Opinion Quarterly* 15 (1951-52). 734 - 761.
- Fowler, Floyd J. Jr. *Survey Research Methods, Second Ed.* New York: Sage Publications, 1993.

- Groves, Richard M. and Nancy H. Fultz. "Gender Effects Among Telephone Interviewers in a Survey of Economic Attitudes." *Sociological Methods & Research* 14 (1985): 31 - 52.
- Groves, Robert M. *Survey Errors and Survey Costs*. New York: John Wiley & Sons, 1989.
- Hanson, Robert H. and Eli S. Marks. "Influence of the Interviewer on the Accuracy of Survey Results." *Journal of the American Statistical Association* 53 (1958): 635 - 655.
- Hatchett, Shirley and Howard Schuman. "White Respondents and Race-of-Interviewer Effects." *Public Opinion Quarterly* 39 (1975/6): 523 - 528.
- Henerson, Marlene E., Lynn Lyons Morris and Carol Taylor Fitz-Gibbon. *How to Measure Attitudes*. New York: Sage Publications, 1978.
- Hippler, Hans-J. and Norbert Schwarz. "Response Effects in Surveys." In *Social Information Processing and Survey Methodology*. Edited by Hans-J. Hippler, Norbert Schwarz and Seymour Sudman. New York: Springer-Verlag, 1987.
- Kalton, Graham, Martin Collins and Lindsay Brook. "Experiments in Wording Opinion Questions." *Applied Statistics* 27, no. 2 (1978): 149 - 161.
- Kish, Leslie. "Studies of Interviewer Variance for Attitudinal Variables." *Journal of the American Statistical Association* 57(1962): 92 - 115.
- Laurent, Andre. "Effects of Question Length on Reporting Behavior in the Survey Interview." *Journal of the American Statistical Association* 67 (1972): 298 - 305.
- Lessler, Judith T. and William D. Kalsbeek. *Nonsampling Error in Surveys*. New York: John Wiley & Sons, 1992.

- Matell, Michael S. and Jacob Jacoby. "Is There an Optimal Number of Alternatives for Likert Scale Items? Study I: Reliability and Validity." *Educational and Psychological Measurement* 31 (1971): 657-74.
- Molenaar, Nico J. "Response-Effects of 'Formal' Characteristics of Questions." In *Response Behaviour in the Survey Interview*. Edited by W. Dijkstra and J. VanderZouwen. Toronto: Academic Press, 1982.
- Morton-Williams, Jean and Wendy Sykes. "The Use of Interaction Coding and Follow-up Interviews to Investigate Comprehension of Survey Questions." *Journal of the Market Research Society* 26, no. 2 (1984): 109-127.
- Noelle-Neumann, Elisabeth. "Wanted: Rules for Wordng Structured Questions." *Public Opinion Quarterly* 34 (1970/71): 191 - 201.
- Northrup, David A. and Anne E. Oram, *The 1993 Canadian Election Study, Incorporating the 1992 Referendum Survey on the Charlottetown Accord*. North York: Institute for Social Research, York University, 1994.
- Pepper, Susan and Lubomir S. Prytulak. "Sometimes Frequently Means Seldom: Context Effects in the Interpretation of Quantitative Expressions." *Journal of Research in Personality* 8 (1974): 95 - 101.
- Rasinski, Kenneth A. "The Effect of Question Wording on Public Support for Government Spending." *Public Opinion Quarterly* 53 (1989): 388 - 394.
- Reinecke, Jost and Peter Schmidt. "Explaining Interviewer Effects and Respondent Behavior: Theoretical Models and Empirical Analysis." *Quality and Quantity* 27 (1993): 219 - 247.
- Sanchez, Maria Elena, "Effects of Questionnaire Design on the Quality of Survey Data." *Public Opinion Quarterly* 56 (1992): 206-217.

- Schuman, Howard. "Context Effects: State of the Past/State of the Art." In *Context Effects in Social and Psychological Research*. Edited by Norman Schwarz and Seymour Sudman. New York: Springer-Verlag, 1992.
- Schuman, Howard, Graham Kalton, and Jacob Ludwig. "Context and Contiguity in Survey Questionnaires." *Public Opinion Quarterly* 47 (1983): 112-115.
- Schuman, Howard, Stanley Presser, and Jacob Ludwig. "Context Effects on Survey Responses to Questions About Abortion." *Public Opinion Quarterly* 45 (1981): 216-223.
- Schuman, Howard. "Ordinary Questions, Survey Questions, and Policy Questions." *Public Opinion Quarterly* 50 (1986): 432 - 442.
- Schwarz, Norbert and Hans-J. Hippler. "What Response Scales May Tell your Respondents: Informative Functions of Response Alternatives." In *Social Information Processing and Survey Methodology*. Edited by Hans-J. Hippler, Norbert Schwarz and Seymour Sudman. New York: Springer-Verlag, 1987.
- Singer, Eleanor, Martin R. Frankel and Marc B. Glassman. "The Effect of Interviewer Characteristics and Expectations on Response." *Public Opinion Quarterly* 47 (1983): 68 - 83.
- Smith, Tom W. "In Search of House Effects: A Comparison of Responses to Various Questions by Different Survey Organizations." *Public Opinion Quarterly* 42 (1978): 443 - 459.
- Smith, Tom W. "That Which We Call Welfare By Any Other Name Would Smell Sweeter: An Analysis of the Impact of Question Wording on Response Patterns." *Public Opinion Quarterly* 51 (1987): 75-83.
- Strack, Fritz. "'Order Effects' in Survey Research: Activation and Information Functions of Preceding Questions." In *Context Effects in Social and Psychological Research*.

Edited by Norman Schwarz and Seymour Sudman. New York: Springer-Verlag, 1992.

Strack, Fritz and Leonard L. Martin. "Thinking, Judging, and Communicating: A process Account of Context Effects in Attitude Surveys." In *Social Information Processing and Survey Methodology*. Edited by Hans-J. Hippler, Norbert Schwarz and Seymour Sudman. New York: Springer-Verlag, 1987.

Tamai, John and Don A. Dillman. "Questionnaire Context as a Source of Response Differences in Mail and Telephone Surveys." In *Context Effects in Social and Psychological Research*. Edited by Norman Schwarz and Seymour Sudman. New York: Springer-Verlag, 1992.

Tourangeau, Roger. "Context Effects on Responses to Attitude Questions: Attitudes as Memory Structures." In *Context Effects in Social and Psychological Research*. Edited by Norman Schwarz and Seymour Sudman. New York: Springer-Verlag, 1992.

Tucker, Clyde. "Interviewer Effects in Telephone Surveys." *Public Opinion Quarterly* 47 (1983): 84 -95.

END

0|4-06-9|6

FIN