

1-1-2001

## Measuring the effects of access control, confidentiality statements, and results display in Web-based marketing research surveys

Arthur Wilson Peoples  
*Iowa State University*

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>

### Recommended Citation

Peoples, Arthur Wilson, "Measuring the effects of access control, confidentiality statements, and results display in Web-based marketing research surveys" (2001). *Retrospective Theses and Dissertations*. 21472.

<https://lib.dr.iastate.edu/rtd/21472>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

Measuring the effects of access control, confidentiality statements,  
and results display in Web-based marketing research surveys

by

Arthur Wilson Peoples

A thesis submitted to the graduate faculty  
in partial fulfillment of the requirements for the degree of  
MASTER OF SCIENCE

Major: Business

Major Professor: Thomas E. DeCarlo

Iowa State University

Ames, Iowa

2001

Graduate College  
Iowa State University

This is to certify that the Master's thesis of

Arthur Wilson Peoples

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy

To Eric, Sara, and Robert Jackman;  
and to my sons Jeffrey and Daniel Peoples

## TABLE OF CONTENTS

<b>LIST OF FIGURES</b>	vi
<b>LIST OF TABLES</b>	vii
<b>NOMENCLATURE</b>	ix
<b>ABSTRACT</b>	xi
<b>CHAPTER 1. STATEMENT OF RESEARCH PROBLEM</b>	1
<b>CHAPTER 2. LITERATURE REVIEW</b>	3
Introduction	3
Privacy Issues	3
Privacy and Web-based Research	5
Improving Response Rate	9
<b>CHAPTER 3. MATERIALS AND METHODS</b>	11
Survey Development and Design	11
Experimental Design for Testing Hypothesis 1	12
Sample and Procedure for Testing Hypothesis 1	13
Preventing Duplicate Submissions	14
Controlling the Sample	16
Experimental Design for Testing Hypotheses 2 and 3	19
Sample and Procedure for Testing Hypotheses 2 and 3	20
<b>CHAPTER 4. RESULTS AND DISCUSSION</b>	21
Hypothesis 1A	21
Hypothesis 1B	23
Identification Follow-up Question	24
Hypothesis 2A	27
Hypothesis 2B	29
Hypothesis 3	31
Internal and External Validity	33

<b>CHAPTER 5. CONCLUSIONS</b>	35
Access Control	35
Confidentiality Statements	36
Results Display	36
<b>APPENDIX A. PILOT TEST RESULTS</b>	37
<b>APPENDIX B. E-MAIL INVITATIONS</b>	39
<b>APPENDIX C. SURVEY HOME PAGES</b>	42
<b>APPENDIX D. SURVEY INSTRUMENT</b>	45
<b>APPENDIX E. FOLLOW-UP QUESTIONNAIRES</b>	48
<b>APPENDIX F. RESULTS DISPLAY</b>	50
<b>REFERENCES CITED</b>	55

## LIST OF FIGURES

Figure 3.1 Survey Model with Access Control ID	15
Figure 3.2 Survey Model without Access Control ID	17
Figure B1. Access ID only E-mail Invitation	39
Figure B2. No Access ID E-mail Invitation	39
Figure B3. Results Display E-mail Invitation	40
Figure B4. Confidentiality Statement E-mail Invitation	40
Figure B5. Confidentiality Statement and Results Display E-mail Invitation	41
Figure C1. Access ID only Survey Home Page	42
Figure C2. No Access ID Survey Home Page	42
Figure C3. Results Display Survey Home Page	43
Figure C4. Confidentiality Statement Survey Home Page	43
Figure C5. Confidentiality Statement and Results Display Survey Home Page	44
Figure D. Main Questionnaire	45
Figure E1. Access ID only and No Access ID Follow-up Questionnaire	48
Figure E2. Results Display Follow-up Questionnaire	48
Figure E3. Confidentiality Statement Follow-up Questionnaire	49
Figure E4. Confidentiality Statement and Results Display Follow-up Questionnaire	49
Figure F. Results Display Graph	50

## LIST OF TABLES

Table 4.1 Descriptive Statistics for Access Control Response Rate	21
Table 4.2 Estimates for Access Control Response Rate	21
Table 4.3 ANOVA for Access Control Response Rate	22
Table 4.4 Pairwise Comparison for Access Control Response Rate	22
Table 4.5 Descriptive Statistics for Access Control Missing Data Rate	23
Table 4.6 Estimates for Access Control Missing Data Rate	23
Table 4.7 ANOVA for Access Control Missing Data Rate	24
Table 4.8 Pairwise Comparison for Access Control Missing Data Rate	24
Table 4.9 Descriptive Statistics for Access Control Identification Follow-up Question	26
Table 4.10 Estimates for Access Control Identification Follow-up Question	26
Table 4.11 ANOVA for Access Control Identification Follow-up Question	27
Table 4.12 Pairwise Comparison for Access Control Identification Follow-up Question	27
Table 4.13 Descriptive Statistics for Confidentiality Statement Response Rate	28
Table 4.14 Estimates for Confidentiality Statement Response Rate	28
Table 4.15 ANOVA for Confidentiality Statement Response Rate	28
Table 4.16 Pairwise Comparison for Confidentiality Statement Response Rate	29



Table 4.17 Descriptive Statistics for Confidentiality Statement Missing Data Rate	30
Table 4.18 Estimates for Confidentiality Statement Missing Data Rate	30
Table 4.19 ANOVA for Confidentiality Statement Missing Data Rate	31
Table 4.20 Pairwise Comparison for Confidentiality Statement Missing Data Rate	31
Table 4.21 Descriptive Statistics for Results Display Response Rate	32
Table 4.22 Estimates for Results Display Response Rate	32
Table 4.23 ANOVA for Results Display Response Rate	33
Table 4.24 Pairwise Comparison for Results Display Response Rate	33
Table A1. First Pilot Test	37
Table A2. Second Pilot Test	38

## NOMENCLATURE

Access control	A method to ensure that only authorized persons can access a Web site.
Access control ID	A set of numbers and/or characters that are entered into a text box on the login screen of a Web site. The ID is checked against an ID list, and if there is a match the person is allowed to access the Web site.
Active Server Pages	A Web server technology that allows developers to build advanced functionality into Web pages such as reading and writing to a database.
Cookie	A text file with information that a Web server places on a user's computer when the user visits a Web site. When the user visits the Web site again, the information in the cookie is read by the server.
Internet protocol address	The address of a computer connected to the Internet that is placed in a request to or response from another computer telling the data to what computer on the Internet it should be directed.
Results display	A graphical display on a Web page of the cumulative results of all participants in a Web-based survey.
Robots meta tag	A HTML tag placed on a Web page that instructs

search engines not to list that Web site in any of its search results.

Search engine

A Web site that produces a list of links to Web sites that include part or all of a user's search criteria.

Web-based survey

A survey posted on the Internet or World Wide Web.

## ABSTRACT

The purpose of this research is to examine three important issues that affect the validity of Web-based marketing research surveys. Questionnaires posted on the Internet are the quickest, most efficient method of gathering survey data, but their validity is questionable. The issues examined in the study are as follows:

1. Sample control using access control IDs is investigated to determine the effect on response rate and quality. Quality in the study is measured by the percentage of unanswered questions.

2. The effect of confidentiality statements on response rate and quality is examined.

3. Increasing motivation for user participation through the promise and display of current cumulative results is investigated. The variable response rate is used to measure the effect on motivation.

Data was gathered through a set of marketing research surveys posted on the World Wide Web. Subjects were invited to participate through personalized e-mail invitations. The analysis suggests that the researcher can employ access control IDs for controlling the sample without a significant loss of response rate and quality caused by users' concerns over identification with their personal information. The study shows that posting a confidentiality statement to alleviate consumers' concerns over identification with and use of their personal information has no significant effect on response rate and quality. Finally, the research demonstrates a

lack of evidence that promising the display of survey results immediately after the user's completion of Web-based surveys increases response rates.

## CHAPTER 1. STATEMENT OF RESEARCH PROBLEM

The marketing research survey conducted on the World Wide Web provides an efficient, inexpensive tool for market researchers to quickly gather information necessary to plan marketing strategies for their products, services, and ideas. Web-based surveys are employed by numerous marketing research companies such as Burke Marketing Research at [www.burke.com](http://www.burke.com), NFO Interactive at [www.nfoi.com/nfointeractive](http://www.nfoi.com/nfointeractive), The NPD Group at [www.npd.com](http://www.npd.com), and Greenfield Online at [www.greenfiled.com](http://www.greenfiled.com). However, certain validity problems need investigation and resolution before this marketing research tool can give researchers confidence in its results (Cleland, 1999; McCullough, 1998; Stanton, 1998; Schillewaert, Langerak, and Duhamel 1998). One problem, in particular, that has been overlooked is the issue of privacy and confidentiality. The dearth of research in this important topic is surprising given that one of the greatest issues with the Internet and World Wide Web is that of the privacy of the user. Perceptions of privacy have had a significant influence on previous Web-based studies. For example, responses to computer-based attitude perceptions are likely to vary depending upon respondents' computer-based beliefs about anonymity (Kantor 1991). Moreover, Stanton (1998) points out that using access controls raises troubling anonymity problems. In the present study, the lack-of-anonymity effect is defined as the bias that results from respondents' belief that they can be identified with the information they provide in the survey. The respondent's lack of anonymity in a Web-based survey that requires the user to enter an access control ID,

therefore, is a problem that needs investigation to determine the effect on response rate and quality. Research into the effect of posting a privacy and confidentiality statement on a Web-based survey with an access control ID requirement is another area that needs examination. Improvement in the response rate of a Web-based survey that employs interactive functions, such as promising and displaying current results of the survey, also warrants exploration. The purpose of this study, therefore, is to compare Web-based surveys that contain access control IDs, privacy and confidentiality statements, and results display to those that do not contain these features.

## CHAPTER 2. LITERATURE REVIEW

### *Introduction*

Various studies have compared Web-based research survey results with more traditional paper and pencil survey results (e.g. Stanton, 1998; Huei-Fen, 1998; Zhang, 2000). In general, these studies suggest that the Web-based survey mode may provide the same quality and useful data as the paper and pencil mode. While the issue of privacy was not explicitly addressed in previous research, we can gain insight into the usefulness of Web-based surveys for marketing research from a brief examination of these studies. First though, we will examine the concept of privacy.

### *Privacy Issues*

From a legal perspective, information practices that gather information to generalize across groups of consumers where the focus is on market segments and not individual consumers would have a reduced applicability regarding the privacy concept. However, information-gathering practices that involve individual-level consumer information (i.e., personal information that can be identified with a single named person) are important legal issues (e.g., Nowak and Phelps 1997).

At the core of these issues are consumers' concerns about the use of their personal information. In a recent study of consumers' perceptions of marketing information practices in 17 industries, direct marketers were the least trusted because of their propensity to capture and share information with third parties. The more trusted organizations included those with very little to sell and not having a



profit motive (Milne, Boza, 1999). In a recent Graphic, Visualization, and Usability (GVU) poll, 72 percent of the respondents favored new online privacy protection legislation while 82 percent disapproved of the sale of their personal information. Another poll taken by Business Week showed that 53 percent of the people questioned, favored legislation governing the way personal information can be collected and used online. That was triple the percent of respondents that favored self-regulation by trade groups. Despite these findings, the importance of the issue to consumers has not been appreciated by most marketers. For example, Geocities, a well known personal Web site host, settled with the FTC, who charged them with misleading members by selling their personal information to marketers without their consent (Hagel and Singer, 1999). Thus, consumers are likely to be suspicious of marketer-generated Web-based surveys, causing an increased potential for the lack-of-anonymity effect.

One technique marketers could use to reduce the lack-of-anonymity effect is to assure the respondent of complete anonymity. However, in the case of the Web-survey that does not identify the respondent, the survey can be submitted more than once by the same person and bias the results. In a 1999 study on Internet survey research, requiring a unique case ID eliminated this problem (Zhang, 2000). Unique case IDs are also used to control sampling for Web-based surveys, however, the potential for elevating respondents' suspicions about anonymity and confidentiality exists (Stanton 1998).

## ***Privacy and Web-based Research***

Recent studies have examined the validity of Web-based research surveys in areas such as university course evaluation research (Huei-Fen, 1998), organizational research (Stanton, 1998), and library information science research (Zhang, 2000). These studies illustrate some unique challenges to Web-based surveys.

A 1998 study comparing Web-based course evaluation surveys to traditional pencil and paper course evaluations found the response rate of the Web surveys lower than the paper and pencil surveys. The same study found higher quality of responses in the Web-based surveys as measured by the number of evaluations that included comments and the number of words included in each comment (Huei-Fen, 1998). Although the study did not offer this as an explanation for the lower response rate in the Web-based version, the fact that students were required to enter their Student ID number before completing the evaluation should certainly be considered a factor in the lower response rate. The students' lack of anonymity may have caused some who would want to criticize an instructor to decline participating in the Web-based survey.

In the organizational research study noted above, data collected using a Web-based instrument showed comparable item variability and similar internal covariance patterns with data collected via a paper and pencil mode suggesting an equivalence in data quality. Results of the study also showed significantly fewer missing data fields in the Web version. Evidence that factor structure of items forming a scale would not differ across the two survey formats was also found as

well as evidence that correlations between scales were the same between the two modes (Stanton, 1998). In the report, the analysis intentionally discarded the mean differences between the groups because access controls were not used and there was no straightforward way of controlling who responded to the Web-based survey. In addition, no individual identifying information was gathered in either version of the questionnaire. The subjects in Stanton's (1998) study would have had a desire for anonymity since the instrument was designed to "explore the determinants of individuals' perceptions of fairness in their day to day interactions with their supervisors" (p. 714). Because access control IDs were not used for the Web version, the respondents most likely felt their responses were anonymous just as in the paper and pencil version. This could explain the equivalence in the data quality. Had access control IDs been employed in the Web version, the data quality may not have been equivalent if the subjects perceived a lack of anonymity. Kantor (1991), for example, found that identified subjects scored higher on questions about supervisor satisfaction on the Job Descriptive Index survey than partially identified (subjects that were asked demographic questions that they thought might be used to identify them) and unidentified respondents. Thus, comparison of a Web-based survey with and without an access control ID requirement could be used to measure the lack-of-anonymity effect. This effect could influence a respondent's willingness to participate or complete all questions in a marketing survey that contains sensitive questions, such as personal finances, age, occupation, marital status, or personal buying behavior and preferences.

To investigate for the lack-of-anonymity effect, using data collected from a Web-based survey with an access control ID requirement compared to data from the same survey without an access control ID requirement, the following hypotheses can be expressed (a discussion of how, during an experiment, duplicate submissions are eliminated and sampling is controlled for the Web survey without an access control ID requirement, is presented in chapter three):

H1a: The response rate of a Web-based marketing research survey administered with an access control ID requirement is significantly less than the response rate of the same survey administered without an access control ID requirement.

H1b: The mean percentage of missing data fields in a marketing research survey with an access control ID requirement is significantly greater than the mean percentage of missing data fields in the same survey administered without an access control ID requirement.

If Access control IDs do affect response rate because of a lack of anonymity, a method for countering this effect would be needed to assure validity of the results. Including a statement of privacy and confidentiality would seem to be the obvious approach, but would this remove the lack-of-anonymity effect? Research has shown differing results on the effectiveness of including these statements to gain higher levels of cooperation in surveys (Rasinski, Willis, Baldwin, Yeh, and Lee, 1999).

Singer, Hippler, and Norbert (1992) found that detailed assurance of confidentiality raised respondents' apprehension about surveys that actually did not contain sensitive questions. They called for future studies using sensitive questions to examine the effects of confidentiality assurances on respondent participation.

In an exploratory research study by the Panel on Privacy and Confidentiality as Factors in Survey Response, a majority of respondents felt the inclusion of a promise of confidentiality would affect the accuracy of responses on questions about finances on census questionnaires. However, participants in small-group discussions were cynical about assurances of confidentiality in the same type of surveys. In a response behavior survey in the same study, questionnaires were administered by interview. The experiment failed to show that a promise of confidentiality was a major factor in determining response or nonresponse. However, it did show that confidentiality is a response factor for some people (National Research Council (U.S.) Panel on Privacy and Confidentiality as Factors in Survey Response, 1979).

Although the effect of privacy and confidentiality statements is unclear, a test to determine whether they would counter any lack-of-anonymity effect in surveys with an access control ID requirement could further help in discovering ways to improve the validity of Web-based marketing research surveys. Based on the above discussion, the following hypotheses are offered:

H2a: The response rate of a Web-based marketing research survey administered with a posted privacy and confidentiality statement (both

in the solicitation e-mail and on the Web site) is significantly greater than the response rate of the same survey administered without a posted privacy and confidentiality statement.

H2b: The mean percentage of missing data fields in a Web-based marketing research survey administered with a posted privacy and confidentiality statement (both in the solicitation e-mail and on the Web site) is significantly less than the mean percentage of missing data fields in the same survey administered without a privacy and confidentiality statement.

### ***Improving Response Rate***

Web-based surveys have the potential for improving response rate and quality through providing motivation for participation through dynamic interactive functions (Schmidt, 1997). However, these functions have been used very little in previous studies because programming them requires significant technical expertise. Therefore, their effect on survey results is presently unclear (Zhang, 2000). Research into the response rates of postal mail surveys has shown evidence that informing subjects of the survey results encourages participation (Mertens, 1998). For example, in a case study of research into scholarly use of Internet-based electronic resources, participants were given a choice of responding via a Web-based questionnaire or by mail. In both cases, respondents were given the option of reviewing the results of the survey as well as their own replies. 62.2% of the sample replied via the Web and 15.4% via mail. Of those using the Web survey, 26.4%

viewed the overall survey results, while an additional 9.6% chose to view the overall results and their own replies. Of those that participated via mail, none chose to receive their own replies or the overall survey results. (Zhang, 2000). These findings strongly suggest that those for whom reviewing results is a motivation, would rather receive them immediately after completion of the survey than wait for them to arrive at a later date. It is likely that Web-based surveys would also benefit from results display. A possible reason for people being motivated by results display could be the desire to see how they compare with others, or just to see how others feel about particular topics.

In sum, the notion that the promise and display of survey results improves the rate of response by comparing a Web survey with and without results display is expressed by the following hypothesis:

H3: The response rate of a Web-based marketing research survey that promises and displays the current cumulative results from all other respondents after completion of the survey is significantly greater than the response rate of the same survey that does not promise and display the results.

## CHAPTER 3. MATERIALS AND METHODS

### *Survey Development and Design*

A Web-based marketing survey questionnaire on financial services served as the instrument to gather the data to test the hypotheses. Small group discussions conducted by the National Research Council (U.S.) Panel on Privacy and Confidentiality as Factors in Survey Response (1979) revealed that questions about income and other financial matters were considered private and inappropriate for inclusion in surveys. Based on this information, a pilot test questionnaire was developed that included questions on people's personal finances. A pilot test survey for measuring the clarity and sensitivity of the proposed questions was administered to four Iowa State University graduate students. The first section of proposed questions involved consumers' opinions regarding the perceived importance of the benefits provided by their banks and insurance companies. Also included were questions that asked what companies they patronized, whether they transacted with their financial institutions on the Internet, and if so, what software they used. The second section asked various demographic questions.

The results suggested some needed modifications to the instrument. The students did not consider the questions about the benefits of the subject's financial institutions at all sensitive. Based on these results, a revised questionnaire was developed and subsequently tested on four different Iowa State University graduate students. The second pilot test indicated that the most sensitive questions asked the amount of money consumers had in their various financial accounts (savings,



checking, investments, etc.) and what credit cards they carried. The final survey instrument was developed by adding these questions. Additionally, half of the questions that involved consumers' opinions regarding the perceived importance of the benefits provided by their financial institutions were eliminated and half were retained. Those retained were merely used as dummy questions to keep the length of the questionnaire the same. A 5-point Likert scale was used (with 1=Very and 5=Not at all) for these questions. The demographic questions in the second section included the subject's income, occupation, education, age, etc.

The questionnaire Web page used radio buttons and check boxes for capturing the respondents' answers. Active Server Pages code was implemented to insert the data into a database on the Web site host Web server. The code allowed respondents to skip any number of questions. However, the instructions at the beginning of the survey did not explicitly tell the subjects that they could or could not skip questions. The survey was set up with a main questionnaire page that included the entire form on one page, and required the subject to scroll a full screen four times on a 15 inch monitor (3 times on a 17 inch monitor). After the submit button was clicked, a follow-up attitudinal questionnaire page appeared with questions that were appropriate for the treatment that was assigned to that particular respondent.

### ***Experimental Design for Testing Hypothesis 1***

The design for testing hypothesis 1 involved a single experimental group assigned the single treatment of an access control ID. A single control group was assigned the absence of the access control ID treatment. The main and follow-up

questionnaires were identical for both groups. For hypothesis 1a, the response rate served as the response variable, and the absence or presence of the access control ID treatment served as the independent variable. The response rate for this study equals the ratio of the number of survey submissions to the number of successfully delivered e-mail invitations. For hypothesis 1b, the missing data rate served as the response variable. The missing data rate for this study equals the ratio of the number of unanswered questions to the total number of questions in the main questionnaire (excluding the before mentioned dummy questions). A one-way analysis of variance was employed to analyze the collected data.

### ***Sample and Procedure for Testing Hypothesis 1***

One thousand Iowa State University Alumni were randomly assigned to one of the two groups so that each group contained exactly 500 subjects. The subjects in the experimental group were e-mailed an offer to participate in the Web-based survey that included the Web site address for the experimental group home page and a ten-digit randomly generated access code unique to each recipient. Once the experimental group subjects pulled up the home page of the survey, they were directed to enter the access code before proceeding to the questionnaire. Once the questionnaire was completed and submitted by clicking on the submit button, the follow-up screen appeared asking the subjects if they felt they were anonymous when submitting the survey, or whether they thought that they could be identified. Upon clicking the submit button on this screen, a confirmation page appeared

thanking the respondents for their participation. The data and access ID were then transmitted into the database.

The control group subjects were e-mailed an offer to participate in the survey that included the Web site address for the control group version home page. The e-mail did not include an access ID. When the control group subjects downloaded the home page of the survey, there was a button to continue to the survey questionnaire page, but no access ID field was present. The follow-up questionnaire and confirmation screen were the same as in the experimental group version.

### ***Preventing Duplicate Submissions***

For the experimental group, duplicate submissions were prevented when respondents accessed the questionnaire after having previously submitted the survey. When the access ID was entered, a lookup in the database was performed. If a response record with that access ID existed, and the subject went on to complete the survey, the new record was not written to the database. The confirmation screen appeared after the resubmission, so the subject was not aware that the resubmission was not recorded. If respondents clicked the back button of their browser after completing the survey the first time and resubmitted, the same lookup was performed using the access ID of the original submission, and that record was not written to the database either (see Figure 3.1).

To prevent duplicate submissions from the control group, a cookie was placed on the computer of each visitor to the control group survey when they reached the survey home page. Upon submitting the questionnaire, the cookie was edited

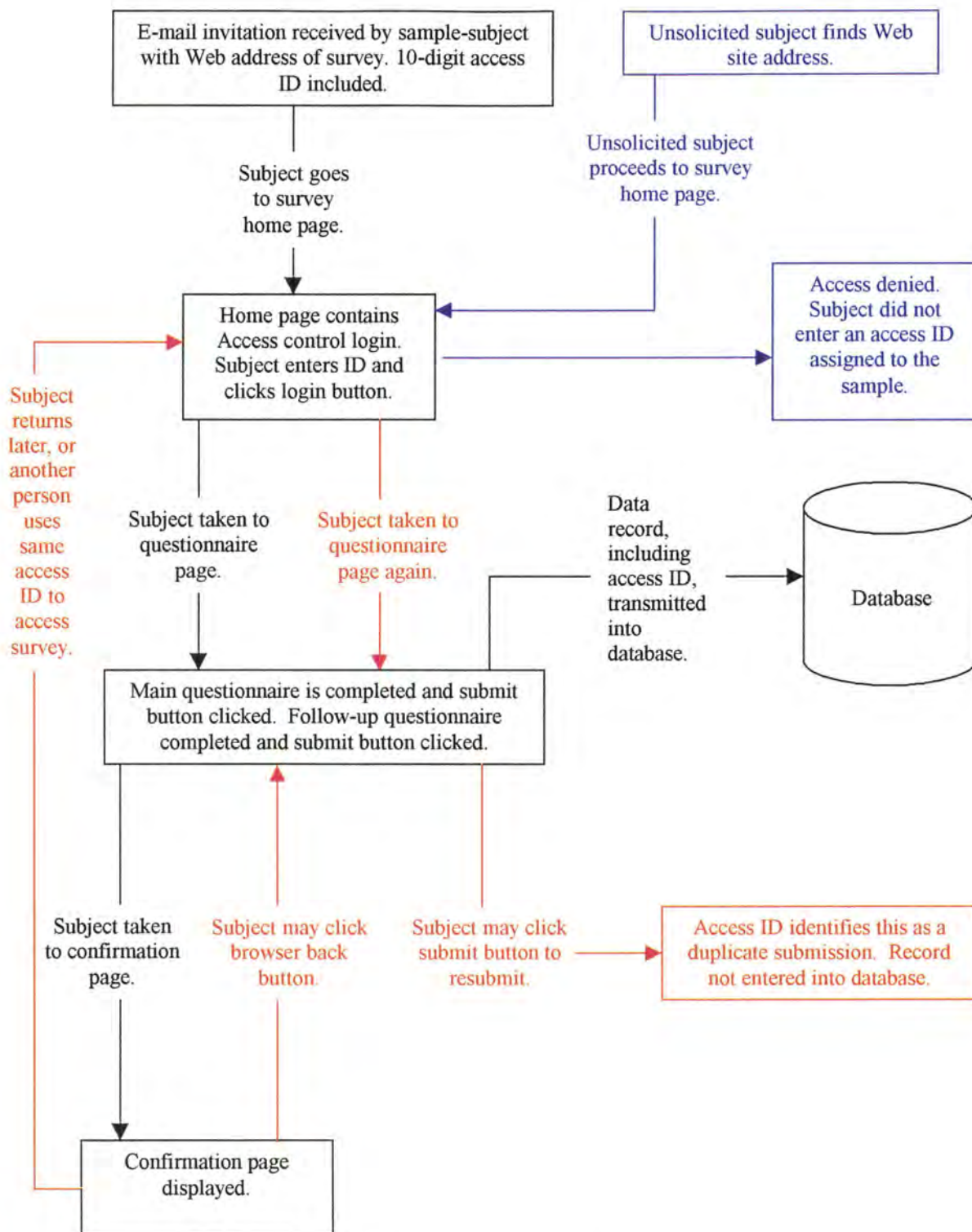


Figure 3.1 Survey Model with Access Control ID

to indicate that the survey had been taken on that computer. If a resubmission was executed using the same computer (either by starting over or using the back button of the browser), the cookie identified it as a resubmission, and the data record was not written to the database (see Figure 3.2). As with the experimental group, the confirmation screen appeared after the resubmission, so the subject was not aware that it was not recorded. The only way for the control group subjects to transmit duplicate submissions was to either have cookies disabled in their Web browser settings, or to delete the cookie after submitting the survey. This would require awareness that a cookie was placed on their computers (which is done behind the scenes) and knowledge of how to find and delete the cookie after each submission. If they set the Web browser so that cookies were not accepted, the submission of the survey was flagged in the database as a submission from a computer that had cookies disabled. The Internet Protocol address of the computer that submitted the survey was also written to the database. The IP address of those submissions with cookies disabled was compared to the IP addresses of other submissions in the database in the same approximate time period to determine if a duplicate submission was recorded. If this were the case, the record was disqualified.

### ***Controlling the Sample***

The access ID requirement prevented unsolicited persons that had downloaded the experimental group home page (whether by using a search engine, or intentionally or unintentionally typing in the correct Web address) from accessing the

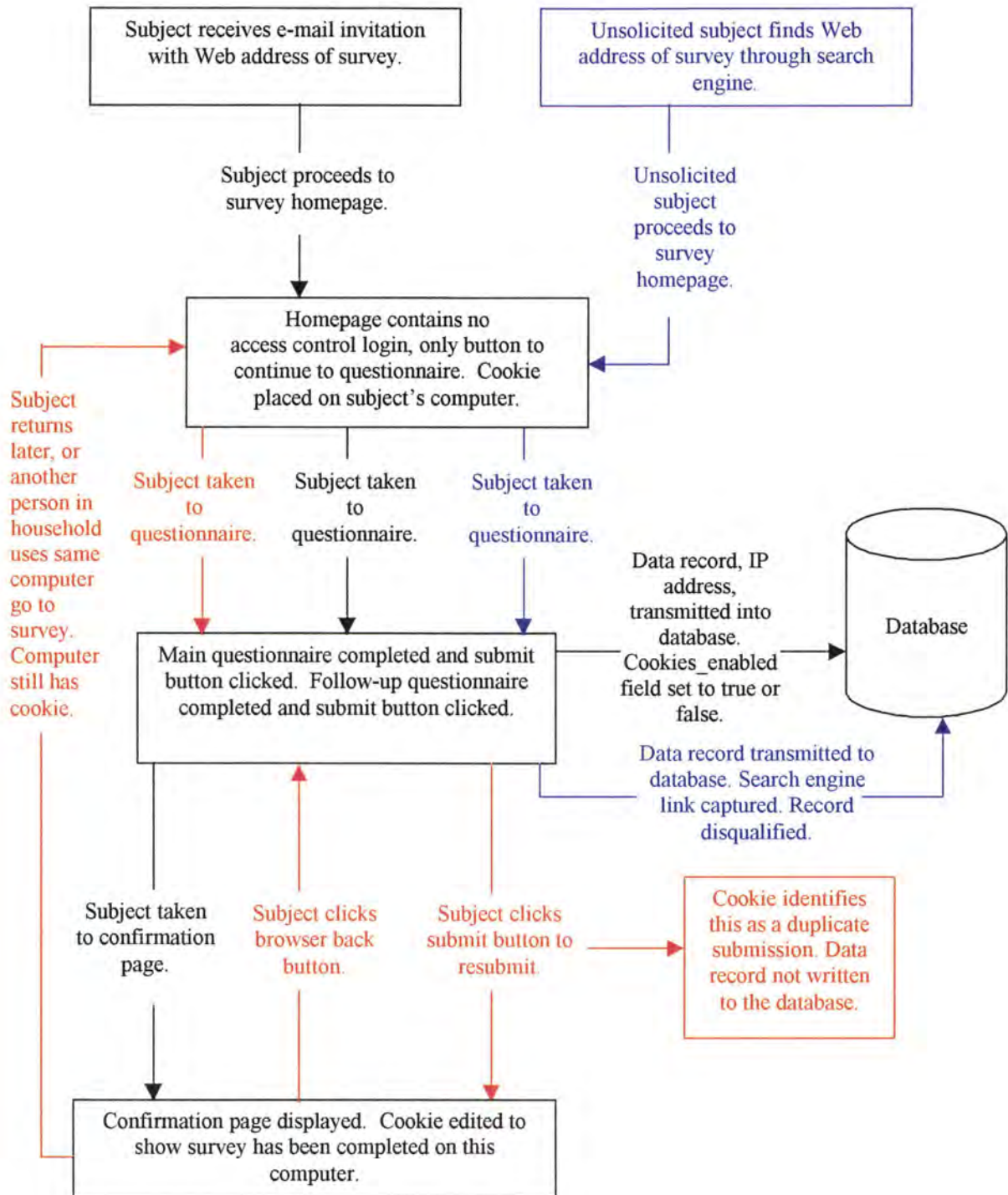


Figure 3.2 Survey Model without Access Control ID

experimental group questionnaire page. If persons clicked the login button without entering an access ID or tried to guess an ID, a lookup was performed and if the ID was not in the database table that contained the access IDs of the sample participants, a screen appeared informing them that they had entered an incorrect access ID and instructing them to backup and try again.

The following procedures prevented submission of the control group survey by unsolicited persons. First, steps were taken to prevent search engines from indexing the Web site. A robots meta tag named NOINDEX was placed in the Active Server Pages code of the control group survey home page that did not allow the pages to be indexed. This procedure prevented unsolicited persons from accessing the control group survey through a link from some search engines. If persons accessed the survey through a search engine that did not support the NOINDEX function, the link that they clicked from the search results was captured and entered into the database. An examination of the link would identify it as an access from a search engine, and the submission would be disqualified.

Only a remote chance existed that anyone not invited to participate in the control group survey would accidentally or intentionally type in the correct Web site address and fill it out. The data was gathered during a relatively short time span (one month), and the home page filenames (the filename is the rightmost entry in the Web address) of the control group Web pages served the same purpose as a password (if a person does not know the filename they can't pull up the home page). The only probable way for persons who have not received the e-mail solicitation to access the control group survey would have been for the sample subjects to forward

the Web address to them or to verbally tell them. This is probably the only case where selection bias may have been introduced. If subjects forwarded the solicitation e-mail for the experimental group version to unsolicited persons, and those people accessed and completed the survey, the submissions were not written to the database if the original recipients had already participated. If the original recipients had not completed the survey, and the persons who received the forwards filled them out using the original recipients' access IDs, these data records were used in the analysis. However, this removed the treatment for the lack-of-anonymity effect since the persons were using access IDs not specifically e-mailed to them (they identified the original recipients).

It should be noted that the forwarding of the solicitation e-mail for a Web-based survey is comparable to a person receiving a paper and pencil survey through the mail, then photocopying the questionnaire and giving it to someone else. However, forwarding e-mail takes literally no effort, where photocopying and distributing copies of a paper and pencil survey would take much more effort. With this observation in mind, a remote possibility of selection bias in the control group version of the Web survey exists.

### ***Experimental Design for Testing Hypotheses 2 and 3***

A 2x2 factorial experiment was used to test hypotheses 2 and 3. For hypotheses 2a and 3, the response variable was the response rate, and the two independent variables were the presence or absence of the confidentiality statement and the presence or absence of the results display. For hypothesis 2b, the missing



data rate was the response variable. Here, the response rate and missing data rate definitions are the same as in the test for hypothesis 1. All four factor combinations included an access control ID, so that variable was held constant. An analysis of variance was employed to determine the presence of significant main and interaction effects. The main questionnaire was the same for all four factors. The questions in the follow-up questionnaire depended on what treatment was present.

### ***Sample and Procedure for Testing Hypotheses 2 and 3***

One thousand five hundred thirty six Iowa State University Alumni were randomly assigned to one of the four groups so that each group contained exactly 384 subjects. The sample sizes were determined by the number of delivered e-mails from each group. All factor combinations included the access ID requirement since it was the most reliable way to control the sample. The first combination involved an e-mail invitation and home page that did not include a confidentiality statement or the promise of a results display, while the second combination did not include a confidentiality statement but did include the promise of the results display. The third combination included a confidentiality statement but no promise of a results display, and the fourth combination included both the confidentiality statement and the promise of a results display. The prevention of duplicate submissions was accomplished with the access control ID just as in the experimental group for testing hypothesis 1.

## CHAPTER 4. RESULTS AND DISCUSSION

### *Hypothesis 1A*

Of the 500 e-mail invitations sent to the experimental group with the access ID, 119 were returned undeliverable making the sample size  $n = 381$  (see Table 4.1). Of this number, 85 responded to the survey making the response rate 22.3% with an error of estimation of  $\pm 4.3\%$  (see Table 4.2). Of the 500 e-mail invitations sent to the control group without the access ID, 116 were returned undeliverable making the sample size  $n = 384$ . Of this number, 93 responded to the survey making the response rate 24.2% with an error of estimation of  $\pm 4.3\%$ . Four of these responses had cookies disabled but no duplicate IP addresses were found, therefore, none were disqualified. No responses captured a search engine link, so none of the responses were from unsolicited individuals that found the Web survey through a search engine.

**Table 4.1 Descriptive Statistics for Access Control Response Rate**

Access Control	Mean	Std. Deviation	N
No access ID	.24	.43	384
Access ID	.22	.42	381
Total	.23	.42	765

Dependent variable: response rate

**Table 4.2 Estimates for Access Control Response Rate**

Access Control	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
No access ID	.242	.022	.200	.285
Access ID	.223	.022	.181	.266

Dependent variable: response rate

Hypothesis 1a is restated as:  $H_0: p_{\text{access ID}} = p_{\text{no access ID}}$

$H_a: p_{\text{access ID}} < p_{\text{no access ID}}$

A one-way ANOVA was run to test the hypothesis. The test statistic has the value  $F = 0.390$ , and the observed significance level is 0.533 (see Table 4.3). The null hypothesis is not rejected, indicating that there is insufficient evidence to show that the response rate for Web-based surveys with an access control ID is less than the response rate for Web-based surveys without an access control ID.

**Table 4.3 ANOVA for Access Control Response Rate**

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	6.97E-02	1	6.97E-02	.390	.533
Intercept	41.403	1	41.403	231.410	.000
Access Control	6.97E-02	1	6.97E-02	.390	.533
Error	136.513	763	.179		
Total	178.000	765			
Corrected Total	136.583	764			

Dependent variable: response rate

The mean difference in response rate ( $p_{\text{no access ID}} - p_{\text{access ID}}$ ) is 1.9%. The 0.95 confidence interval is -4.1% to 7.9%, therefore, we are 95 percent confident that the true difference in response rate between Web-based surveys with and without an access ID requirement is between -4.1% and 7.9% (see Table 4.4).

**Table 4.4 Pairwise Comparison for Access Control Response Rate**

(I) Access Control	(J) Access Control	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
No access ID	Access ID	1.909E-02	.031	.533	-4.095E-02	7.913E-02
No access ID	Access ID	1.909E-02	.031	.533	-4.095E-02	7.913E-02

Dependent variable: response rate

### ***Hypothesis 1B***

The mean percentage of missing data fields for the  $n = 85$  responses from the experimental group with the access control ID was 4.4% (see Table 4.5), with an error of estimation of  $\pm 2.8\%$  at the 0.95 confidence level (see Table 4.6). The mean percentage of missing data fields for the  $n = 93$  responses from the control group without the access control ID was 2.9% with an error of estimation of  $\pm 2.7\%$  at the 0.95 confidence level.

**Table 4.5 Descriptive Statistics for Access Control Missing Data Rate**

Access Control	Mean	Std. Deviation	N
No access ID	2.9391E-02	9.7039E-02	93
Access ID	4.3922E-02	.15922	85
Total	3.6330E-02	.1303	178

Dependent variable: missing data rate

**Table 4.6 Estimates for Access Control Missing Data Rate**

Access Control	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
No access ID	2.939E-02	.014	2.692E-03	5.609E-02
Access ID	4.392E-02	.014	1.599E-02	7.185E-02

Dependent variable: missing data rate

Hypothesis 1b is restated as:  $H_0: p_{\text{access ID}} = p_{\text{no access ID}}$

$H_a: p_{\text{access ID}} > p_{\text{no access ID}}$

A one-way ANOVA was used to test the hypothesis. The test statistic has the value  $F = 0.551$ , and the observed significance level is 0.459 (see Table 4.7). The null hypothesis is not rejected, indicating that there is insufficient evidence to show that the percentage of missing data fields for Web-based surveys with an access

**Table 4.7 ANOVA for Access Control Missing Data Rate**

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	9.38E-03	1	9.38E-03	10.551	.459
Intercept	.239	1	.239	14.023	.000
Access Control	9.38E-03	1	9.38E-03	.551	.459
Error	2.996	176	1.70E-02		
Total	3.24	178			
Corrected Total	3.005	177			

Dependent variable: missing data rate

control ID is greater than the percentage of missing data fields for Web-based surveys without an access control ID. The mean difference in the percentage of missing data fields is 1.5%, and the 0.95 confidence interval on the difference ( $p_{\text{access ID}} - p_{\text{no access ID}}$ ) is -2.4% to 5.3% (see Table 4.8). Therefore, we are 95 percent confident that the true difference in the percentage of missing data fields in Web-based surveys with and without an access control ID is between -2.4% and 5.3%.

**Table 4.8 Pairwise Comparison for Access Control Missing Data Rate**

(I)Access Control	(J)Access Control	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Access ID	No access ID	1.453E-02	.020	.459	-2.411E-02	5.317E-02

Dependent variable: missing data rate

### ***Identification Follow-up Question***

83 of the 85 respondents from the access ID group answered the follow-up question. 69.9% felt they could be identified and 30.1% thought they were anonymous. 92 of the 93 respondents from the no access ID group answered the follow-up question. 56.5% felt they could be identified and 43.5% felt they were

anonymous. The sample sizes are too small to determine if there is a significant difference in the number of respondents that felt they could be identified. However, combined with some previously unused data, the sample sizes are sufficient. The data collected for testing hypothesis 1 came from a sample of respondents from a second mailing of invitations. Data from the first mailing was not used because some subjects from the control group without the access ID were receiving an error message when they clicked the button to submit the survey. Unknown at the time to this researcher, the data field that captured the link from search engines was also capturing the link in email invitations that were sent to Web-based email addresses (hotmail, yahoo mail, etc.). Some of these Web-based email programs were adding hundreds of additional characters to the link. Unfortunately, the database field that stored these links was only set to 100 characters. Consequently, some responses were not recorded for the control group (this was discovered when some of the subjects replied to the email saying they received the error message). The server log files were unavailable, so it was impossible to determine the number of failed response attempts. Because response rate was affected, this data was not used, and a second set of invitations had to be sent out after setting the data field to the memo data type (64,000 characters).

The assumption is made that in this prior mailing the mean percentage that felt they could be identified is the same for those whose response attempt failed (due to the error message) as for those that succeeded. Only the sample size would have been affected. Based on this assumption, the data from these responses was combined with the data from the second sample producing the following results:

136 respondents from the access ID group answered the follow-up question (see Table 4.9). 70% felt they could be identified, with a  $\pm 8.1\%$  error of estimation (see Table 4.10). 156 respondents from the no access ID group answered the question, and 58% felt they could be identified, with a  $\pm 7.6\%$  error of estimation.

**Table 4.9 Descriptive Statistics for Access Control Identification Follow-up Question**

Access Control	Mean	Std. Deviation	N
No access ID	.57692	.49564	156
Access ID	.69853	.46059	136
Total	.63356	.48266	292

Dependent Variable: Identification Question

**Table 4.10 Estimates for Access Control Identification Follow-up Question**

Access Control	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
No access ID	.577	.038	.501	.653
Access ID	.699	.041	.618	.779

Dependent Variable: Identification Question

A one-way ANOVA was run to test for a significant difference. The test statistic is  $F = 4.670$ , and the observed significance level is 0.032 (see Table 4.11). There is sufficient evidence to show that the mean percentage of people who feel they can be identified in a Web-based survey with an access ID is greater than in a Web-based survey without an access ID. The mean difference ( $p_{\text{access ID}} - p_{\text{no access ID}}$ ) is 12.2% and the 0.95 confidence interval for the difference is 1.1% to 23.2% (see Table 4.12). Therefore, we are 95% confident that the true mean difference in the percentage of respondents that feel they could be identified with or without access control is between 1.1% and 23.2%.

**Table 4.11 ANOVA for Access Control Identification Follow-up Question**

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	1.074	1	1.074	4.670	.032
Intercept	118.198	1	118.198	513.775	.000
ACCESSID	1.074	1	1.074	4.670	.032
Error	66.717	290	.230		
Total	185.000	292			
Corrected Total	67.791	291			

Dependent Variable: Identification Question

**Table 4.12 Pairwise Comparison for Access Control Identification Follow-up Question**

(I) Access Control	(J) Access Control	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Access ID	No access ID	.122	.056	.032	1.086E-02	.232

Dependent Variable: Identification Question

### ***Hypothesis 2A***

Of the 768 e-mail invitations sent to the experimental group with the confidentiality statement, 182 were returned undeliverable, making the sample size  $n = 586$  (see Table 4.13). Of this number, 103 responded to the survey making the response rate 17.6% with an error of estimation of  $\pm 3.1\%$  at the .095 confidence level (see Table 4.14). Of the 768 e-mail invitations sent to the control group without the confidentiality statement, 173 were returned undeliverable, making the sample size  $n = 595$ . Of this number, 109 responded to the survey making the response rate 18.4% with an error of estimation of  $\pm 3.1\%$  at the 0.95 confidence level.



**Table 4.13 Descriptive Statistics for Confidentiality Statement Response Rate**

Confidentiality Statement	Mean	Std. Deviation	N
No confidentiality statement	.18	.39	595
Confidentiality statement	.18	.38	586
Total	.18	.38	1181

Dependent variable: response rate

**Table 4.14 Estimates for Confidentiality Statement Response Rate**

Confidentiality Statement	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
No confidentiality statement	.184	.016	.153	.214
Confidentiality Statement	.176	.016	.145	.207

Dependent variable: response rate

Hypothesis 2a is restated as:  $H_0: \rho_{\text{confidentiality statement}} = \rho_{\text{no confidentiality statement}}$

$H_a: \rho_{\text{confidentiality statement}} > \rho_{\text{no confidentiality statement}}$

The test statistic for the ANOVA analysis has the value  $F = 0.104$ , and the observed significance level is 0.747 (see Table 4.15). The analysis showed no significant interaction effect between the confidentiality statement and results display treatments ( $F=.292$ ,  $\text{Sig}=.589$ ). The null hypothesis is not rejected, indicating that there is insufficient evidence to show that the response rate for Web-based surveys

**Table 4.15 ANOVA for Confidentiality Statement Response Rate**

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	.396	3	.132	.896	.443
Intercept	38.202	1	38.202	259.084	.000
Confidentiality Statement	1.54 -02	1	1.5E-0	.104	.747
Results Display	.339	1	.339	2.297	.130
Confidentiality Statement * Results Display	4.30E-02	1	4.3E-0	.292	.589
Error	173.548	1177	.147		
Total	212.000	1181			
Corrected Total	173.944	1180			

Dependent variable: response rate

with a posted confidentiality statement is greater than the response rate for Web-based surveys without a posted confidentiality statement. The 0.95 confidence interval for ( $p_{\text{confidentiality statement}} - p_{\text{no confidentiality statement}}$ ) is -5.1% to 3.7%, therefore we are 95 percent confident that the true difference in response rate is between -5.1% and 3.7% (see Table 4.16).

**Table 4.16 Pairwise Comparison for Confidentiality Statement Response Rate**

(I) Confidentiality Statement	(J) Confidentiality Statement	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					L-Bound	U-Bound
Confidentiality statement	No confidentiality statement	-7.22E-03	.022	.747	-5.109E-02	3.664E-02

Dependent variable: response rate

Of the 103 responses in the confidentiality statement group, 54.4% indicated in the follow-up questionnaire that the confidentiality statement was very important in their willingness to participate in the survey, while 24.2% said that it was somewhat important. 9.7% were neutral, while 3.9% thought it was not very important. Only 2% felt that it was not at all important and 5.8% did not answer the question.

### ***Hypothesis 2B***

The mean percentage of missing data fields for the 103 responses with the confidentiality statement was 3.7% (see Table 4.17) with a  $\pm 2.2\%$  error of estimation (see Table 4.18). The mean percentage of missing data fields for the 109 responses without the confidentiality statement was 3.1% with a  $\pm 2.2\%$  error of estimation.

**Table 4.17 Descriptive Statistics for Confidentiality Statement Missing Data Rate**

Confidentiality Statement	Mean	Std. Deviation	N
No Confidentiality statement	3.0581E-02	8.8672E-02	109
Confidentiality statement	3.7540E-02	.13191	103
Total	3.3962E-02	.11157	212

Dependent variable: missing data rate

**Table 4.18 Estimates for Confidentiality Statement Missing Data Rate**

Confidentiality Statement	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
No confidentiality statement	3.054E-02	.011	9.355E-03	5.172E-02
Confidentiality statement	3.658E-02	.011	1.467E-02	5.848E-02

Dependent variable: missing data rate

Hypothesis 2b is restated as:  $H_0: p_{\text{confidentiality statement}} = p_{\text{no confidentiality statement}}$

$H_a: p_{\text{confidentiality statement}} < p_{\text{no confidentiality statement}}$

The test statistic has the value  $F = 0.153$ , and the observed significance level is 0.696 (see Table 4.19). No significant interaction existed between the confidentiality statement treatment and results display ( $F = .231$ ,  $\text{Sig.} = .631$ ). The null hypothesis is not rejected, indicating that there is insufficient evidence to show that the percentage of missing data fields for Web-based surveys with a posted confidentiality statement is less than for Web-based surveys without a posted confidentiality statement. The estimated mean difference is -0.6% and the 0.95 confidence interval for  $(p_{\text{no confidentiality statement}} - p_{\text{confidentiality statement}})$  is -3.7% to 2.4% (see Table 4.20). Therefore, we are 95 percent confident that the true difference in the proportion of missing data fields is between -3.7% and 2.4%. Of the 103 responses in the confidentiality statement group, 46.6% said that the confidentiality statement was very important in their willingness to answer all questions in the

**Table 4.19 ANOVA for Confidentiality Statement Missing Data Rate**

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	1.117E-02	3	3.722E-03	.296	.828
Intercept	.237	1	.237	18.850	.000
Confidentiality Statement	1.919E-03	1	1.919E-03	.153	.696
Results Display	5.963E-03	1	5.963E-03	.474	.492
Confidentiality Statement * Results Display	2.907E-03	1	2.907E-03	.231	.631
Error	2.615	208	1.257E-02		
Total	2.871	212			
Corrected Total	2.627	211			

Dependent variable: missing data rate

**Table 4.20 Pairwise Comparison for Confidentiality Statement Missing Data Rate**

(I) Confidentiality Statement	(J) Confidentiality Statement	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					L-Bound	U-Bound
No confidentiality statement	Confidentiality statement	-6.038E-03	.015	.696	-3.651E-02	2.444E-02

Dependent variable: missing data rate

survey, while 26.2% said it was somewhat important. 14.6% were neutral and 4.9% thought it was not very important. 1.9% did not think the statement was important at all in their willingness to answer all the questions and 5.8% skipped the question.

### ***Hypothesis 3***

Of the 768 e-mail invitations sent to the results display group, 194 were returned undeliverable, making the sample size  $n = 574$  (see Table 4.21). Of this number, 113 responded to the survey making the response rate 19.7% with a  $\pm 3.2\%$  error of estimation (see Table 4.22). Of the 768 e-mail invitations sent to the control group

**Table 4.21 Descriptive Statistics for Results Display Response Rate**

Results Display	Mean	Std. Deviation	N
No Results display	.16	.37	607
Results display	.20	.40	574
Total	.18	.38	1181

Dependent variable: response rate

**Table 4.22 Estimates for Results Display Response Rate**

Results Display	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
No results display	.163	.016	.132	.194
Results display	.197	.016	.165	.228

Dependent variable: response rate

without the results display, 161 were returned undeliverable, making the sample size  $n = 607$ . Of this number, 99 responded to the survey making the response rate 16.3% with an error of estimation of  $\pm 3.1\%$ .

Hypothesis 3 is restated as:  $H_0: p_{\text{results display}} = p_{\text{no results display}}$

$H_a: p_{\text{results display}} > p_{\text{no results display}}$

The test statistic for the ANOVA analysis has the value  $F = 2.297$ , and the observed significance level is 0.130 (see Table 4.23). As stated earlier, the ANOVA found no significant interaction effect between the results display and confidentiality statement treatments. The null hypothesis is not rejected, indicating that there is insufficient evidence to show that the response rate for Web-based surveys with a promised results display is greater than the response rate for Web-based surveys without a promised results display. The 0.95 confidence interval on the difference ( $p_{\text{results display}} - p_{\text{no results display}}$ ) is -1% to 7.8% (see Table 4.24), therefore, we are 95 percent confident that the true difference in response rate is between -1% and 7.8%.

**Table 4.23 ANOVA for Results Display Response Rate**

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	.396	3	.132	.896	.443
Intercept	38.202	1	38.202	259.084	.000
Confidentiality Statement	1.54 -02	1	1.5E-02	.104	.747
Results Display	.339	1	.339	2.297	.130
Confidentiality Statement * Results Display	4.30E-02	1	4.3E-02	.292	.589
Error	173.548	1177	.147		
Total	212.000	1181			
Corrected Total	173.944	1180			

Dependent variable: response rate

**Table 4.24 Pairwise Comparison for Results Display Response Rate**

(I) Results Display	(J) Results Display	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Results display	No results display	3.389E-02	.022	.130	-9.977E-03	7.775E-02

Dependent variable: response rate

Of the 113 respondents with the promise of a results display, only 1% said it had an important motivating effect in getting them to participate in the survey, while 31% indicated it had somewhat of an effect. 21.2% were neutral on the question, while 25.7% said it did not have very much of a motivating effect. 16.7% said it had no effect and 4.4% did not answer.

### ***Internal and External Validity***

The internal validity of the study was established through random assignment of the subjects to the treatments. It was also set up by controlling the sample through the access control IDs and the previously detailed procedures for the non-

access ID groups. However, it should be noted that the samples consisted entirely of highly educated subjects, many holding advanced college degrees. Regarding the general population, further research is needed to establish any possible external validity of the results. While researchers can easily obtain a representative sample of e-mail addresses from limited populations such as special interest groups, obtaining an e-mail probability sample from the general population would be quite difficult if not impossible. Many people have multiple e-mail addresses, and not all people in the general population have computers. Many of those who do, have unpublished e-mail addresses.

## CHAPTER 5. CONCLUSIONS

### *Access Control*

The results of the experiment show that the marketing researcher can control the sample in Web-based surveys using access control IDs without significantly reducing the response rate or quality as defined in this study. This can help improve the representativeness of the sample by greatly reducing, through advanced sampling, self-selection bias. Therefore, Web-based surveys can be as useful for marketing research as traditional methods for obtaining marketing information from special interest consumer groups. Combined with the low cost and speed with which marketing data can be collected on Web surveys, this mode of data collection may become the preferred method for gathering marketing information in the future.

The follow-up survey showed that the use of access IDs raises respondents' awareness that they can be identified with the information that they submit. However, this raised awareness does not appear to affect response rate and quality. Therefore, there appears to be no real lack-of-anonymity effect. Unfortunately, the survey only gathered answers to the follow-up question from those who participated, whereas the feelings about anonymity with or without access control IDs of those who did not respond would have also been helpful in the analysis. Further research is needed to discover the beliefs of those who would not participate in Web-based surveys. This could be accomplished by implementing an attitudinal survey through a more traditional mode such as paper and pencil.



### ***Confidentiality Statements***

The analysis suggests that confidentiality statements have no real impact on the response rate and quality of Web-based surveys. This apparently was not because people believed the statements were empty words, for the vast majority in both cases indicated they were very important in getting them to participate in the survey and to answer all the questions. The literature review indicated that confidentiality statements actually raised concerns for some people, which may have balanced with those whose concerns were alleviated. Although including a confidentiality statement in the solicitation e-mail may not be necessary, it is still considered good practice to include a link to a privacy and confidentiality statement on a Web site that gathers personal information.

### ***Results Display***

Finally, the analysis suggests that the researcher cannot improve response rate by promising and displaying the current cumulative results of a survey after the respondents submit the questionnaire. The results of the follow-up attitudinal question support this conclusion. This could be viewed as good news for the researcher because displaying the results on the Web can expose the data from the research to competitors if they discover the Web site. Without any benefit to displaying results, the Web-based marketing researcher need not take this risk. Possible future research could include allowing participants to view current results prior to participation to determine the effect on response rate.

## APPENDIX A. PILOT TEST RESULTS

Proposed Survey Questions	Makes Sense? <sup>a</sup>	Average Score <sup>b</sup>
1-A. How important are convenient banking hours in your choice of banks?	100%	6.75
1-B. How well does your current bank provide convenient banking hours?	100%	7.00
2-A. How important is convenient location in your choice of banks?	100%	7.00
2-B. How well does your current bank provide a convenient location?	100%	7.00
3-A. How important is reliable 24 hour online access in your choice of banks?	100%	7.00
3-B. How well does your current bank provide reliable 24 hour online access?	100%	6.50
4-A. How important are low service charges in your choice of banks?	100%	6.75
4-B. How well does your current bank provide low service charges?	100%	7.00
5-A. How important is courteous customer service in your choice of banks?	100%	6.25
5-B. How well does your current bank provide courteous customer service?	100%	6.75
6. Do you bank online?	100%	6.25
7. If yes, what software do you use?	100%	6.50
8. Do you invest online?	100%	6.00
9. If yes, what online broker(s) do you use?	100%	6.00
10-A. How important is fast and fair claim service in your choice of an auto/homeowners insurance company/agency?	100%	6.75
10-B. How well does your current auto/homeowners insurance company/agency provide fast and fair claim service?	100%	6.25
11-A. How important is it for your auto/homeowners insurance agent to always be in the office during office hours?	100%	6.25
11-B. How well does your current auto/homeowners insurance agent provide the benefit of always being in the office during office hours?	100%	6.50
12-A. How important are low rates in your choice of an auto/homeowners insurance company/agency?	100%	6.50
12-B. How well does your current auto/homeowners insurance company/agency provide low rates?	100%	6.75
13-A. How important is a knowledgeable insurance agent in your choice of an auto/homeowners insurance company/agency?	100%	6.25
13-B. How knowledgeable is your current auto/homeowners insurance agent?	100%	6.50
14-A. How important is courteous customer service in your choice of an auto/homeowners insurance company/agency?	100%	5.75
14-B. How well does your current auto/homeowners insurance company/agency provide courteous customer service?	100%	6.00
15. Have you ever obtained an auto/homeowners insurance quote online?	100%	6.25
16. What company do you insure your auto(s) with?	100%	5.75
17. What company do you carry your homeowners or renters insurance with?	100%	5.75
18. What is your field of occupation?	100%	6.00
19. What is your household income?	100%	3.25
20. What is your age?	100%	4.25
21. What is your gender?	100%	5.75
22. How many persons live in your household?	100%	6.00
Is the length of the questionnaire too long, too short or just right?	100%	Just right

a. Percent that says the question makes sense. b. 1 = Extremely sensitive 7 = Not at all sensitive

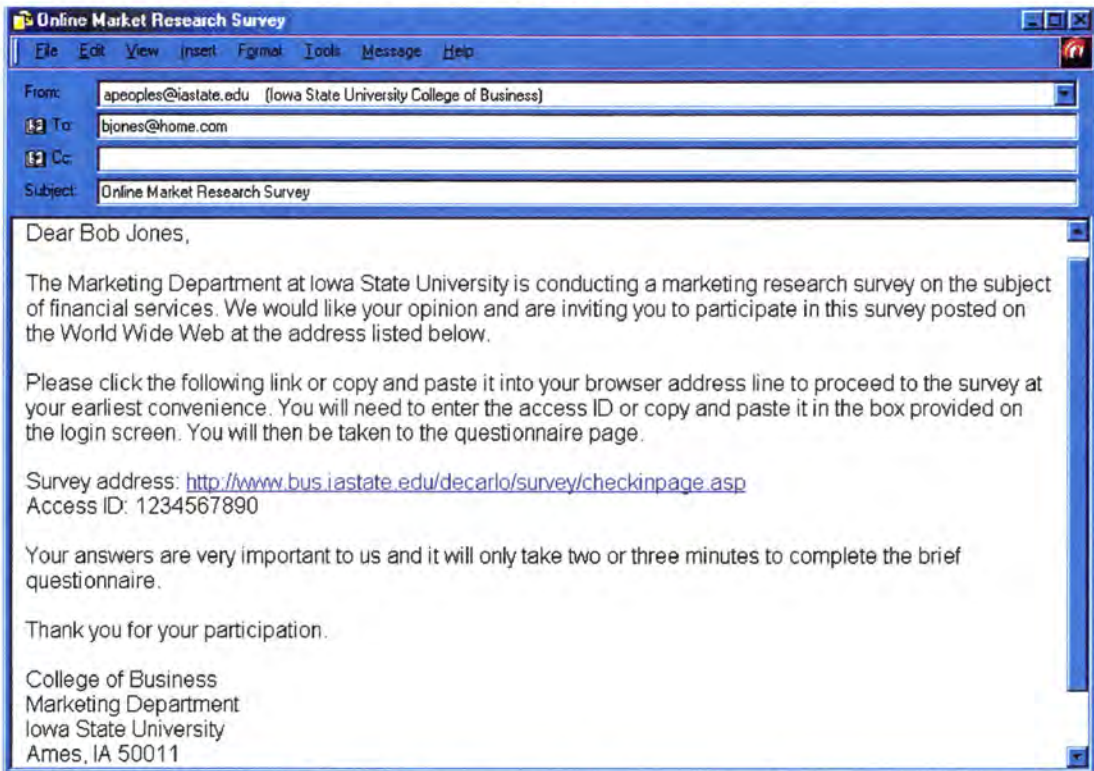
**Table A1. First Pilot Test**

<b>Proposed Survey Questions</b>	<b>Makes Sense?<sup>a</sup></b>	<b>Average Score<sup>b</sup></b>
1-A. How important are convenient banking hours in your choice of banks?	100%	6.75
1-B. How well does your current bank provide convenient banking hours?	100%	6.50
2-A. How important is convenient location in your choice of banks?	100%	6.75
2-B. How well does your current bank provide a convenient location?	100%	6.50
3-A. How important is reliable 24 hour online access in your choice of banks?	100%	6.75
3-B. How well does your current bank provide reliable 24 hour online access?	100%	6.50
4-A. How important are low service charges in your choice of banks?	100%	6.75
4-B. How well does your current bank provide low service charges?	100%	6.50
5-A. How important is courteous customer service in your choice of banks?	100%	6.75
5-B. How well does your current bank provide courteous customer service?	100%	6.25
6. Do you bank online?	100%	6.25
7. If yes, what software do you use?	75%	6.25
8. What is your average checking account balance?	100%	1.25
9. How much do you have in savings (CDs, Money Markets, savings accounts, etc.)?	100%	1.25
10. What credit cards do you carry?	100%	4.00
11. Do you invest online?	100%	5.25
12. If yes, what online broker(s) do you use?	100%	5.75
13. How much do you have invested (stocks, bonds, options, etc.)?	100%	1.00
14. Have you ever obtained an auto/homeowners insurance quote online?	100%	5.75
15. What company do you insure your auto(s) with?	100%	5.75
16. What company do you carry your homeowners or renters insurance with?	100%	5.50
17. What is your field of occupation?	100%	5.50
18. What is your household income?	100%	1.00
19. What is your age?	100%	4.75
20. What is your gender?	100%	6.25
21. What is your marital status?	100%	5.50
22. What is your education level?	100%	5.75
23. How many persons live in your household?	100%	5.50
Is the length of the questionnaire too long, too short or just right?	100%	Just right

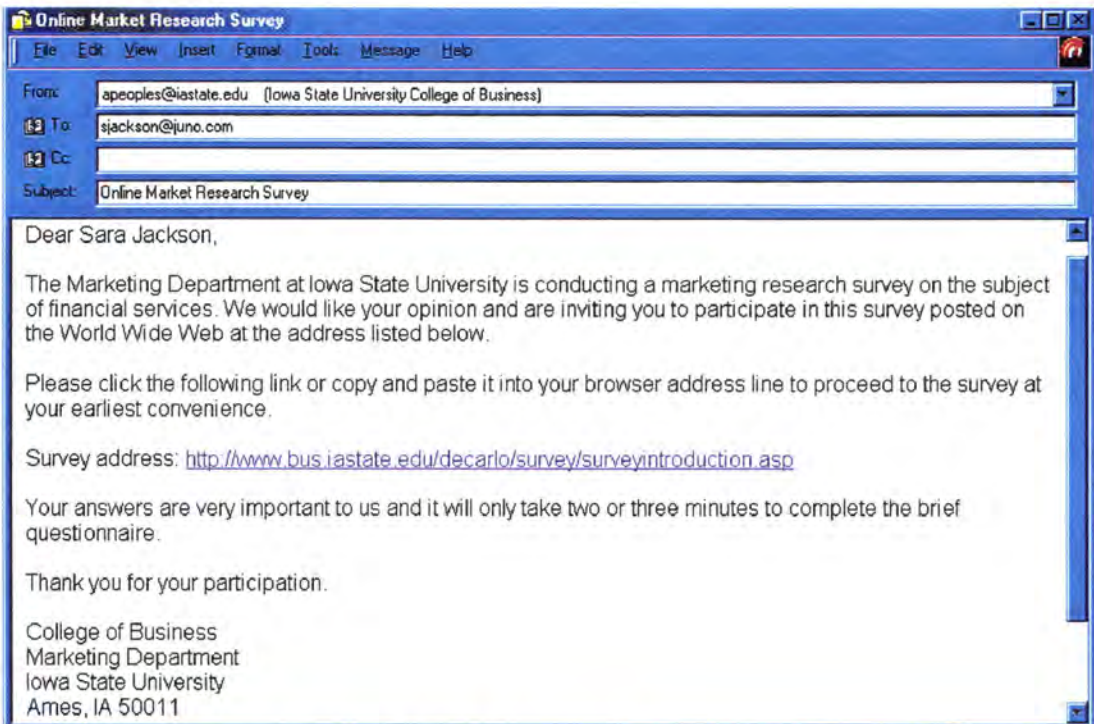
a. Percent that says the question makes sense. b. 1 = Extremely sensitive 7 = Not at all sensitive

**Table A2. Second Pilot Test**

## APPENDIX B. E-MAIL INVITATIONS



**Figure B1. Access ID only E-mail Invitation**



**Figure B2. No Access ID E-mail Invitation**

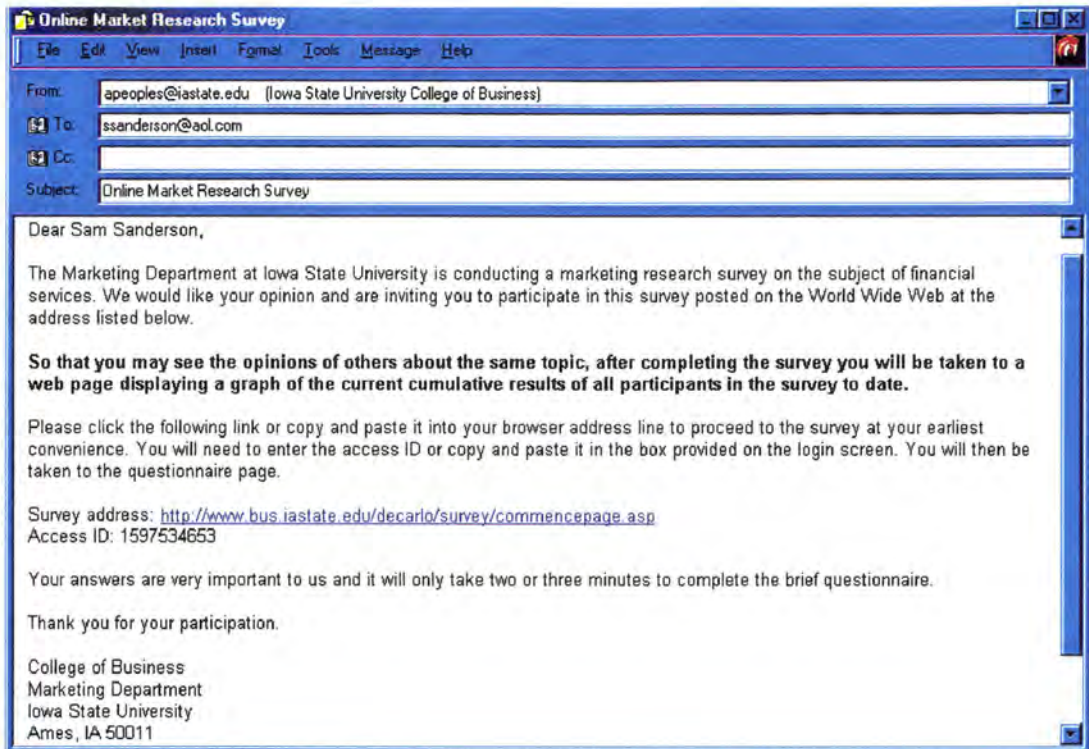


Figure B3. Results Display E-mail Invitation

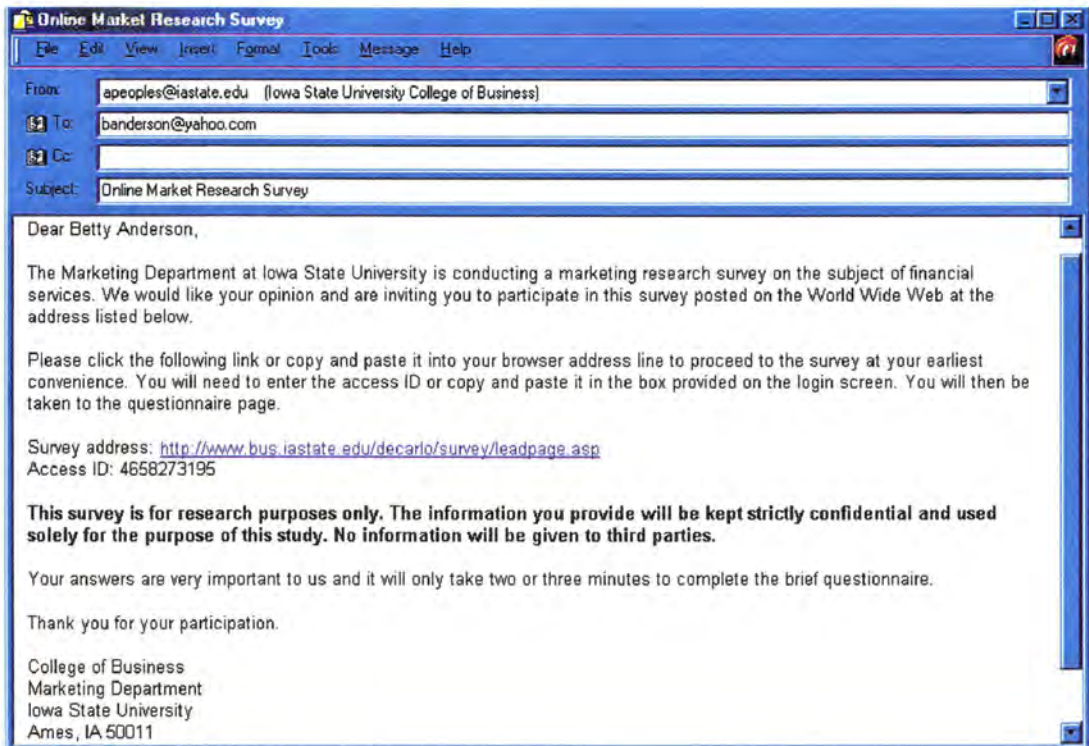


Figure B4. Confidentiality Statement E-mail Invitation

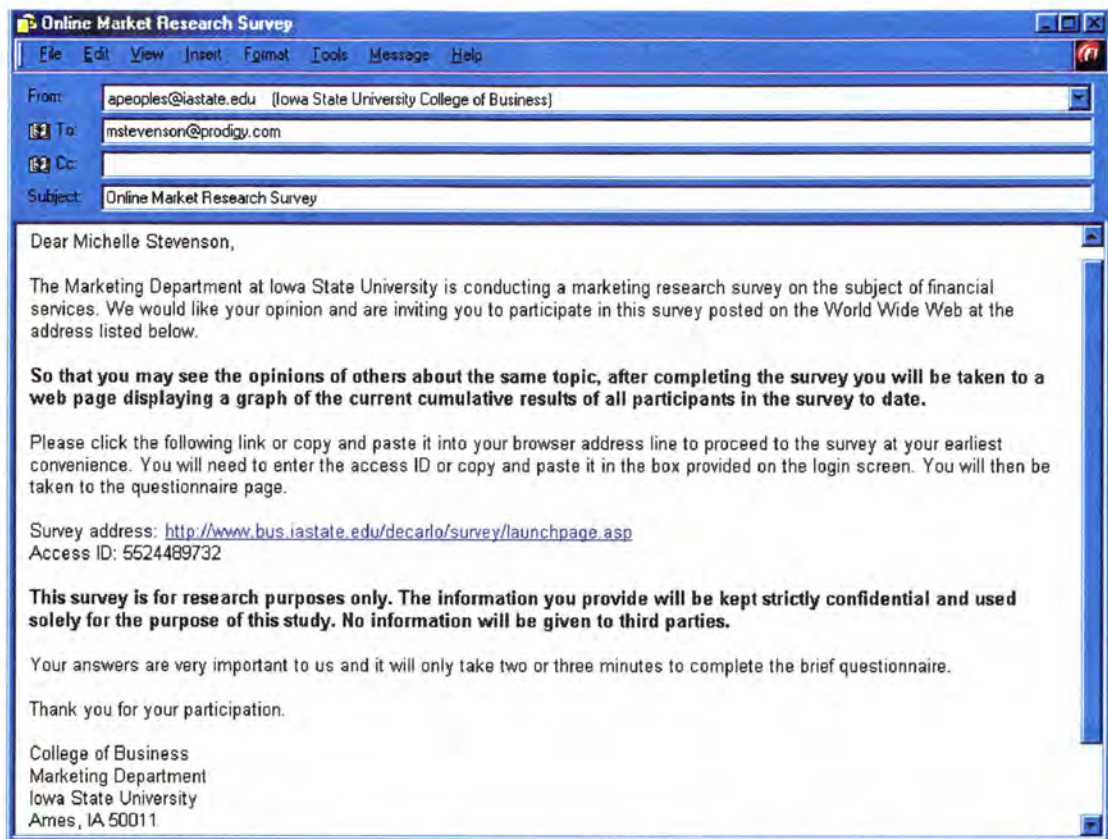


Figure B5. Confidentiality Statement and Results Display E-mail Invitation

## APPENDIX C. SURVEY HOME PAGES

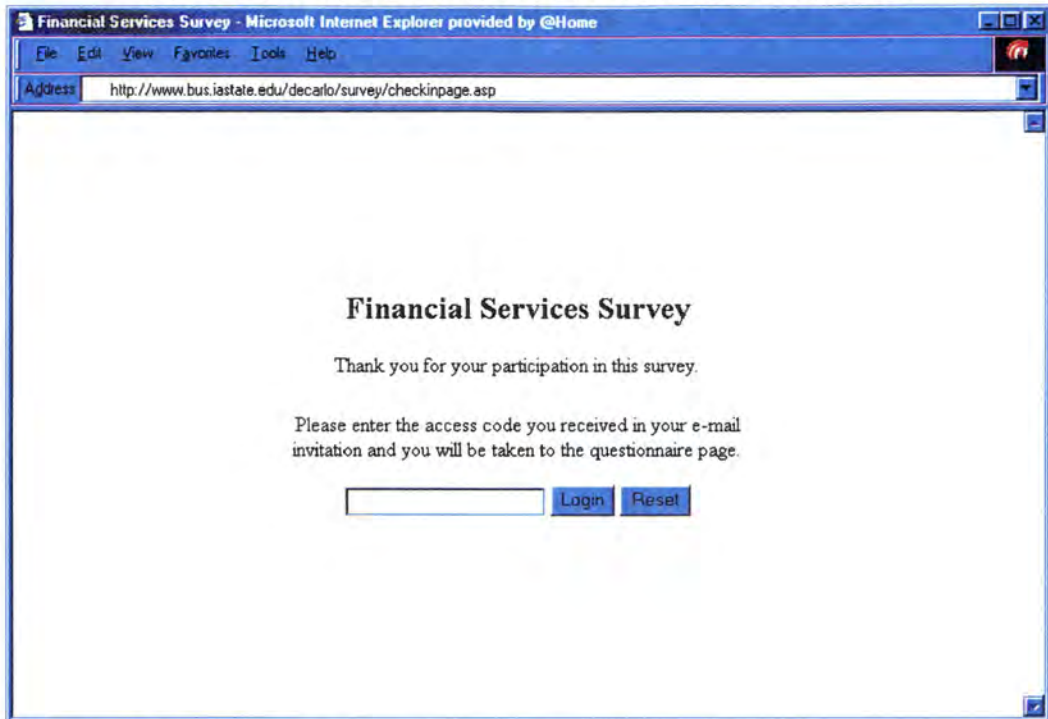


Figure C1. Access ID only Survey Home Page

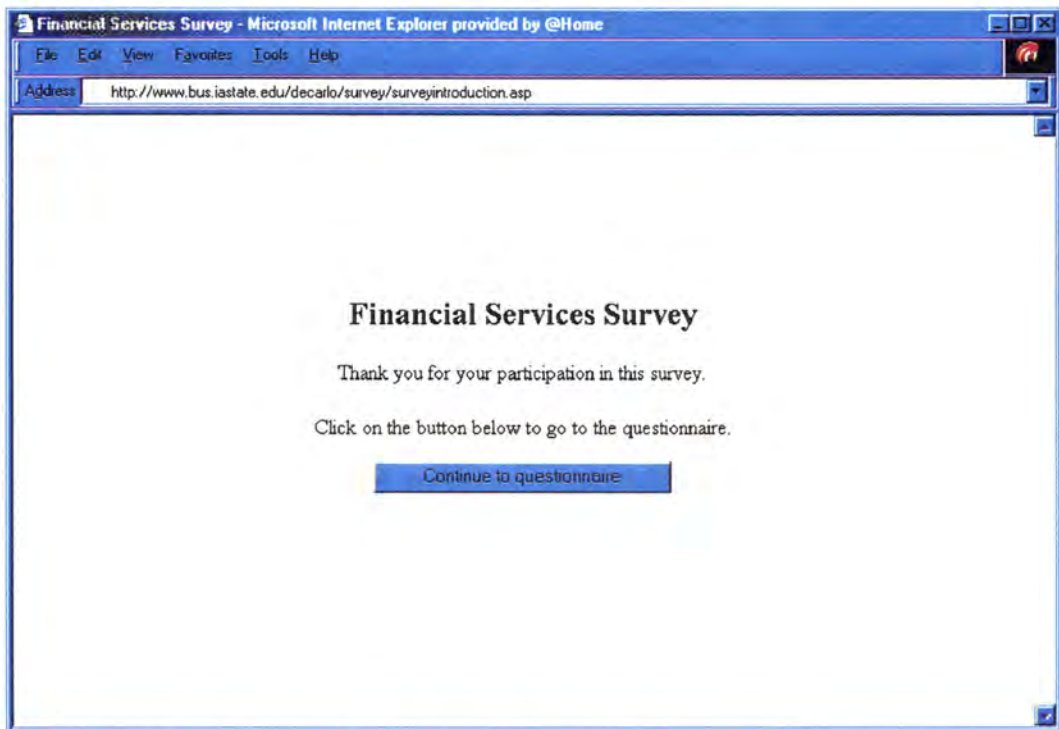
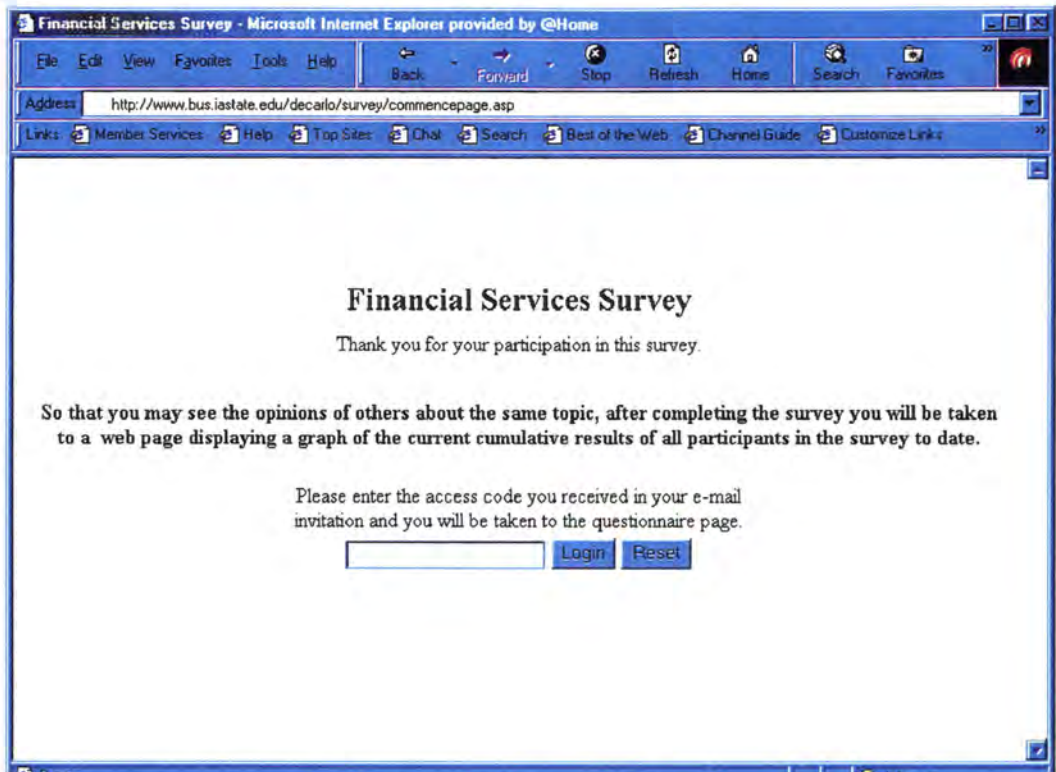
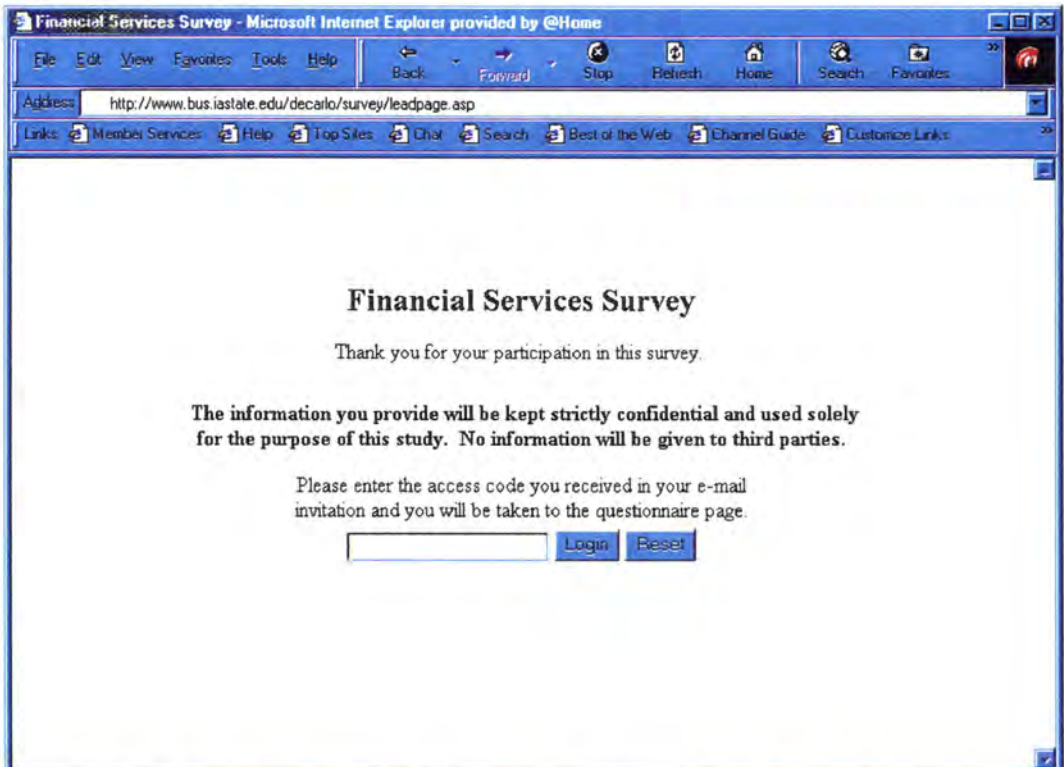


Figure C2. No Access ID Survey Home Page



**Figure C3. Results Display Survey Home Page**



**Figure C4. Confidentiality Statement Survey Home Page**



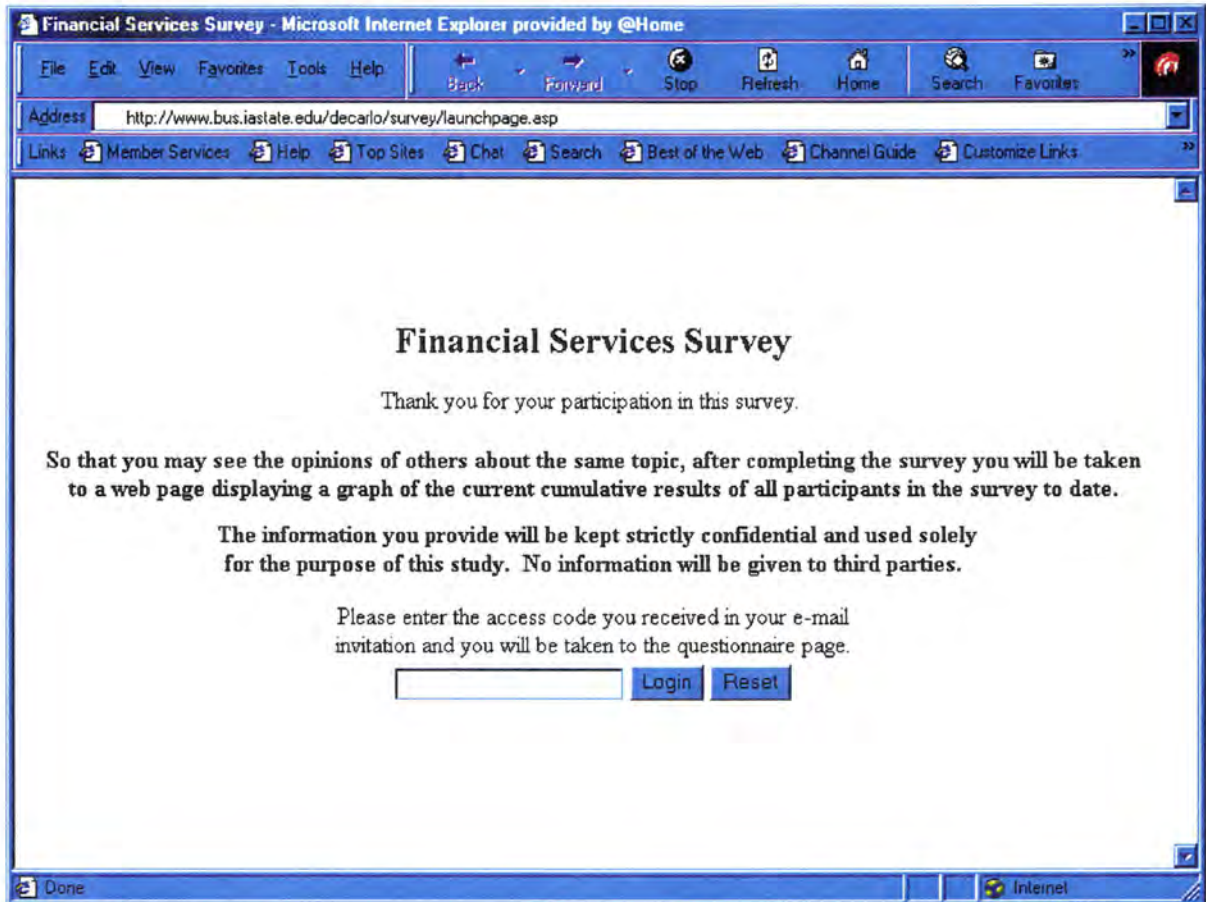


Figure C5. Confidentiality Statement and Results Display Survey Home Page

## APPENDIX D. SURVEY INSTRUMENT

Financial Services Questionnaire - Microsoft Internet Explorer provided by @Home

File Edit View Favorites Tools Help

Address <http://www.bus.iastate.edu/decarlo/survey/survey.asp?sv=collection>

### Financial Services Questionnaire

For the following questions please use the following scale to evaluate the importance each benefit might have on your decision to do business with a particular financial institution. After each importance question will be a question of how well your current financial institution provides that benefit. Use the second scale to answer those questions.

---

**Use the following scale to rate the importance of each benefit in your choice of a financial institution:**

<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
Very	Somewhat	Neutral	Not Very	Not at all

---

**Use the following scale to rate how well your current financial institution provides that benefit.**

<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
Very	Somewhat	Neutral	Not Very	Not at all

---

1-A. How important are convenient banking hours in your choice of banks?  
 1       2       3       4       5

B. How well does your current bank provide convenient banking hours?  
 1       2       3       4       5

2-A. How important is reliable 24 hour online access in your choice of banks?  
 1       2       3       4       5

B. How well does your current bank provide reliable 24 hour online access?  
 1       2       3       4       5

3-A. How important is courteous customer service in your choice of banks?  
 1       2       3       4       5

B. How well does your current bank provide courteous customer service?  
 1       2       3       4       5

4-A. How important is a knowledgeable insurance agent in your choice of an auto/homeowners insurance company/agency?  
 1       2       3       4       5

B. How knowledgeable is your current auto/homeowners insurance agent?  
 1       2       3       4       5

5-A. How important is it for your auto/homeowners insurance agent to always be in the office during office hours?  
 1       2       3       4       5

B. How well does your current auto/homeowners insurance agent provide the benefit of always being in the office during office hours?  
 1       2       3       4       5

Figure D. Main Questionnaire

Financial Services Questionnaire - Microsoft Internet Explorer provided by @Home

File Edit View Favorites Tools Help Address Links

6-A. How important is courteous customer service in your choice of an auto/homeowners insurance company/agency?

1       2       3       4       5

B. How well does your current auto/homeowners insurance company/agency provide courteous customer service?

1       2       3       4       5

7. Do you bank online?      8. If yes, what software do you use?

Yes    No       Quicken    Microsoft Money    Software provided by my bank

Other (please specify)

9. What is your average checking account balance?       \$0- \$500    \$500-\$999    \$1,000-\$1,499

\$1,500-\$1,999    \$2,000-\$2,499    \$2,500+

10. How much do you have in savings (CDs, Money Markets, savings accounts, etc.)?       \$0-\$1,000    \$1,000-\$4,999    \$5,000-\$9,999

\$10,000-\$19,999    \$20,000-\$29,999    \$30,000+

11. How much do you have invested (stocks, bonds, options, etc)?       \$0- \$5,000    \$5,000-\$9,999    \$10,000-\$19,999

\$20,000-\$29,999    \$30,000-\$39,999    \$40,000-\$49,999

\$50,000-\$74,999    \$75,000-\$99,999    \$100,000+

12. Do you invest online?      13. If yes, what online broker(s) do you use?

Yes    No       E\*Trade    Ameritrade    Suretrade    Datek Online    Charles Schwab

Other (please specify)

14. What credit cards do you carry?

None    Visa    MasterCard    Discover    American Express

Other Please Specify

15. Have you ever obtained an auto/homeowners insurance quote online?    Yes    No

16. What company do you insure your auto(s) with?

State Farm    Allstate    American Family    Farmers    Independent Agent

Other (please specify)        I have no auto insurance

17. What company do you carry your homeowners or renters insurance with?

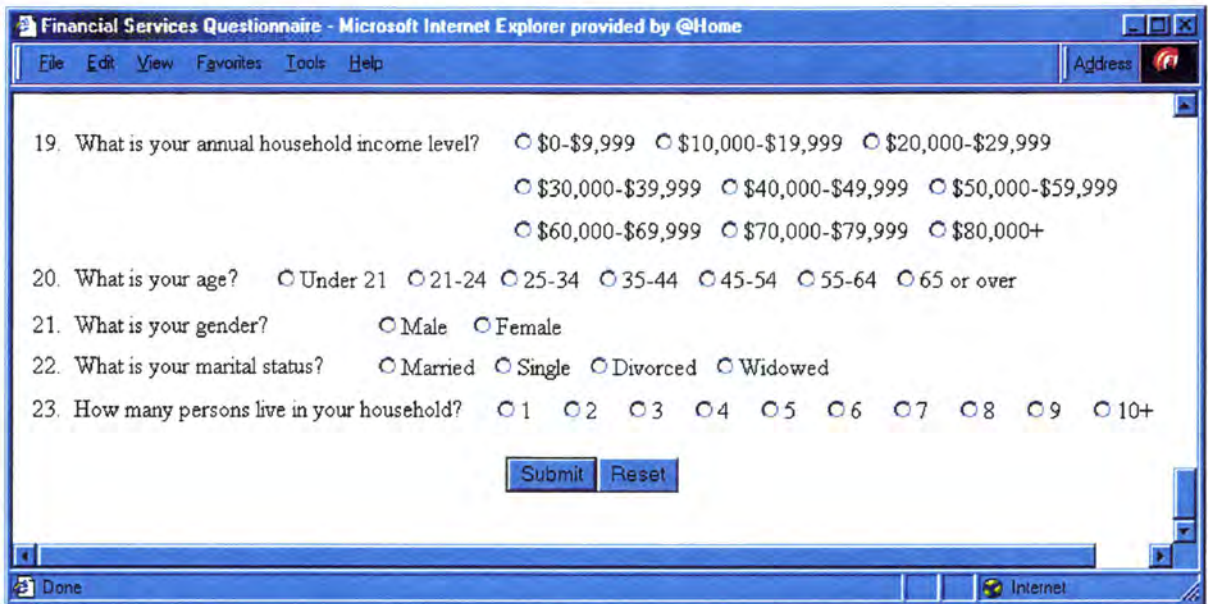
State Farm    Allstate    American Family    Farmers    Independent agent

Other (please specify)        I have no homeowners or renters insurance.

18. What is your field of occupation?    Professional    Clerical    Technical    Managerial

Labor    Administration    Education    Student    Other (please specify)

Figure D. continued



Financial Services Questionnaire - Microsoft Internet Explorer provided by @Home

File Edit View Favorites Tools Help Address

19. What is your annual household income level?  \$0-\$9,999  \$10,000-\$19,999  \$20,000-\$29,999  
 \$30,000-\$39,999  \$40,000-\$49,999  \$50,000-\$59,999  
 \$60,000-\$69,999  \$70,000-\$79,999  \$80,000+

20. What is your age?  Under 21  21-24  25-34  35-44  45-54  55-64  65 or over

21. What is your gender?  Male  Female

22. What is your marital status?  Married  Single  Divorced  Widowed

23. How many persons live in your household?  1  2  3  4  5  6  7  8  9  10+

Submit Reset

Done Internet

Figure D. continued

## APPENDIX E. FOLLOW-UP QUESTIONNAIRES

We are interested in your feelings about participating in a survey posted on the Internet, particularly the one you just completed. Please answer the following question.

Do you feel that you were anonymous when submitting this survey or do you think you could be identified?

I was anonymous     I could be identified

Figure E1. Access ID only and No Access ID Follow-up Questionnaire

We are interested in your feelings about participating in a survey posted on the Internet, particularly the one you just completed. Please answer the following questions.

#1. Do you feel that you were anonymous when submitting this survey or do you think you could be identified?

I was anonymous     I could be identified

#2. Did the promise of displaying a graph of the cumulative results of the survey to date have an important motivating effect to get you to participate in the survey?

Very Much     Somewhat     Neutral     Not very     Not at all

Click on the submit button to proceed to the display of the survey results.

Figure E2. Results Display Follow-up Questionnaire

We are interested in your feelings about participating in a survey posted on the Internet, particularly the one you just completed. Please answer the following questions.

#1. Do you feel that you were anonymous when submitting this survey or do you think you could be identified?

I was anonymous    I could be identified

The following statement was included in your e-mail invitation and in the survey: **This survey is for research purposes only. The information you provide will be kept strictly confidential and used solely for the purpose of this study. No information will be given to third parties.**

#2. How important was this statement in your willingness to participate in the survey?

Very important    Somewhat important    Neutral    Not very important    Not at all important

#3. How important was this same statement in your willingness to answer all questions in the survey?

Very important    Somewhat important    Neutral    Not very important    Not at all important

Figure E3. Confidentiality Statement Follow-up Questionnaire

We are interested in your feelings about participating in a survey posted on the Internet, particularly the one you just completed. Please answer the following questions.

#1. Do you feel that you were anonymous when submitting this survey or do you think you could be identified?

I was anonymous    I could be identified

The following statement was included in your e-mail invitation and in the survey: **This survey is for research purposes only. The information you provide will be kept strictly confidential and used solely for the purpose of this study. No information will be given to third parties.**

#2. How important was this statement in your willingness to participate in the survey?

Very important    Somewhat important    Neutral    Not very important    Not at all important

#3. How important was this same statement in your willingness to answer all questions in the survey?

Very important    Somewhat important    Neutral    Not very important    Not at all important

#4. Did the promise of displaying a graph of the cumulative results of the survey to date have an important motivating effect to get you to participate in the survey?

Very Much    Somewhat    Neutral    Not very    Not at all

Click on the submit button to proceed to the display of the survey results.

Figure E4. Confidentiality Statement and Results Display Follow-up Questionnaire

## APPENDIX F. RESULTS DISPLAY

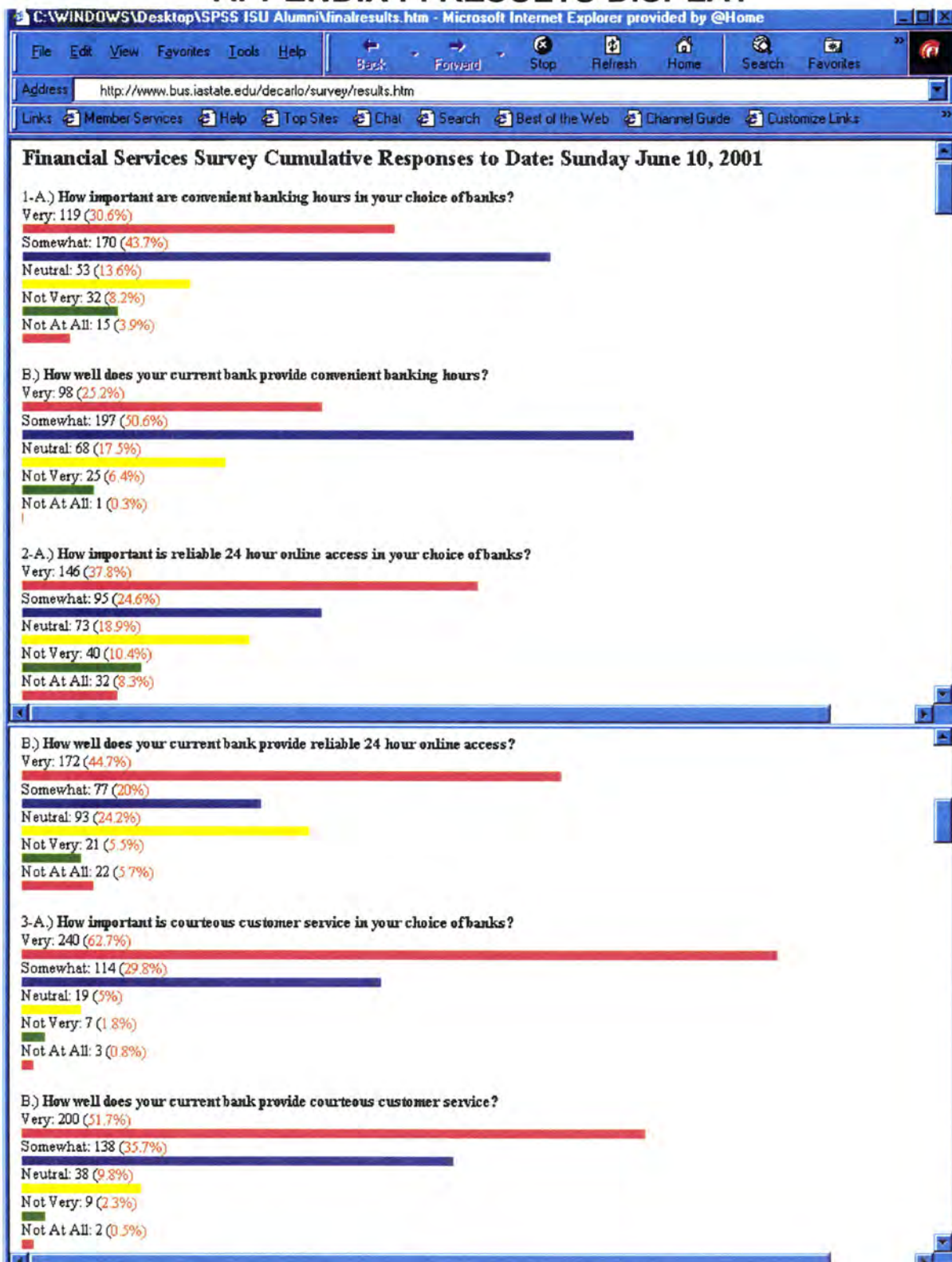


Figure F. Results Display Graph



Figure F. continued



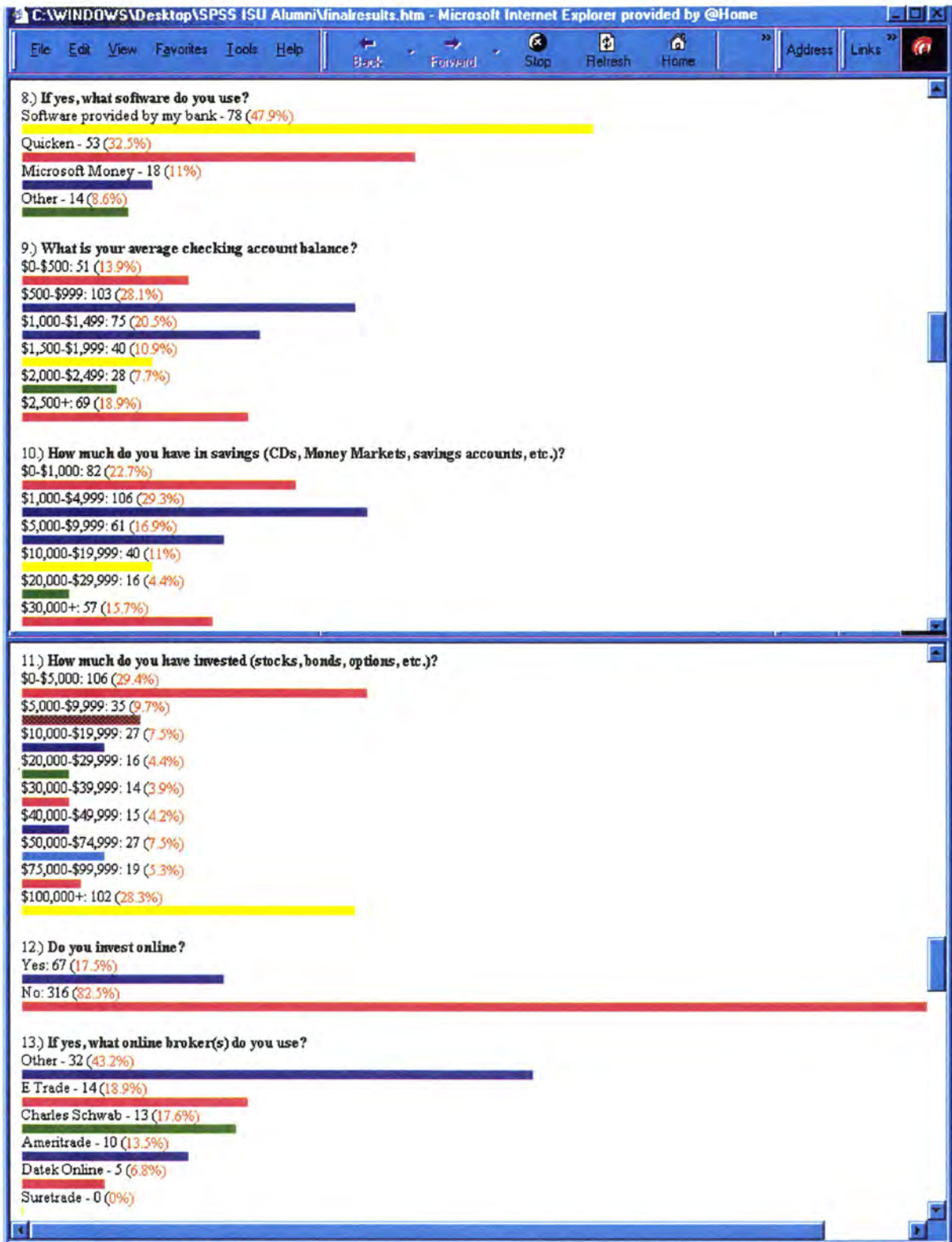


Figure F. continued

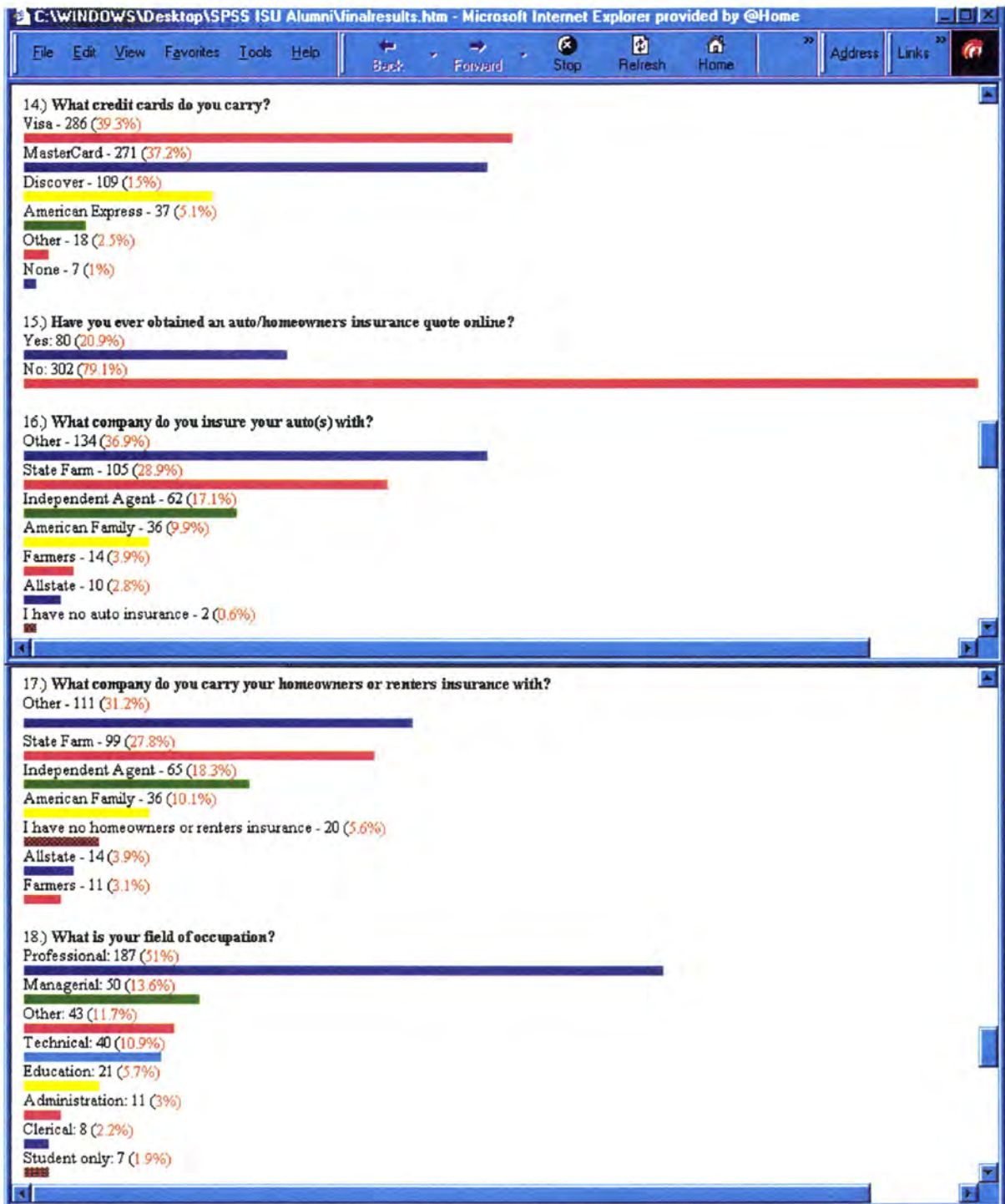


Figure F. continued

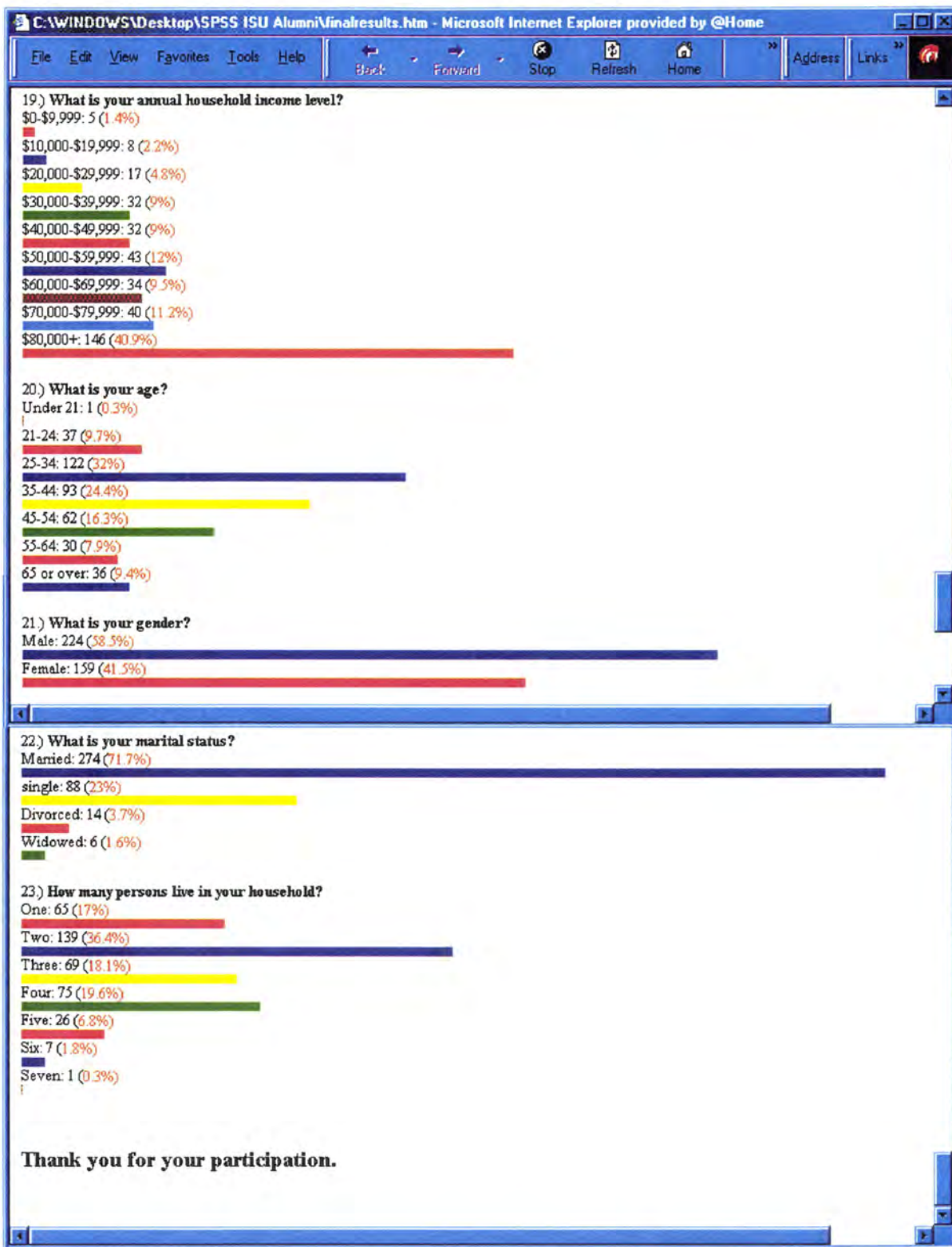


Figure F. continued

## REFERENCES CITED

- Cleland, Kim (1996). "Online research costs about half that of traditional methods," *Advertising Age's Business Marketing* 81(4), B8-B9.
- Hagel, John III and Singer, Marc (1999). "Private Lives," *The McKinsey Quarterly* (1), 7-15.
- Huei-Fen, Mary Wu (1998). "Use of Web-based Survey Instrument to Collect, Analyze and Present Course Evaluation Data," Ph.D. diss., George Washington University.
- Kantor J. (1991). "The effects of computer administration and identification on the Job Descriptive Index," *Journal of Business & Psychology* 5, 309-323.
- McCullough, Dick (1998). "Market research on the Web: guidelines for success," *Communication World* 15(9), 29.
- Mertens, Donna M. (1998). *Research Methods in Education and Psychology: Integrating Diversity with Quantitative & Qualitative Approaches*. Thousand Oaks, CA: Sage Publications.
- Milne, George R. & Boza, Maria-Eugenia (1999). "Trust and Concern in Consumers' Perceptions of Marketing Information Management Practices," *Journal of Interactive Marketing* 13(1), 5-24.
- Nowak, Glen J. & Phelps, Joseph (1997). "Direct Marketing and the Use of Individual-Level Consumer Information: Determining how and When "Privacy" Matters," *Journal of Direct Marketing* 11(4), 94-108.
- National Research Council (U.S.) Panel on Privacy and Confidentiality as Factors in Survey Response (1979). *Privacy and Confidentiality as factors in Survey Response* Washington, D.C.: National Academy of Sciences.
- Rasinski, Kenneth A., Willis, Gordon B., Baldwin, Alison K., Yeh, Wenchi and Lee, Lisa (1999). "Methods of Data Collection, Perceptions of Risks and Losses, and Motivation to Give Truthful Answers to Sensitive Survey Questions," *Applied Cognitive Psychology* 13, 465-484.
- Schillewaert, Langerak, and Duhamel (1998). "Non-probability sampling for WWW surveys: a comparison of methods," *Journal of the Market Research Society* 40(4) 307(1).

Singer, Eleanor, Hippler, Hans-Jurgen and Schwarz, Norbert (1992). "Confidentiality Assurances in Surveys: Reassurance or Threat?" *International Journal of Public Opinion Research* 4(3) 256-268.

Schmidt, William C. (1997). "World-Wide Web survey research: Benefits, potential problems, and solutions," *Behavior Research Methods, Instruments, & Computers* 29(2) 274-279.

Stanton, Jeffrey M. (1998). "An Empirical Assessment of Data Collection Using the Internet," *Personnel Psychology* 51, 709-725.

Zhang, Yin (2000). "Using the Internet for Survey Research," *Journal of the American Society for Information Science* 51(1), 57-68.