Communications of the Association for Information Systems

Volume 9 Article 13

October 2002

The Impact of Demographics on Choice of Survey Modes: Demographic Distinctiveness between Web-Based and Telephone-based Survey Respondents

Pratim Datta
Louisiana State University, pdatta2@lsu.edu

Kenneth R. Walsh
University of New Orleans

Dek Terrell
Louisiana State University, mdterre@lsu.edu

Communications of the Association for Information Systems

THE IMPACT OF DEMOGRAPHICS ON CHOICE OF SURVEY MODES: DEMOGRAPHIC DISTINCTIVENESS BETWEEN WEB-BASED AND TELEPHONE-BASED SURVEY RESPONDENTS.

Pratim Datta
Information Systems and Decision Sciences
Louisiana State University
pdatta2@lsu.edu

Ken Walsh
Department of Management
University of New Orleans

Dek Terrell
Department of Economics
Louisiana State University

ABSTRACT

Web surveys potentially cost less to administer and are more convenient for participants than either telephone or paper-based surveys, but concern remains about the representativeness of the respondents to the population. This research investigates the effectiveness of using Web-based survey methods by comparing response rates and demographic characteristics of the telephone-based and web-based survey respondents. The paper first reviews survey research methodologies, including recent trends in web-based surveys that take advantage of the power of modern computing and telecommunications technology. The issue of respondent characteristics in web-based surveys is then explored by comparing respondent demographics of an ongoing telephone survey to those of a subsequent Web survey of the same population. Exploratory and confirmatory statistical analyses are used to triangulate the findings and test the hypotheses. The data suggests that demographics of the respondents of the two methods are similar across race and age, but differ significantly across income and education levels, implying a converging digital divide.

KEYWORDS: web surveys, telephone surveys, demographic studies, statistical analyses, digital divide

I. INTRODUCTION

Computing and telecommunications increased the reach of and access to markets around the globe, making it cheaper, easier, and faster to gather feedback from target groups through surveys. In a 1998 issue of MIS Quarterly, Newsted et al. [1998] pointed out the growing popularity of survey research in the IS community, prompting the need for informing IS researchers of the effectiveness of survey instruments for better IS research. Their effort led to the "development of a website of survey instrument for IS researchers as 'Living Scholarship'" [Newsted et. al, 1998 p. 554]. As survey instruments become more refined and evolve to take advantage of newer technologies, especially the Internet, both practitioners and researchers in IS

are faced with a choice of using web-based surveys as their survey instrument of choice. However, there are associated ambiguities concerning its effectiveness.

Access to the underlying technologies of computing and telecommunications is a salient determinant of the efficacy of survey instruments. With an estimated 96% of U.S. households having access to telephones, telephone-based surveys gained strong support within the survey research community [American Statistical Institute, 1999]. However, the growth of the Internet increased web access across traditional geographic, and temporal boundaries. Computer-mediated environments create opportunities as they are neither temporally nor geographically bound [Hoffman and Novak, 1996; Hoffman et al., 1995]. These characteristics have led academics and practitioners to develop an interest in understanding the legitimacy of web-based survey methodologies over traditional survey methods. Rising interest in web-based surveys [Dillman et al. 1998] reached a point where web surveys were even held as the researcher's panacea [Couper, 2001; Dillman and Bowker, 2001], reaching "hundreds of thousands of people" and processing "results for a cost that would previously only have allowed a few hundred responses to be tabulated" [Dillman and Bowker, 2001, p. 4]. Yet, little empirical work on the utility of web-surveys is published and little is known about how well web surveys represent the population in terms of coverage [Couper, 2000]. To address this vacuum, this study compared demographic factors and response rates of telephone and web-based surveys.

In testing how demographics affect response rates, this research sheds light on the continued existence of a "digital divide," the demographic heterogeneity of use and access to the digital medium (National Telecommunications and Information Administration, 2000). High telephone penetration rates made the telephone a useful tool for reaching representative population samples. However, home adoption rates of computers and Internet access are much lower, raising questions about the representativeness of samples collected using such technology. Are web-based survey respondents representative of the same population as telephone-based survey respondents, and, if not, what are the distinct demographic factors that underlie the disparity in accessing information technology? This study sampled the population of Louisiana to reveal whether web-based and telephone-based survey respondents share the same demographic makeup.

Section II reviews the available survey methods and identifies trends. Section III discusses web surveys and presents the hypotheses relating to demographic characteristics of respondents involved in different survey methods. Research methodology is described in Section IV. The article ends with the results (Section V), limitations, and discussion of the findings (Section VI).

II. SURVEY METHODS

A survey is defined as a "method of gathering information from a representative sample of a population being studied for a specific purpose" [Ferber et al. 1980, p. 2]. Unlike a census, surveys focus on a representative sample rather than the entire population. In a survey, the sample must be chosen randomly to prevent sample selection biases. The need to maintain anonymity and to create a composite profile of the population is important issues in conducting bona fide surveys [Ferber et al., 1980].

Five basic steps constitute survey research [Ary et al. 1996]:

- Planning,
- Sampling,
- Constructing the instrument,
- Conducting the survey, and
- Processing the data.

It is in these steps where the available survey methodologies elicit both problems and promises. Available survey methodologies include mail, personal interviews, telephone, and web-based surveys. Trochim [1991] divides surveys into two broad categories of questionnaire-based methods and interview-based methods. Questionnaire-based methods prompt respondents to

complete the surveys while interview-based methods require the administrator to complete the survey. Mail and web-based surveys constitute self-administered questionnaire-based survey methods, while personal and telephone interviews constitute interview-based survey methods.
As technology evolved and temporal and geographical constraints were reduced, use of survey methods evolved in parallel. Fowler [1993] notes that while personal interviews would have been the survey "method of choice" twenty-five years ago, more recently, cost advantages and the widespread availability of telephones made telephone surveys the preferred method. By 1993, Fowler reported that better and faster responses could be attained with telephone rather than mail surveys or personal interviews [Fowler, 1993, p. 67-68].

Telephone-based surveys show a variety of advantages that prompted survey researchers to adopt them. Some of the advantages are:

- Higher response rates and cooperation [Fowler and Mangione, 1982; Groves and McGonagle, 1999; Nederhof, 1988; Day et al., 1995].
- High population representation [American Statistical Institute, 1999].
- Better resolution of respondent misunderstanding [Couper and Groves, 1992] together with clearing up ambiguous questions or misunderstandings within the survey.
- Easier administration of telephone reminder calls which, in turn, generate higher response rates. Telephone calls are seen to be as effective as certified mailings [Nederhof, 1988; Day et al., 1995].

Some of the disadvantages are:

- Reluctance to participate because of reluctance to answer specific questions, or an inability to answer these questions [Blair and Chun, 1992; Triplett et al., 1996; De Heer and Israels, 1992; De Leeuw and De Heer, 1999]
- The young and the educated segments of the population are harder to reach because of their busy lifestyle revolving around work, education, and/or outdoor activities [Groves and Couper 1998].
- Higher costs per completed interview because of the inherent overheads and fixed costs as well as labor hours for interviewers and facility coordinators [Stevens, 2000].
- Problems associated with telephone surveys encountering telephone answering machines [Oldendick and Link 1994; Piazza 1993], call screening via Caller ID machines [Link, 1999], and a lack of participation associated with the increase in telemarketing [Direct Marketing Association, 2000].

The high (96%) level of penetration of telephone access in U.S. households is the primary cause for the popularity of telephone surveys as a survey research tool [The American Statistical Institute, 1999]. Telephone survey methods were refined over time with the introduction of new statistical and technological techniques for data collection. Robust statistical sampling methods such as the Mitofsky-Waksberg method [Waksberg, 1978] and the use of commercial lists were incorporated by telephone surveys for efficient representative sampling.

With the convergence of computing and communication technologies, improved telephone survey methods such as Computer Assisted Telephone Interviewing (CATI) emerged. In 1990, the U.S. Subcommittee on Computer-Assisted Survey Information Collection concluded, "CATI has become a standard collection vehicle grounded strongly in a firm body of research" [Gonzalez, 1990, p. 1]. A full-featured CATI system enables interviewers to enter responses to each question directly into the computer, supported by a user-friendly format that reduces common errors such as entering incorrect dates. CATI "provides an exact report on the disposition of each call made and the scheduling algorithm helps to ensure that respondents represent the target population" [University of Kentucky Survey Research Center, 2000]. CATI, therefore, among other potential benefits, promised a better survey by utilizing robust statistical techniques along with computing technology [Wilson et al., 2001].

Thus, in the wake of robust methodologies such as CATI and the level of penetration, telephone-based surveys still remain the predominant currency among other household survey

methodologies, especially when timeliness is an issue, and the length of the survey is limited [American Statistical Institute, 1999].

Further technology trends paved the way for more effective modes of data collection for surveys. Rapid advances in Internet technology spurred the use of the medium for data collection, especially with evidence that Internet surveys resulted in lower cost, quicker response time, and higher probability of responses to open-ended questions [Bachman, 2001]. Further, since 1995, email and Internet use spread to a "wider population" [Bachman, 2001]. As access to telephones and the Internet increased, telephone and Web-based surveys increased in scale and scope.

III. WEB BASED SURVEY METHODS AND RESEARCH QUESTIONS

The paradigmatic shift from mail, interviews, and telephone surveys towards newer forms of survey tools are trends in market research [Mehta and Sivadas, 1995]. Online tools are not constrained by geographic location, proximity, or time zone [Foster, 1995]. Furthermore, with diminishing demographic disparities, online research tools are likely to become more popular [Coomber 1997]. Couper [2001] points out that web surveys already impact the survey industry with a rapidity and breadth like few other innovations before. The proportion of computer owners among the U.S. adult population is around 51% and approximately 41.5% among them have access to the Internet, with email being the most widely used application (approximately 79%) among computer users [U.S. Census Bureau, 2000]. The potential for online survey research led to the introduction of a variety of tools such as e-mail questionnaires, web-based surveys, and converted CATI systems to translate and administer questionnaire over the Web. Web survey systems use user-friendly interfaces to create visually appealing yet complex questionnaires that are more cost effective than any of the other survey tools [Watt, 1997]. Furthermore, from a survey researcher's standpoint, Web survey research was found to yield meaningfully comparable data about both Internet users and larger populations [Witte et al., 2000]. However, little is known about how demographically representative web-survey respondents are to the population as a whole.

DSS Research [2001] finds the following advantages of online survey research:

- Web Surveys are more cost and time effective than offline surveys due to the cost reductions in postage, transcription costs, and stationery costs increasing with sample size despite being temporally unconstrained [Watt, 1997; Bainbridge, 1999; Brewer, 2000].
- Communicating with respondents and delivery of surveys is easier and faster [Coomber, 1997; Chen and Hinton, 1999], especially without the hindrances of answering machines and caller IDs.
- Web-based surveys are a valuable access tool for targeting high income, young, and educated demographic subgroups that are normally harder to reach via telephone surveys but are continuously expanding their technological domain [Kehoe and Pitkow, 1996]. They are also the early adopters of new products and technologies.
- Sensitive questions and difficult to reach groups can be better targeted and reached via Web surveys [Coomber, 1997]. Individuals are more likely to reveal themselves in self-administered surveys in their own privacy rather than being guided or supervised by an interviewer. Open-ended questions are also better answered without the onus of time conveyed by an interviewer.
- Web surveys use technologies such as Java applets, JavaScript, VBScript, and ActiveX to enhance and customize web pages to make them interactive and ad hoc, giving respondents a degree of control [Markham, 1998]. Survey questions can be adaptive to previous responses while eliminating systemic problems.

On the other hand, the following are disadvantages of Web surveys:

- Web surveys may fail to represent the U.S. population properly because the digital divide in the population limits Web access. The more educated and relatively higher income

people are the predominant users of the service [Coomber, 1997]. It may take a generation before Web surveys are reasonably representative [Eaton 1997],

- Concerns about the security of Web surveys may result in dampening the benefits of self-administered surveys with sensitive questions, leading to a low response rate [Dommeyer and Moriarty, 2000; Couper, 2000].
- Together with problems with effective sampling, coverage, statistical reliability, and non-response errors, Web surveys suffer from measurement errors that may result from lack of comprehension, survey design problems, and lack of motivation. [Couper et al., 1999; Couper, 2000; Bainbridge, 1999].
- Web survey respondents are sometimes impatient with high-burden Web interactions and problems associated with browser interactions. A survey by the University of Michigan on the effects of password entry, browser modification, and notices found that both perceived and real burden associated with Web interactions impacted response rates in Web surveys [Crawford et al., 2001].

This article focuses on the differences in demographic characteristics between CATI telephone and web-based survey respondents. A Forrester research study on "consumer technographics" [cited in Weiss, 2001] looked only at Internet users and how they make online-purchase decisions. In this study we examine the demographic composition that results in the access and acceptance of the Internet and the web-based survey medium. This study raises pertinent questions on whether demographic characteristics of web-based survey respondents are indeed representative of the population; thus reexamining the demographic variables that define the "digital divide," a pertinent issue that looks at demographic factors that are associated with acceptance and use of Internet technologies by the population.

The digital divide is primarily linked to the disparities in access to technologies, largely because of penetration. Disparate access to technology was found to be a primary detriment to the adoption of such single modes of survey [Dillman, 2000]. –While the penetration of telephone services is 96% of the U.S. households [American Statistical Institute, 1999] only about 25% of households can access the Internet [U.S. Census Bureau, 2000 cited in Couper, 2001], making demographic characteristics an issue linked to the penetration of technologies. Telephone surveying, because of its high degree of household penetration, is largely considered an effective survey mode. Web surveying, on the other hand, still suffers from a low degree of household access and penetration, creating potential disparities in representing the population. Coomber [1997] argued that the demographic disparities in the digital divide are diminishing rapidly, thus paving the way for web-based survey respondents to be representative of the population. Given this scenario, one could infer that web-based surveys could represent the population well and be an effective survey mode. Other researchers cite that demographic factors such as age [Pew Internet Project, 2001], race [Rogers and Oder, 1999], income [Kehoe and Pitkow, 1996; Eaton, 1997], and education [Kehoe and Pitkow, 1996; Eaton, 1997] vary among respondents with or without access to and use of the Internet, affecting the choice of survey modes.

Thus, to examine the demographic composition of survey respondents, this paper tests the following four hypotheses:

H_1: Age is a distinguishing demographic characteristic between web survey respondents and telephone survey respondents.
H_2: Race is a distinguishing demographic characteristic between web survey respondents and telephone survey respondents.
H_3: Income is a distinguishing demographic characteristic between web survey respondents and telephone survey respondents.
H_4: Education is a distinguishing demographic characteristic between web survey respondents and telephone survey respondents.

IV. RESEARCH METHODOLOGY

To study differences in response populations between telephone-based and web-based survey methods we teamed with researchers conducting the monthly Louisiana Consumer Confidence Survey. This alliance gave us access to a telephone survey using a random dialing method to

create a well-randomized sample and a survey that collected demographic information. In the first phase of the project, we used questions added to the consumer confidence survey to determine the respondent's ability and willingness to participate in a web-based survey. Email addresses were collected for use in phase 2, a follow up web-based survey. Couper [2000] refers to a similar multistage and multimode sampling method incorporated by Intersurvey. The collection of email addresses from a random telephone survey was used so that biases were limited to those relating to willingness and ability to participate in a Web based survey rather than using email lists from ISPs that would have introduced additional biases.

In considering interface and survey design issues, we focused on simple single-response multiple-choice questions where respondents were asked to identify one response from either categorical or yes/no questions. The only exception was an open-response asked during the first stage telephone interviewing using CATI where respondents were asked to provide their email addresses. In analyzing response effects in computer-administered questioning, response expressions were converted into numeric integer values for analysis. Furthermore, the interface design mirrored the primary consumer confidence survey offered using CATI.

In the telephone survey, respondents were told that their telephone number was generated at random and they could decline to answer any of the questions. Therefore, the telephone survey was confidential except that respondents were asked if they were willing to participate in an online survey and, if so, were asked their email address. The email question was asked last in the survey so as not to affect response rates. Our procedure stored responses to the telephone survey without the telephone number used to dial the respondent. The email information from CATI respondents was used to create an email mailing list. Each email recipient was sent the URL of the Web based survey accessible via a randomly generated ID without storing any identifying information. To control for multiple responses, the Web based system monitored IP addresses independently of survey results. No duplicate IP addresses were used to enter the system. No incentives were used in either survey. The format for data collection was as follows: Age was denoted as a discrete variable ranging between 18 and 110. Education was also denoted as a discrete variable ranging between 1 and 18 based on the years of education. Race was categorized as a seven-item scale with 1 for White, 2 for Black, 3 for Hispanic, 4 for Asian, 5 for American Indian, 6 for Other, and 7 for Multiracial. Income was categorized into a 10-item scale ranging from less than $10,000 to greater than $150,000. The survey administered stipulated confidentiality, leaving the reporting of demographic characteristics optional for the respondents. Among the responses, only the variable "Age" was found to have 22 missing values In this case, the missing values for age were replaced using a sample mean replacement (mean imputation). In respect to the rest of the variables, all entries were complete.

In phase 2, after collecting telephone surveys for one year, a web-based survey was conducted. All respondents to the telephone survey who provided email address were sent an email with a URL pointing to the web-based survey. The web-based survey asked the same questions as the telephone survey, but omitted the added questions regarding ability and willingness to participate in Web-based surveys. An email requesting participation and containing the URL of the survey was sent on a Wednesday at 2:30pm and the survey was made available until an extended period-of-time had elapsed with no new respondents. Data was collected and recorded in the same format designated for telephone surveys.

V. RESULTS

The primary 12-month survey data from CATI yielded 6844 household responses. Out of these 6844 responses, 2280 or 33.3% of the total households had Web access. Of those with Web access, 639 or 28.0% of the total households with Web access indicated a willingness to participate in an online survey and provided an email address. Using email data of the 639 respondents, an email was sent to each asking them to participate in a web-based survey. The URL to the survey was provided in the email and respondents were asked to click on it to proceed to the Web based survey. Web Surveyor, a program for generating web-based surveys, was used as the enabling tool. Immediately after the emails were sent, 181 emails were returned due to delivery failures stemming from wrong domain name or user name. The 181 delivery failures were coded according to Simple Mail Transfer Protocol (SMTP) error codes to check for their

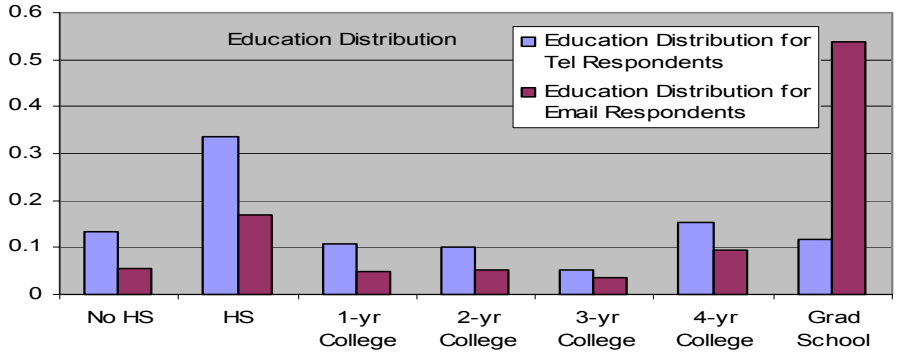
causes and their distribution. Among the 181 email delivery failures, 40 were attributed to bad usernames and the remaining 141 attributed to bad domain names. The rest of the 458 emails generated 51 responses for the web survey, an 11.14% response rate. Alreck and Settle [1995], Gay and Diehl [1992] posit a sampling proportion of 10% of the population to be representative of the population statistic.

Multiple statistical analyses were employed to triangulate the consistency of the observed results. Triangulation techniques are employed by testing the model using student's t-tests for difference of means, a follow-up exploratory ANOVA (F-tests), furthered by a confirmatory structural equation model (SEM). Each of these statistical techniques is used for analyses of the consumer-confidence survey data and examining the findings. We begin by evaluating Hypotheses H_1 to H_4 using t-tests and one-way analysis of variance (ANOVA) (with the exception of chi-square tests for nominal data) to observe if the means of each demographic variable are significantly different across telephone-based and web-based survey respondents. The summary statistics are presented in Table 1.

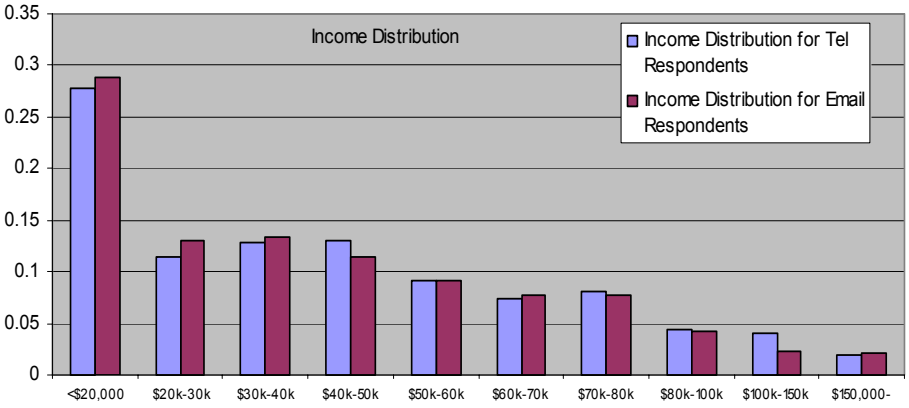
Table 1.Summary Statistics and Plots for Interval Data
(Standard Deviations Shown in Parentheses)

Means and Standard Deviations of Interval Data	Age	Income	Education
Sample (n=6844)	44.06 (16.93)	3.91 (2.36)	13.32 (2.78)
Telephone Survey (unwilling to participate in online survey) (n=6205)	44.03 (16.94)	3.89 (2.36)	13.37 (2.72)
Telephone Survey (willing to participate in online survey) (n=639)	43.7 (16.74)	3.92 (2.19)	13.57 (2.74)
Telephone Survey (w/ non-valid email access or no email address) (n=181)	43.67 (16.81)	3.94 (2.20)	13.53 (2.75)
Telephone Survey (with valid email) (n=458)	46.39 (14.97)	5.32 (2.16)	15.43 (2.35)
Final Survey Sample (n=6256)	44.06 (16.93)	3.91 (2.36)	13.39 (2.72)
Web Survey (n=51)	47.33 (15.22)	5.41 (2.37)	15.33 (2.15)

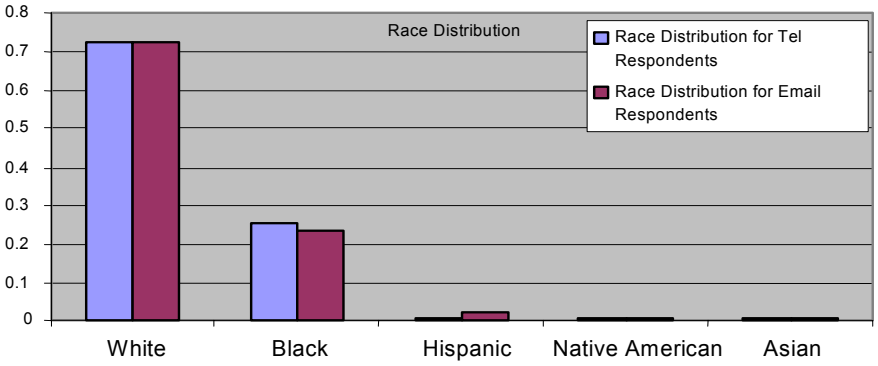
Descriptive and summary statistics together with related plots of the hypothesized demographic variables across web-based and telephone based survey respondents are shown in Table 1 and Figure 1. The mean level of education for telephone respondents was 13.37 years, matching qualifications for college freshmen while the mean level of education for web survey respondents is 15.39 years, matching a college senior. Standard deviation for education is also higher among telephone respondents. The minimum educational level recorded for telephone respondents is a one-year elementary education compared to a minimum high school education recorded for web survey respondents. The means for income for telephone respondents show a lower distribution of income than that of web survey respondents with an approximate difference of $ 16,000. Lastly, central tendency statistics for age demographics is slightly higher at 47.93 years for web survey respondents compared to 44.03 years for telephone respondents.



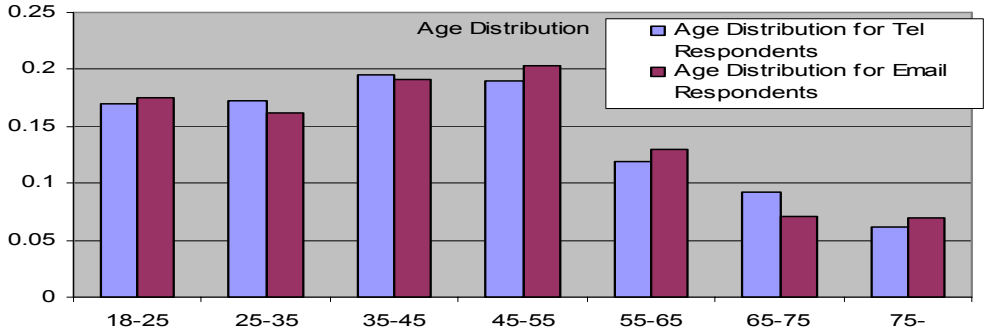
a. Education Categories



b. Income Categories



c. Race Categories



d. Age Categories

Figure 1. Demographic Distribution

The cumulative total response plot in Figure 2 indicates the disparate densities in responses with a distinct drop in responses after the first week of receipts. Out of 51 responses, only 3 (5.9%) were returned 9, 12, and 22 days after the first response was received- portraying a steep decrease in response rate with an increasing time lag).

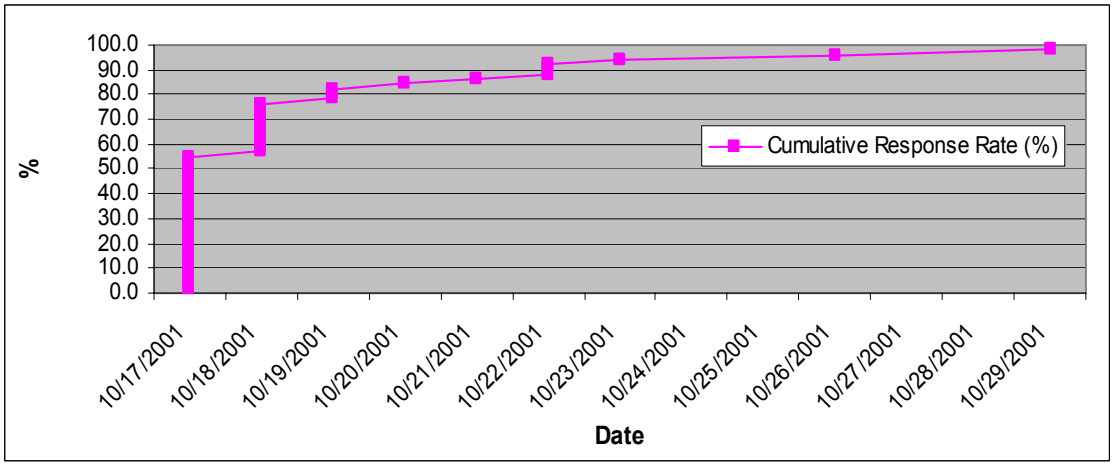


Figure 2, Timeline of Web Cumulative Responses

A preliminary t-test is performed to determine whether the variances between the non-respondent sample (n= 588) with email address and respondent sample with email address (n= 51) are significant. The results show that the variances are not significant at $\alpha_{0.05}$ across age (p-value > 0.05), race (p-value > 0.05), income (p-value > 0.05), or education (p-value > 0.05), making the demographic composition of the respondent sample with email addresses similar to that of the non-respondent sample with email addresses.

It is important to understand if there are discernible differences between respondents and non-respondents. The effect of nonresponse on survey estimates depends on whether the non-respondents are biased and systematically different from the respondents (Fowler, 1993). In

addressing non-response in our paper, we find no evidence of any systematic non-respondents bias.

Similar to the findings by the NTIA [National Telecommunications and Information Administration, 1999, cf. Couper, 2000], the plots highlight that the characteristics of higher education, higher income, whites, and young to middle aged individuals stand out as prominent demographic segments with greater access to technology.

Next, to investigate the importance of demographic characteristics in yielding significant differences in the variable means as our hypotheses, t-tests followed by a one-way ANOVA are performed.

Because we are trying to evaluate the differences in means between two groups (telephone-based vs. web-based respondent demographics) we begin by conducting a parametric student's t-test of difference of means with $\alpha_{0.05.}$ The results indicate that the means for age (t-test = 1.386; p-value > 0.05) and race (t-test = 1.239; p-value > 0.05) are not significantly different while the means for education (t-test = 5.1432; p-value < 0.01) and income (t-test = 4.585; p-value < 0.01) are significantly different at even a 99% level of confidence. However, t-tests are generally used for small sample sizes and reliant on the assumptions of normality and homogeneity of variances. In general, for every t-test there is a matching F-test with 1 numerator degree of freedom. When this is the case, the t-test will have degrees of freedom equal to the F-test's denominator degrees of freedom and the observed F-value will be the square of the observed t-value ($F_{(1, df)} = t^2_{(df)}$).
We conducted a one-way ANOVA for a matching F-test. In comparing two means (i.e. telephone-based and web-based respondent demographics), the ANOVA gives the same results as the Student's t-test for independent samples while providing a greater degree of robustness and analytical power (Levin, 1999). ANOVA tests each factor while controlling for the other, thus making it more statistically powerful than t-tests. Furthermore, F-tests are remarkably robust to deviations from assumptions such as normality and homogeneity of variances [Box, 1954]. The results from the ANOVA (F-tests) mirror the results from the t-tests. Findings show that hypotheses H_1 and H_2 (age and race) are not supported while hypotheses H_3 and H_4 (education and income) are supported at a 95% level of confidence ($\alpha_{0.05}$).

Education (F-test- 26.453; p-value < 0.05) and Income (F-test- 21.019; p-value < 0.05) show significant differences between web and telephone survey respondents at a 0.05 level of significance. However, as indicated in the ANOVA test for age (F-test- 1.92; p-value > 0.05) and the non-parametric Chi-Square test for race (χ^2 = 12.676; p-value > 0.05) (F-test- 21.019; p-value > 0.05), neither race nor age are not found to be significant at a 0.05 level of significance, while income and education were found significant at even a 0.01 level of significance. The tests imply that age and race demographics do not significantly differ between telephone-based and web-based survey respondents, while income and education are significantly different between telephone-based and web-based respondents.

The results from the t-tests, ANOVA, and chi-square tests show significant differences among some of the demographic characteristics between web-based and telephone-based survey respondents. The results support the hypotheses as follows:
H_1: Not Supported. Age is not a distinguishing demographic characteristic between web survey respondents and telephone survey respondents.
H_2: Not Supported. Race is not a distinguishing demographic characteristic between web survey respondents and telephone survey respondents.
H_3: Supported. Income is a distinguishing demographic characteristic between web-survey respondents and telephone survey respondents.
H_4: Supported. Education is a distinguishing demographic characteristic between web survey respondents and telephone survey respondents.

We checked our findings by enhancing the exploratory findings from t-tests and ANOVA with a confirmatory structural equation model (SEM). The focus is on enhancing the exploratory findings by revealing and confirming associations between the demographic variables in context. Figure 3 shows paths and associations between the variables, creating a model that explicates the linear

structural relationships between the variables, indicating predictive validity [Joreskog and Sorbom, 1999; Mueller, 1996]. In this model, the latent variable labeled "other" is stipulated to absorb random variation of other variables for which no suitable predictors were denoted in this preliminary model. In order to identify the model, a regression weight of 1 is denoted between the variable of "Error" and survey response mode of Web/Tel. A summarized covariance and correlation matrix is presented in Table 2.

Table 2. Correlation and Covariance Matrix

Correlation Matrix				
	Age	Race	Income	Education
Age	1			
Race	-0.115**	1		
Income	-0.054**	-0.069**	1	
Education	-0.063**	-0.039**	0.316**	1
Covariance Matrix				
	Age	Race	Income	Education
Age	1			
Race	-1.707**	1		
Income	-2.16**	-0.143**	1	
Education	-2.923**	-0.094**	2.028**	1
**p<0.01				

Amos Structural Equation Modeling package [Arbuckle, 1999] was used for confirmatory analysis (Indices shown in Table 2). Schumacker and Lomax [1996] note that because there is no specific statistical test of significance to identify the validity of the model uniquely, multiple tests of goodness of fit needed to be used. Within these goodness-of-fit indices, Bentler [1992] points out that the Comparative Fit Index (CFI) and the Normed Fit Index (NFI) are the preferred measures for ascertaining model validity [Tan, 2001]. In our model, the structural model for the hypothesized variables indicates that the hypothesized model covariances do fit the actual covariances (CFI= 0.997; NFI= 0.997). The fit is also reiterated in the other indices such as Relative Fit Index (0.991), Incremental Fit Index (0.997), and Tucker-Lewis Fit Index (0.991) that show values greater than 0.90 together with a low RMSEA (< 0.08); confirming a good fit between the model and the actual dataset (Table 3). Figure 3 shows the structural model, confirming past findings and reiterating that while income and education were associated significantly with the choice of web-based surveys and telephone-based surveys, age and race demographics showed a weak and insignificant association. The strength of significant positive association between education and income imply education and/or income as the salient variables that can influence access to technology, creating a digital divide.

Table 3. Fit Measures

Global Fit Indices for the Structural Model	
Relative Fit Index (RFI)	0.991
Normed Fit Index (NFI)	0.997
Incremental Fit Index	0.997
Tucker-Lewis Fit Index	0.991
Comparative Fit Index (CFI)	0.997
Root Mean Square Error of Approximation (RMSEA)	0.7

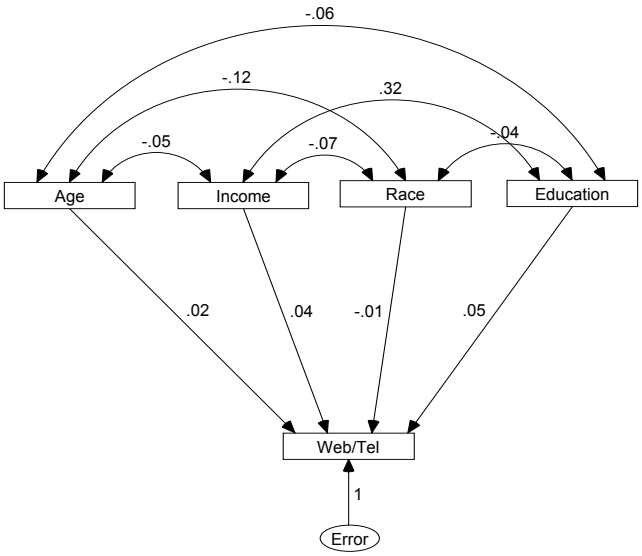


Figure 3. Structural Equation Model

VI. DISCUSSION AND LIMITATIONS

Respondents of telephone-based surveys and web-based surveys are found to be significantly different in terms of income and education. The disparity may indicate the digital divide and information inequality, notions referred to by Novak and Hoffman [1998], Bohland et al. [2000], thus making web-based survey respondents not representative of the whole population at this time and relatively reducing the effectiveness of web-based surveys.
ANOVA results in this paper show similar findings. Income shows a significant association with web survey respondents, affirming a positive relationship of income to access and use of technology [King, 1998; John J. Heldrich Center for Workforce Development, 1999]. Similarly, education is seen to have a significant association with web survey respondents, reaffirming the findings from an investigation by the National Telecommunications and Information Administration [1999]:

> "*only those who have already attained educational success are using the Internet in large numbers. Consequently, Americans with less education, who could perhaps benefit most from the Internet's educational value, are being left behind.*"

Thus, researchers need to understand that education and income of respondents, their target population, determine the effectiveness of the Web as a survey medium. However, the digital divide appears to be narrowing, as indicated in the lack of significance in race and age demographics. Similar findings are reported on race by Conhaim [2001] and on age by Rogers and Oder [2001].

The use of the consumer confidence survey to study the response population in web-based surveys does indicate the existence of a digital divide, manifested by income and education. Still, it is prudent to point out several issues and limitations. The multistage survey sampling technique used CATI to record preliminary data including email addresses. In coding and assigning the undeliverable email addresses as domain name or user name related, it was obvious that a large proportion of those delivery failures were attributable to careless user inputs. This asserts the need for a greater focus by researchers on reducing input errors. Furthermore, the fact that web survey respondents had to take the survey twice could have reduced their participant motivation

and may have contributed to measurement error that stems from a deviation of responses from their true values [Couper, 2000]. However, this limitation may have been assuaged to a degree because of the nature and importance of the content of the consumer confidence survey and the preset interest of respondents in undertaking the survey. Nevertheless, the request for repeating exactly the same survey may have dampened the survey response rate for potential respondents, resulting in a relatively low participation.

Regarding the technical difficulties that may impede user participation [Couper, 2000], we hope to have alleviated the problem by delivering our survey content as a "text only" format. We maintained confidentiality and privacy by explicitly informing respondents that their information would be private, automating access code generation from an email listserv, and not linking telephone numbers to email addresses.

Limitations arising from coverage error also need mentioning. Coverage error results when there is a mismatch between the target population and frame population" [Couper, 2000: p. 467]. This error arises because significant numbers of people do not have access, or choose not to use the Internet. It is perhaps in light of the issue of coverage error that one realizes the importance of understanding the digital divide. The evidence of a digital divide shows that coverage errors still remain a salient concern for researchers in matching their frame population to their target population. This fnding echoes Couper's [2000: p. 466] statement that "coverage error represents the biggest threat to the representativeness of sample surveys conducted via the Internet." We have, however, alleviated the problem of nonresponse error [Couper, 2000] for the frame with email addresses by performing an ANOVA test that showed no significant demographic differences between the respondents and the non-respondents on the variable of interest (individuals with email addresses) [Groves and Couper, 1998].

Another limitation stems from sampling error, which, according to Couper [2000: 467], "arises from the fact that not all members of the frame population are measured," creating problems in constructing a frame population. Considering the fact that this study considers probability-based sample design using probability samples of full population [Couper, 2000], we tried to reduce sampling error by including the entire frame population of households from which we derived our frame that constituted households with access to the web and email that were willing to participate in web surveys.

A final limitation may lie with the generalizability of the survey. The findings may be generalizable only within the state of Louisiana or possible the United States, and may not be valid beyond the boundaries of the study. In particular, the findings may or may not be generalizable for other nations around the globe.

VII. CONCLUSION

The paper set forth to answer a dilemma that is faced by IS researchers and practitioners in deciding on web-surveys as an appropriate survey instrument representative of their population of interest. Using the population of Louisiana as a test bed, we find that the efficacy and appropriateness of web-surveys rests primarily on the population's ability to access the underlying technology. This study showed that web-surveys, although gaining momentum, are still short of representing the entire population and limited to those with higher income and/or education. However, the evidence that race and age does not significantly vary among telephone and web-survey respondents adds a positive note to our finding that the digital divide in narrowing.

In revisiting the demographic factors that are salient to the digital divide, the results show mixed inferences about the digital divide and therefore, the effectiveness of web-based surveys as being representative of the population as a whole. Race and age demographics that were once significantly attributed to the digital divide lost their significance, reducing disparities among respondents in these categories. However, income and education remain significant and conspicuous as discerning and distinctive demographic factors among web-based and telephone-based survey respondents, making it hard to infer that web-based survey respondents are representative of the population.

This survey elicits many other interesting research issues. Questions arise about the harmonizing of human factors and technology to improve information and data quality, thus reducing input errors. Another question comes from the turbulent technological flux in which ISP's and email providers show a high attrition rate. How consistent are e-mail lists or how fast do e-mail addresses change through time?

With surveys dominating the way first hand information is obtained about target populations, effectiveness of survey modes remains an important issue for the present and the future. With the Internet having democratized the survey taking process, web surveys are proliferating rapidly [Couper, 2000]. Perhaps this is the juncture where we need to examine and match the effectiveness and access of the web surveys to the target population being sought, finding issues, prospects, and limitations about the research methodology. This article is an attempt to reconcile the disparities by eliciting the demographic factors that impede the "democratization" of the web as a survey medium.

Editor's Note: This article was fully refereed. It was received on April 24, 2002. It was with the authors approximately 3 months for two revisions. It was published on October 27, 2002.

REFERENCES

EDITOR'S NOTE: The following reference list contains the address of World Wide Web pages. Readers who have the ability to access the Web directly from their computer or are reading the paper on the Web, can gain direct access to these references. Readers are warned, however, that

1. these links existed as of the date of publication but are not guaranteed to be working thereafter.
2. the contents of Web pages may change over time. Where version information is provided in the References, different versions may not contain the information or the conclusions referenced.
3. the authors of the Web pages, not CAIS, are responsible for the accuracy of their content.
4. the author of this article, not CAIS, is responsible for the accuracy of the URL and version information.

Alreck, P.L. and Settle, R.B. (1995) *The Survey Research Handbook*, 2nd edition. Chicago: Irwin.
Arbuckle, J. R. (1999). *Amos Structural Equation Modeling Software.* Smallwaters Corporation: Chicago, IL.

Ary, D., Jacobs, L.C. and Razavieh, A. (1996). *Introduction to Research in Education*. Fort Worth, Harcourt Brace.
Bachman, D.P. (2001). E-mail and snail mail face off in rematch. *Marketing Research*. Chicago. Winter 1999/Spring 2000

Bainbridge, William Sims. (1999). Cyberspace: Sociology's Natural Domain. *Contemporary Sociology-A Journal of Reviews* 28: 664-667.
Blair, Johnny, and Young I. Chun. (1992). Quality of Data from Converted Refusers in Telephone Surveys. Paper presented at the *American Association for Public Opinion Research Meeting*, St. Petersburg, FL.

Bohland, J., Papadakis, M., and Worrall, R. (2000). *Creating the CyberSouth*. A report commissioned by the Southern Growth Policies Board. September.

Bowen, Bruce. (1996) *An Introduction to Survey Research, Polling, and Data Analysis.* California: SAGE Publications.

Box, G. E. P. (1954). Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems. *Annals of Statistics*, 25: 290-302.

Brewer, Laura C. (2000). The Politics and Practice of Doing Online Surveys: A Case Study of Higher Education Students in Web-Based Courses. *American Sociological Association 2000.*

Chen, Peter and SM Hinton. 1999. Realtime Interviewing Using the World Wide Web. *Sociological Research Online* 4.

Cobanoglu, C., Warde, B., and Moreo, P.J. 2001. A Comparison of Mail, Fax, and Web-Survey Methods. *International Journal of Market Research*. 43, 4: 441-452.

Conhaim, W. W. (2001). Update on the digital divide in the United States. *Link - up*; Medford; Jul/Aug 2001.

Coomber, R. (1997). 'Using the Internet for Survey Research' *Sociological Research Online*. 2, 2. http://www.socresonline.org.uk/socresonline/2/2/2.html (current Sept. 8, 2002).

Couper, M and Groves, R. (1992). The Role of the Interviewer in Survey Participation. *Survey Methodology*, 18, 2: 263-271.

Couper, M. P. (2000). Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly*. Vol. 64, 4: 464-494.

Couper, M. P. (2001). The Promises and Perils of Web Surveys. *ASC Conference: The Challenge of the Internet*. 11-12 May, 2001.

Couper, M. P., M. Traugott, and M. Lamias. (1999). Effective Survey Administration on the Web. Paper presented at the *Midwest Association for Public Opinion Research*, November.

Crawford, Scott, Mick P. Couper, and Mark J. Lamias. (2001). Web Surveys: Perceptions of Burden. *Social Science Computer Review* 19:146-162.

Day, Neil Atherton, David R. Dunt, and Susan Day. (1995). Maximizing Response to Surveys in Health Program Evaluation at Minimum Cost Using Multiple Methods. *Evaluation Review* 19(4): 436-450.

De Heer, W. F., and Israels. A. Z. (1992). Nonresponse Trends in Europe. Paper presented at the *Joint Statistical Meetings of the American Statistical Association*, Boston.

De Leeuw, Edit, and Wim F. De Heer. (1999). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. Paper presented to the *International Conference on Survey Nonresponse*, Portland, Oregon.
Dillman, D. A. (1978). *Mail and Telephone Surveys; The Total Design Method*. New York: Wiley.

Dillman, D. A. (2000). *Mail and Internet Surveys: The Tailored Design Method*. New York: Wiley.

Dillman, D. A. and Bowker, D. K. (2001). *The Web Questionnaire Challenge to Survey Methodologists*. Dimensions of Internet Science. Edited by Ulf-Dietrich Reips and Michael Eichengrund, Germany: Bosnjak. Pabst Science Publishers.

Dillman, D. Tortora, A., Conradt, J., and Bowker. D. K. (1998). Influence of Plain vs. Fancy Design on Response Rates for Web Surveys. Paper presented at the *Joint Statistical Meetings of the American Statistical Association*, Dallas, August.

Direct Marketing Association. (2000). Economic Impact: US. *Direct Marketing Today*. http://www.the-dma.org (current Sept. 8, 2002).

Dommeyer, C. J., and E. Moriarty. (2000). Increasing the Response Rate to E-Mail Surveys. Paper presented at the annual meeting of the *American Association for Public Opinion Research*, Portland, OR, May.

Drury, D. and Farhoomand, A. (1997). Improving Management Information Systems Research: Question Order Effects in Surveys. *Information Systems Journal,* 21, 3.

DSS Research. (2000). *Mail vs. Telephone.* Texas: DSS Research.

Eaton, B. (1997). Internet Surveys: Does WWW Stand for `Why Waste the Work? *Quirk's Marketing Research Review*. URL: http://www.quirks.com (current Sept. 8, 2002).

Ferber, R., Sheatsley, P., Turner A., and Waksberg, J. (1980). *What is a Survey?* Subcommittee of the Section on Survey Research Methods. American Statistical Association Washington, D.C.

Foster, G. (1994). 'Fishing the Net for Research Data' *British Journal of Educational Technology*. 25, 2: 91-97.

Frey, James H. (1983), *Survey Research by Telephone,* Sage Publications, Inc.

Gay, L.R. and Diehl, P.L. (1992). *Research Methods for Business and Management*. New York: Macmillan.

Gonzalez, M. E. (1990). Statistical Policy Working Paper 19 - *Computer Assisted Survey Information Collection*. Federal Committee on Statistical Methodology. http://www.fcsm.gov/working-papers/wp19.html (current Sept. 8, 2002).

Groves, R. M., and M. P. Couper (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.

Hoffman, D. L. and T. P. Novak. (1996, February). *A New Marketing Paradigm for Electronic Commerce*. http://www2000.ogsm.vanderbilt.edu/novak/new.marketing.paradigm.html (current Sept. 8, 2002).

Hoffman, D. L., T. P. Novak, and P. Chatterjee (1995). Commercial scenarios for the Web: opportunities and challenges. *Journal of Computer-Mediated Communication*, 1(3). http://shum.huji.ac.il/jcmc/vol1/issue3/vol1no3.html (current Sept. 8, 2002).

John J. Heldrich Center for Workforce Development (1999). *Working Hard but Staying Poor in the Shadow of the Boom: New Survey of America's Working Poor Provided In-Depth Economic Portrait*. Edward J. Bloustein School of Planning and Public Policy at Rutgers and the Center for Survey Research and Analysis at the University of Connecticut. July 22.

Jöreskog, K. G. and Sörbom, D., Du Toit, S. and Du Toit, M. (1999*). LISREL 8: New Statistical Features*. Chicago: Scientific Software International.

Keeter, S. et al. (2000). Consequences of Reducing Nonresponse in a National Telephone Survey; *Public Opinion Quarterly*, Chicago; Summer 2000; Vol. 64, 2: 125-148.

Kehoe, Colleen M. and Pitkow, James E. (1996). *Surveying the Territory*: GVU's Five WWW User Surveys, Graphics, Visualization, and Usability Center, Georgia Institute of Technology, Atlanta, GA, Technical Report GIT-GVU-97-14, 1996.

King, R. (1998). *To Have and Have not—The Digital Divide Between Rich and Poor Run Deep—Particularly in America's Cities*. August 1998. http://www.techweb.com/se/directlink.cgi?TLC19980801S0035 (current Sept. 8, 2002).

Levin, Irwin P. (1999). *Relating Statistics and Experimental Design*. Thousand Oaks, CA: Sage Publications.

Liefeld, J. P. (1988). Response Effects in Computer-Administered Questioning. *Journal of Marketing Research*. Vol. 25 (4), 405-409.

Link, M. (1999). Call screening: Is it really a problem for survey research?; Public Opinion Quarterly**,** Chicago; Winter 1999; Vol. 63, (4), pp. 577-589.

Markham, A. (1998). *Life Online: Researching Real Experiences in Virtual Space*. Walnut Creek, CA: AltaMira Press.

Mehta, R. and Sivadas, E. (1995). 'Comparing Response Rates and Response Content in Mail versus Electronic Mail Surveys'. *Journal of the Market Research Society*. 37, 4, 429-439.

Mueller, R.O., *Basic Principles of Structural Equation Modeling. An Introduction to LISREL and EQS*. Springer-Verlag, 1996.

National Telecommunications and Information Administration (NTIA) (2000). *Falling through the Net: Towards Digital Inclusion 2000*. U.S. Department of Commerce.

National Telecommunications and Information Administration (NTIA) (1999). *Education: Boosting the Odds for Internet Use. Falling Though the Net: Defining the Digital Divide.* U.S. Department of Commerce.

Nederhof, Anton J. 1988. Effects of a Final telephone Reminder and Questionnaire Cover Design in Mail Surveys. *Social Science Research* 17:353-361.

Newsted, P.R., Huff, S.L., and Munro, M.C. 1998. Survey Instruments in Information Systems. *MIS Quarterly*. December. 22, 4, pp. 553-554.

Novak, T. P. and Hoffman, D. L. (1998). *Bridging the Digital Divide: The Impact of Race on Computer Access and Internet Use*, E-Labs, February 1998.
http://ecommerce.vanderbilt.edu/papers/race/science.html (current Sept. 8, 2002).

Oldendick, Robert W., and Michael W. Link. (1994). The Answering Machine Generation. *Public Opinion Quarterly*. 58(2): 264-73.

Piazza, Thomas. (1993). Meeting the Challenges of Answering Machines. *Public Opinion Quarterly*. 57(2):219-31.

Rogers, M. and Oder, N. (1999). Feds focus on "Digital divide" between Whites and Minorities. *Library Journal*; New York; August.

Rogers, M. and Oder, N. (2001). Studies: Is the Digital Divide Narrowing? *Library Journal*; New York; March.

Schumacker, R.E. and Lomax, R.G. (1996). *A Beginner's Guide to Structural Equation Modeling*. Mahwah, NJ: Erlbaum.

Shepard, Paul. (1999) "Bridging the Racial Net Gap: NAACP Works With AT&T to Create Tech Centers," *The Associated Press*, July 13, 1999.
http://www.abcnews.go.com/sections/tech/DailyNews/digitaldivide990713.html (current Dec. 8,1995).

Stevens, B (2000). Save Money with Online Analysis. *Marketing News*, Chicago; Nov 6, 2000; Vol. 34 (23). pp. 27-28.

Subcommittee on Computer-Assisted Survey Information Collection (1990). *Computer-Assisted Survey Information Collection. Washington, DC*: Office of Management and Budget

Szymanski, D. M. (2001). Modality and offering effects in sales presentations for a good versus a service; *Academy of Marketing Science Journal*, Greenvale; Spring 2001; Vol. 29, Iss. 2; 179-189.

The Survey Research Laboratory (2000). Office of Research and Sponsored Projects. Portland State University. http://www.gsr.pdx.edu/rsp/srl.html (current Sept. 8, 2002).

Triplett, Timothy, Johnny Blair, Teresa Hamilton, and Yun Chiao Kang. (1996). Initial Cooperators vs. Converted Refusers: Are There Response Behavior Differences? Paper presented at the 51st Annual Conference of the *American Association for Public Opinion Research*: Orlando, Florida.

Trochim, W. (2001). *The Research Methods Knowledge Base*, 2nd ed. Atomic Dog Publishing: Cornell University. Ithaca, NY.

United States General Accounting Office (2001). Telecommunications: Characteristics and Choice of Internet Users. *Report to the Ranking Minority Member, Subcommittee on Telecommunications, Committee on Energy and Commerce*, House of Representatives. February.

University of Kentucky Survey Research Center (2000). (http://www.rgs.uky.edu/src/) (current Sept. 8, 2002).

University of Kentucky Survey Research Center. (2001). http://survey.rgs.uky.edu/. Updated August 7th, 2001. (current Sept. 8, 2002).

Waksberg, J. (1978). Sampling methods for Random Digit Dialing. *Journal of the American Statistical Association* 73**,** 40-46.

Watt, J.H. (1997). Using the Internet for Quantitative Survey Research. *Quirk's Marketing Research Review*. July.

Weiss, M. J. (2001). Online America. *American Demographics*. March: 32-39.
Wilson, D. Taylor, A. and Chittleborough, C. (2001). The Second Computer Assisted Telephone Interview (CATI) Forum: The State of Play of CATI survey Methods in Australia. *Australian and New Zealand Journal of Public Health.* June.

Witte, James C., Lisa M. Amoroso, and Philip E. N. Howard. 2000. Research Methodology: Method and Representation in Internet-Based Survey Tools-Mobility, Community, and Cultural Identity in Survey. *Social Science Computer Review* 18:179- 195.

Zatz, D. (2000). Create Effective E-mail Surveys. *HR Magazine*. 45, 1, pp. 97-103.

ABOUT THE AUTHORS

Pratim Datta is a 2nd year Ph.D. student in the Information Systems and Decision Sciences (ISDS) department at Louisiana State University.

Ken Walsh is Assistant Professor of Information Systems at the University of New Orleans.

Dek Terrell is the Freeport-McMoran Chair of Economics and Associate Professor of Economics and the Director of the Division of Economic Development and Forecasting at Louisiana State University.

Communications of the Association for Information Systems

ISSN: 1529-3181