Journal of Official Statistics, Vol. 15, No. 1, 1999, pp. 39–56

A Comparison of Mail and E-mail for a Survey of Employees in U.S. Statistical Agencies

Mick P. Couper,[1] Johnny Blair,[2] and Timothy Triplett[2]

This article reports on the results of a study comparing e-mail and mail for a survey of employees in several government statistical agencies in the U.S. As part of a larger study of organizational climate, employees in five agencies were randomly assigned to a mail or e-mail mode of data collection. Comparable procedures were used for advance contact and followup of subjects across modes. The article describes the procedures used to implement the survey, and discusses the results of the mode experiment. Across all five agencies, higher response rates were obtained for mail (range of 68–76%) than for e-mail (range 37–63%). Data quality (item missing data) was similar across the two modes. Higher-status employees appeared more likely to respond to e-mail than to mail. Controlling for differences in the composition of the samples due to nonresponse, e-mail respondents appeared to be more positive in their responses to questions about climate and morale in their agencies.

Key words: Electronic mail; mode of data collection; organizational surveys.

1. Introduction

The process of converting interviewer-administered surveys from paper-based approaches to computer assisted methods (CATI and CAPI) is well underway, and almost complete in the U.S. and Europe (see Couper and Nicholls, 1998). Until recently mail surveys have remained largely unaffected by this transition. With the proliferation of electronic communications in the last several years, electronic mail (e-mail) surveys are increasingly being promoted as a fast and cost-effective alternative to traditional mail surveys (e.g., Oppermann, 1995; Swoboda, Mühlberger, Weitkunat, and Schneeweiß, 1997). Some claim that e-mail surveys may ''become the standard data collection methods in the 21st century'' (Schuldt and Totten 1994).

Proponents of e-mail surveys generally acknowledge that currently limited access to

[1] Survey Research Center, University of Michigan and Joint Program in Survey Methodology (JPSM), PO Box 1248, Ann Arbor, MI 48106-1248, U.S.A. E-mail: mcouper@isr.umich.edu.
[2] Survey Research Center (SRC), University of Maryland, Campus Drive, College Park, MD 20742.
Acknowledgments: The students in the JPSM practicum played a critical role in the design and conduct of this study. Financial support came from the National Science Foundation, through its support of the JPSM. This was supplemented by contributions from the participating agencies. Agency involvement was coordinated through an advisory council led by Richard Schuchardt of NASS (who served as the liaison to the practicum class), Lynda Carlson and Emilda Rivers of EIA, and Cynthia Clark and Nancy Bates of the Census Bureau. We also thank Skip Camp who, as SRC study manager, invested considerable time and effort in this project. Finally, we appreciate the useful suggestions we received on an earlier version of this article from Don Dillman, Roger Tourangeau, William L. Nicholls II, and an anonymous reviewer. An earlier version of this article was presented at Statistics Canada's *Symposium '97; New Directions in Censuses and Surveys*, Ottawa, November, 1997.

e-mail by the general public makes this method unsuitable for general population surveys. Thus, e-mail surveys have largely been tested in relatively closed populations (such as employees in a single organization) where the penetration of e-mail technology is high. In such cases, where coverage is less of a concern, e-mail is a potentially attractive alternative to traditional mail surveys. Whether e-mail surveys are a transition technology leading to the development of Web-based surveys, or will remain a useful tool in the researcher's toolbox, we believe the trend toward electronic replacements for paper-based methods will continue. It is thus important to explore the utility and effectiveness of this alternative method of data collection. In this article we report on the results of a mode comparison of e-mail versus mail in a study of organizational climate in federal statistical agencies in the United States.

2. Background

There is a small but growing literature comparing e-mail to mail data collection. A key focus of these comparisons has been on the relative response rates achieved using these alternative approaches. A second area of interest is related to efficiency gains from e-mail, both in terms of reduced costs and in terms of timeliness or speed of turnaround. Several studies have also examined differences in the quality of the data obtained by the two methods, whether in terms of item missing data rates or in terms of distributions of substantive responses. Coverage issues have received little attention, the assumption being that access to e-mail is a prerequisite for inclusion in the study.

The few mode experiments that have been conducted vary on so many different dimensions that it is difficult to make direct comparisons between them. It is also difficult to extract definitive trends from this small body of literature. While it is unequivocal that e-mail surveys are faster than mail surveys, and cheaper in terms of data collection costs, the findings on response rates and data quality are more mixed. We summarize the response rate findings for each of these studies briefly in Table 1.

First, we note that only three of the studies (Kiesler and Sproull 1986; Tse, Tse, Yin, Ting, Yi, Yee, Hong 1995; Tse 1996; Schaefer and Dillman 1998) randomly assigned subjects to mode. In the other cases, those with known e-mail addresses were assigned to e-mail while others were assigned to mail. Second, only two of the studies (Parker 1992; Schaefer and Dillman 1998) found higher response rates for e-mail than for mail. We also note the large variation in response rates across these studies, from a low of 18% to a high of 75% for mail, and a low of 6% to a high of 67% for e-mail. These dramatic differences between studies likely reflect many design differences, from the population surveyed, content and length of the instrument to methods of delivery and number of contacts. Clearly, more studies are needed before definitive conclusions are drawn about response rate differences between mail and e-mail.

While several of the authors have speculated on the reasons for the differential non-response in their studies, there has been no systematic attempt to explore the mechanisms underlying these differences. We believe it is important to distinguish between access failure and respondent resistance (to use Sosdian and Sharp's (1980) terminology). In other words, are low e-mail response rates in some studies due largely to delivery problems or to sample person unwillingness to complete an e-mail survey? If the former is

Table 1. Prior studies comparing mail and e-mail

Study	Response Rates		Population	Sample Sizes and Design
	Mail	E-Mail		
Kiesler & Sproull (1986)	75%	67%	Active computer mail users at Carnegie-Mellon U	75 each; telephone call prior to survey to solicit participation; 18-item instrument
Parker (1992)	38%	68%	140 expatriate AT&T employees	100 e-mail for all with e-mail addresses; mail for 40 with no e-mail address; 32-item instrument
Schuldt & Totten (1994)	57%	19%	MIS and Marketing faculty in membership directories	200 mail; 343 e-mail with known addresses (218 of which had usable addresses); 2-page instrument
Kittleson (1995)	77%	28%	Subscribers to e-mail directory of health educators	All 153 sent separate 1-item mail and e-mail instruments on different topics on same day
Tse et al. (1995)	27%	6%	Admin. & teaching staff at Chinese U of Hong Kong with listed e-mail addresses	200 each; 34-item instrument; no followup
Tse (1996)	52%	7%	Same as above	200 mail, 300 e-mail; 1 reminder; 14-item instrument
Mehta & Sivadas (1995)	46%	40%	Active users of BBS newsgroups	60 e-mail, 192 mail with known mail addresses; unsolicited; 5-page (300-line) instrument
Bachman, Elfrink & Vazzana (1996)	66%	53%	Business school deans and chairpersons	244 e-mail for all with e-mail addresses; 244 mail for sample of those with no e-mail address; 2-page instrument
Comley (1997)	18%	9%	3,700 names and addresses purchased from Internet magazine	1,221 e-mail for all with e-mail addresses; mail for balance; 12-page instrument
Schaefer & Dillman (1998)	58%	58%	Washington State University faculty	226 each; 46-item instrument

the case, improvements in e-mail technology and coverage may reduce the gap between mail and e-mail. On the other hand, if there is differential respondent resistance by mode, concerns about nonresponse error may be greater. If there is greater unwillingness to do an e-mail survey than an equivalent mail survey, is this due to technical difficulties in reading, completing and returning the instrument, or are there differential concerns about confidentiality?

In terms of data quality, the results are similarly mixed. Kiesler and Sproull (1986) found that 22% of mail respondents failed to complete or spoiled one or more items, compared to 10% of e-mail respondents. Schaefer and Dillman (1998) found that e-mail was more complete than mail: 69.4% of e-mail respondents completed at least 95% of the items, while only 56.6% of mail respondents did so.

However, Bachman, Elfrink, and Vazzana (1996) found a higher item missing data rate in e-mail (3.1%) than in mail (0.7%). Tse et al. (1995) report slightly higher, but non-significant ($p > 0.05$) rates of item missing data for e-mail (mean of 3.0 items missing out of 36 items) than for mail (mean of 1.15 items missing). Comley (1997) also found non-significant ($p > 0.1$) rates of item omission of 1.2% for e-mail and 0.4% for mail. Finally, Mehta and Sivadas (1995) found no differences in the rates of item missing data across modes.

Only two studies have examined responses to open-ended questions. Schaeffer and Dillman (1998) reported obtaining longer responses to open-ended questions in e-mail (an average of 3.9 lines of text) than in mail (average of 1.7 lines of text). Bachman, Elfrink, and Vazzana (1996) found that 22% of e-mail respondents and 5% of mail respondents provided an answer to the single open-ended question.

The mixed results on these data quality measures again suggest we need to look deeper for possible reasons for these differences. If the instruments were designed to be similar in format, why should we expect differences in missing data rates? Are there technical differences in terms of entry methods that lead to different item completion rates or responses to open-ended questions? Does the scrollable nature of e-mail versus the paging format of mail lead to differences in inadvertent skipping of items? Or do differential concerns about confidentiality lead to lower completion in one mode versus the other? Without further details on each of these studies we cannot answer these questions, but the results suggest it is important to examine possible causes of such differences, whether they be technical issues, design or format differences, or motivational effects associated with mode. In summary, then, there is still much we need to learn about how (and why) e-mail surveys may differ from mail surveys, particularly in terms of nonresponse and data quality. The present study was not designed to answer all these questions, but attempts to add to the research cited above.

In addition to the explicit mode comparisons discussed thus far, there are several recent reports of e-mail only surveys that may provide insight into possible effects of e-mail on survey errors. For example, Oppermann (1995) conducted an e-mail survey of members of the American Association of Geographers with listed e-mail addresses. He selected 665 persons, which yielded 500 usable e-mail addresses. He obtained a response rate of 49%, using two rounds of e-mail messages.

Swoboda et al. (1997) reported on the results of a 9-item e-mail survey sent to addresses obtained from various electronic newsgroups in Munich, Germany. Using a

single-mailing design, with no advance notice or follow-up messages, they sent a total of 8,859 e-mail messages, 5.3% of which were returned as undelivered. They reported a response rate of 20.4%. In addition, to test undeliverable addresses, Swoboda et al. (1997) sent e-mail messages to 118 fictitious addresses. After 2 weeks, 91 had been returned as undeliverable, leaving 27 unaccounted for. This suggests that the count of non-deliverable messages may underestimate technical difficulties in transmitting the instruments.

Smith (1997) sent a 66-item e-mail survey to professional ''webmasters.'' Half the sample of 300 received a direct e-mail survey scripted in HTML, and the other half received a request for participation, following which they were sent the survey. Only 11 of the 150 direct e-mail sample returned the questionnaire, compared to 42 in the other group who requested a survey (of whom 20 actually completed the survey). Smith also reported experiencing several problems with delivery failures and browser incompatibility.

The varying results across these studies suggest that issues of design and implementation may be particularly important for e-mail surveys. Furthermore, given the likely proliferation of e-mail surveys, it is important that additional mode comparison studies be conducted under varying conditions. Several of the above studies (e.g., Comley 1997; Smith 1997; Swoboda et al. 1997) also suggest that many technical issues related to e-mail surveys remained unresolved, especially when the survey is not conducted within a closed system (a single organization in which all persons are using the same e-mail system).

With these issues in mind, we embedded a mode experiment in an organizational climate survey of employees within several statistical agencies in the U.S. The mode experiment was designed to investigate the feasibility of e-mail and evaluate the relative quality of the two methods (mail and e-mail) for surveys of federal employees. In the balance of this article we describe the steps taken to implement the survey, and discuss the results of the mode comparison.

3. Design and Methods

The mode experiment was part of a larger study conducted on behalf of a consortium of federal statistical agencies in the U.S. The survey was conducted during the first quarter of 1997. The study was designed and conducted by graduate students in the Joint Program in Survey Methodology (JPSM), with the assistance of the Survey Research Center (SRC) at the University of Maryland. The overall objectives of the study were to develop and test an organizational climate survey suitable for implementation within federal statistical agencies (see Carlson and Rivers 1997).

The survey content was developed through several iterations of testing, including two focus groups, several cognitive interviews and conventional pretests. The final instrument consisted of 81 Likert-type attitude items (5-point scales) and 10 background items. Nine agencies participated in the larger climate study. The study was restricted to all permanent employees at these agencies. This included part-time workers, but excluded temporary employees such as coders and interviewers, as well as contract workers. All eligible employees in the participating agencies were included in the study.

Table 2. Sample sizes for mode comparison

Agency	Mail	E-mail	Total
A	2,699	2,969	5,668
B	790	396	1,186
C	266	265	531
D	216	221	437
E	216	215	431
Overall	4,187	4,066	8,253

We were assured that all employees at the nine agencies had access to electronic mail, and we were provided with electronic data files containing employee names, office addresses, telephone numbers and e-mail addresses. Given the logistical issues of launching two surveys in each of nine different agencies, we decided to restrict the experimental mode comparison to the five largest agencies. The remaining agencies were given a choice of a single mode; two chose mail and the other two chose e-mail. Table 2 lists the number of employees in each of the five agencies assigned to each mode of administration.

3.1. Development of the e-mail instrument

Because of cost and time constraints we decided against developing our own e-mail survey software and instead examined several commercial products for conducting e-mail surveys. An initial review of technical specifications led to an elimination of all but two products.

There are two basic approaches to developing an e-mail survey. One approach produces an executable file that is downloaded by the respondent and run on his/her machine. The advantage of this approach is that a graphical user interface (GUI) can be employed to facilitate completion of the survey. The executable file can make full use of computer assistance to do completeness, range, consistency and other checks, and permits branching, fills, and other features common to computer assisted self-interviewing (CASI) (see Couper and Nicholls 1998). However, the disadvantages of this approach are that it requires all users to have compatible systems, and requires users to be comfortable with downloading and running executable files.

The alternative approach is to transmit the questionnaire as a text (ASCII) file in the body of the e-mail message. The respondent then simply scrolls through the message responding to items, and returns the message to sender when completed. This approach is largely system-independent (the message can be read with any software/hardware combination) and it is the closest equivalent to a traditional mail survey, with all the benefits and drawbacks such a design brings (see Dillman 1978, 1991). It appears that most of the e-mail studies reported above adopted this latter approach. These two approaches represent different philosophies about what an e-mail survey should be: a CASI survey or an electronic analog of a mail survey.

There is a third approach which is gaining in popularity. This approach uses e-mail to invite respondents to a secure Internet or Web site where they complete the survey (e.g., Batagelj and Vehovar 1998; Clayton and Werking 1998). The Web survey can be designed either as an interactive CASI instrument or as a scrollable mail survey

equivalent. We rejected the Web approach for two reasons. First, we were told that e-mail access was virtually universal in participating agencies, while access to the Web was much less common. Second, concerns were expressed that employees in some agencies were either being monitored in their Web usage, or prevented from gaining access to the Web, because of perceived threats to productivity.

We investigated the feasibility of both e-mail approaches. We first explored a GUI-based system that produces an executable file that runs on DOS or Windows systems. While we could design an elegant instrument, this meant that each user's operating system had to be known in advance. Furthermore, we discovered that the size of the outgoing e-mail file approached 1 Mb per sample person, which was unacceptable, both in terms of the volume of Internet traffic this would create (over 4 Gb for outgoing messages alone), and because of likely agency restrictions on the size of incoming files. We were also unable to test the auto-reply feature in-house, nor did it work for any of the seven pretest subjects (technical contacts at the participating agencies). The vendor's initial solution to this problem was to have respondents change their Windows configuration (.ini) file; this would have been done by e-mailing all sample persons an executable file to automatically update their system configuration prior to receiving the survey instrument. For obvious reasons, this was deemed unacceptable. At this point, three days before the scheduled start of data collection, we were forced to seek an alternative solution.

We switched to another commercially available system that used generic text (ASCII) files. The survey instrument was easy to create and we anticipated few system incompatibilities using this product. One problem we encountered was that the system did not attach unique identifiers to individual messages – a common authentication marker was used for all cases in the survey. This meant that we had to rely on e-mail addresses to match returns back to the sample frame, which proved to be quite difficult in practice.

3.2. Administration of the survey

The mail survey materials were printed in booklet form, on $8\frac{1}{2}$ by 11 inch paper. The questionnaire was 12 pages long, including a cover with the JPSM logo and title of the survey. An ID number was placed on the back of each questionnaire. A cover letter signed by the director of JPSM and a reply-paid envelope were included in the packet. The envelopes were individually addressed and delivered to each agency for distribution using internal mail. Returns were mailed directly to SRC for check-in and data entry.

Closely comparable procedures were used for the e-mail version. However, whereas the items in the paper version were grouped into 14 sections numbered separately, the e-mail software required all 94 items and sub-items to be numbered consecutively. Whereas the response options in the paper version were arranged horizontally and to the right of each item, in the e-mail version they were arranged vertically below each item. The closed-ended questions were answered by placing an X (or any character) inside a set of brackets [ ] alongside the option. Open-ended questions were answered by typing within the brackets. This approach closely resembles that used by Schaefer and Dillman (1998). A message from the director of JPSM accompanied each instrument. The e-mail messages were sent from SRC, with the return address being *agency*@cati.umd.edu (so each agency's returns came to a different mail queue).

Both e-mail and mail questionnaires were delivered to sample persons on approximately the same day. Several days before the delivery, agency staff were informed of the upcoming survey and encouraged to participate. This took the form of agency-wide broadcast e-mail messages, as well as notices posted on bulletin boards around the agencies.

Approximately one week after the initial mailing, a reminder postcard or e-mail message was sent to all sample persons. Two weeks after the reminder, a second mailing or e-mail message containing a replacement questionnaire was sent to all nonrespondents. Finally, telephone reminder calls were attempted for all remaining nonrespondents about 6 weeks after the initial mailing. No attempt at refusal conversion was made during the telephone calls, but replacements questionnaires were offered, and reasons for nonresponse (when provided) were recorded. The data from these reminder calls are used to explore mode differences.

In brief, we attempted to duplicate as best we could the mail procedures for the e-mail mode within the limitations imposed by the differences in media. The questionnaires were comparable in content and differed only slightly in form, as noted above. The field procedures were also designed to be the same in prenotification, personalization, timing of distribution and reminders, use of replacement questionnaires and follow-up.

4. Results

We have a variety of data sources to evaluate the mode comparison. These include a tracking database in which all transactions (outgoing and incoming mail and e-mail) were logged, a small debriefing study of respondents, reminder calls to nonrespondents, and the substantive responses to the survey itself. We use these to explore a variety of possible differences across the modes, particularly in terms of response rates and data quality.

4.1. Overall response rates

First, we examine the response rates by agency and mode. These are presented in Table 3. For each of the five agencies, e-mail produced a significantly ($p < 0.01$) lower response rate than mail. The largest differences in response rate by mode are for agencies A, D, and E. There are several possible explanations. For Agency A and Agency E, the e-mail addresses were constructed from lists of employee names (following agency conventions such as *last.middle.first@agency*). All other agencies provided e-mail addresses for their employees. This suggests that the creation of e-mail addresses was automatic, whether or not an employee actually used e-mail. Such a listing may thus include a large

Table 3. Response rates by agency and mode (in percent)

Agency	Mail	E-mail	Difference
A	68.0	36.7	31.3
B	76.1	62.6	13.5
C	74.4	60.0	14.4
D	75.5	52.9	22.6
E	76.4	54.9	21.5
Overall	70.7	42.6	28.1

number of employees who do not use e-mail and hence never received the questionnaire. But because these employees officially have e-mail addresses registered at the agency, our messages were not returned as undelivered.

Another source of the difference may be technical problems related to the configuration of different e-mail systems across agencies. For example, Lotus CC:Mail automatically converts e-mail messages over a certain size (e.g., 20 Kb) into attachments. Both Agency A and Agency D used CC:Mail, and this may be a possible explanation for their low e-mail response rates. However, Agency B, which had the highest e-mail response rate, also used CC:Mail. We received several reports from employees at agencies A and D that they received attachments, and did not know what to do with them. Subsequent investigation suggests that this does not appear to have been a problem in Agency B, and some users at Agency A received the survey as intended (in the body of the message rather than as an attachment). However, the attachment problem appeared widespread at both Agency A and Agency D. It thus appears that there was variation between subunits of these two agencies, using different e-mail servers. As soon as we learned of this, we sent an additional e-mail message to all sample cases in these two agencies with updated instructions on how to deal with attachments. Similar problems were not experienced at Agency C (using Novell GroupWise) or Agency E (using GroupWise or WPMail).

Further evidence for technical problems related to the size of the e-mail survey instrument (23 Kb) can be found in the response rates to supplement instruments. All agencies were offered the opportunity to include a set of agency-specific supplement questions; only two agencies (A and D) availed themselves of this opportunity. For the mail survey, the supplements took the form of a single-sheet insert printed on color paper. For e-mail, the supplement questions were sent in a separate e-mail message at the same time as the main survey, with sample persons being told they were receiving two different questionnaires to complete. The main and supplement response rates for these two agencies are presented in Table 4. It can be seen that if response rate was defined as **any** completed questionnaire (main or supplement), the overall response rate for Agency A would increase by 19% (to 56%), while that for Agency D would increase by 16% (to 69%), whereas the mail response rates would remain unchanged. These new rates are close to those for e-mail in the two agencies (B and C) which did not experience technical difficulties receiving the e-mail instrument. (An alternative explanation is that, receiving the main and supplement instruments as two separate e-mail surveys, respondents thought they were duplicates, and responded to only one.)

However, even taking these supplement return rates into account, we still find consistently lower response rates for e-mail relative to mail across all agencies. Thus, technical problems associated with the receipt of e-mail do not appear to be the only reason for

Table 4. Main and supplement response rates by agency and mode (in percent)

Agency	Mail			E-mail		
	Main plus supplement	Main only	Supplement only	Main plus supplement	Main only	Supplement only
A	64.9	3.0	0.0	31.2	5.6	19.5
D	72.7	2.8	0.0	46.2	6.8	16.3

the lower e-mail response rates. Hence it is important for us to explore what other reasons there may be for the response rate differential, and what effect this may have on the quality of the data obtained.

4.2. Post-collection processing

Another source of information on the process is the tracking database of all returns, in which cases requiring special attention were noted. A potential advantage of e-mail surveys is the reduction of clerical effort involved in checking returned questionnaires. Table 5 shows the various types of clerical action that were required for those e-mail questionnaires which were returned. These types of action may also be indicators of the types of difficulty experienced by e-mail sample persons. The first column shows the percentage of returns received as an attachment to an e-mail message, while the second column denotes messages that required decoding. In both cases there is a variation across agencies, suggesting different technical approaches to handling e-mail. From the third column, we can see that about 16 percent of cases overall were completed using a word processor (WP) or text editor. Noting that there is overlap in these types of problem (all three could occur on a single return), about 21% of all e-mail returns did not make use of the reply feature as intended. Overall, about 4% of the e-mail respondents printed out the questionnaire and mailed it back (included in the above figure). Furthermore, a large number of cases required additional editing before the data could be appended to the database. The most common reasons were the X placed outside of the brackets or one of the brackets deleted. The fourth column shows that about 27% of cases required such editing, but again there is substantial variation across agencies. The final column identifies the percentage of returned e-mail surveys that required **any** clerical action before appending to the database. The high overall rate (46%) suggests a great deal of attention was required for the e-mail cases, potentially nullifying the savings in post-collection processing. In addition, the two agencies with the lowest e-mail response rates also exhibit the highest rates of clerical action among returns, again suggesting that technical difficulties experienced by sample persons could have affected the response rates.

4.3. Respondent reports of difficulties

We also conducted telephone debriefing interviews with a subset of those who returned completed questionnaires, in order to elicit their reactions to the survey. A total of 694 sample cases were selected from among the respondents, using several replicates to

Table 5. Types of clerical action required for e-mail returns (in percent)

Agency	Attachment	File coded	WP file	Needs edit	Any clerical action	(n)
A	8.2	14.2	17.7	37.9	57.2	(1,091)
B	1.7	0.4	4.3	13.8	20.9	(239)
C	23.7	1.3	23.7	0.0	23.9	(159)
D	12.0	18.5	21.3	19.7	55.6	(117)
E	20.9	0.9	20.9	0.0	20.3	(118)
Overall	9.5	9.9	16.1	27.2	46.5	(1,724)

Table 6. Method of reported e-mail return by agency, debriefing respondents (in percent)

Agency	Reply function	Text editor	Other	(*n*)
A	20.5	67.0	12.5	(88)
B	55.1	37.7	7.2	(69)
C	64.3	31.0	4.8	(42)
D	6.9	65.5	20.7	(28)
E	78.6	14.3	7.1	(29)
Overall	41.8	47.3	10.9	(256)

include both early and late returns. The sample was evenly split between modes, and Agency A was undersampled because of its relatively large size.

Debriefing interviews were conducted by JPSM students, ensuring that no student called a respondent from their own agency or known to them (as many of the students were employees in the agencies surveyed). A small portion of the calls were conducted by members of an undergraduate survey methods class. An overall response rate (complete/eligible) of 77% was obtained. The cooperation rate (complete/contacted) was 90% (including callbacks) or 98% (excluding callbacks). This yielded a total of 244 mail and 256 e-mail respondents who completed the debriefing survey. While we caution about generalizing from this group of cooperative respondents to the full sample, we can nonetheless gain some insight into the process of data collection from these interviews.

E-mail respondents were asked what method they used to complete and return the questionnaire. Their responses are shown in Table 6. These findings parallel those shown in Table 5, and suggest that the difficulty of replying to the survey differed across agencies. In agencies A and D, about two-thirds of respondents reported using a text editor or word processor to complete the survey, despite the fact that the survey was designed to be completed using the e-mail reply function.

We asked debriefing respondents to estimate how long they took to complete the survey. E-mail respondents reported taking significantly longer ($p < 0.01$) than mail respondents (28.3 minutes versus 22.5 minutes on average). Further, 10% of mail respondents and 20% of e-mail respondents said the survey took longer than 30 minutes to complete. While the difficulties in completing the e-mail survey reported above may have contributed to the increased time, there are no significant differences in the reported time of e-mail completion across agencies. In other words, even for those agencies which did not appear to experience technical problems, e-mail was still reported to take longer to complete than mail.

4.4. Confidentiality concerns

One of our initial concerns about e-mail was related to confidentiality. Respondents were being asked to give their candid views on their employers, and the nonanonymity of e-mail may contribute to a reluctance to complete the survey in this mode. These confidentiality concerns come about in two ways. First, all returned questionnaires contained the e-mail address of the respondent. Although mail questionnaires contained identification numbers, any intercepted e-mail questionnaires could be directly linked with the respondent, whereas in mail this could only be done by linking the ID number back to the sample

Table 7. Mean responses to two questions about access to survey responses by mode, debriefing respondents (standard errors in parentheses)

Question	Mail		E-mail	
	mean	(s.e.)	mean	(s.e.)
Supervisor access	6.15	(0.19)	6.61	(0.18)
Access by others in agency	6.32	(0.19)	6.36	(0.18)
(*n*)	(244)		(256)	

frame. The second potential threat to confidentiality comes about through the relatively weak legal protections of e-mail privacy (see Clayton and Werking, 1998; Ramos, Sedivi, and Sweet, 1998). Some (or all) agencies routinely backup all ingoing and outgoing messages onto tape, and these messages could be recovered and examined by agency personnel. In practice this is unlikely to occur, but the perception that e-mail **could** be read by others in the agency may lead to unwillingness to respond using this method.

We asked debriefing respondents how easy they thought it would be for (a) their supervisors and (b) anyone else in their agency to get access to their (mail or e-mail) responses. Using a 10-point scale where 1 means very easy to get access and 10 is very difficult (thus a high score means low confidentiality concern), the average responses by mode are presented in Table 7. Neither set of response differs significantly ($t = 3.66$, $p > 0.05$) by mode. Thus, among the debriefing respondents at least, there does not appear to be greater concern about the confidentiality of their e-mail responses.

Another source of information on potential confidentiality concerns came from reminder calls to nonrespondents. Toward the end of the study, we attempted to contact all remaining nonrespondents to encourage participation, and this information may provide further insight into the reasons for nonreturn of some questionnaires. However, given the high level of nonresponse, time and funds did not permit a concerted effort to contact every nonrespondent. A one-call rule was implemented to ensure that at least one attempt was made for every case. The outcomes of the reminder call attempts are presented in Table 8. The ''other'' category includes wrong numbers, sample persons who had left the agency, and so on.

In Table 9 we present the results of the call for those persons with whom we made contact. First e-mail contacts were more likely to say they did not plan to return the questionnaire (37% versus 23%). Among these, almost half (46% of the refusers and 17% of all those contacted) claimed that they did not receive the questionnaire by

Table 8. Outcome of reminder calls by mode

Outcome	Mail		E-mail	
	percent	(*n*)	percent	(*n*)
Talked with sample person	46.7	(433)	43.2	(964)
Call back	24.7	(229)	30.6	(683)
Left message	12.6	(117)	17.0	(377)
Other	16.0	(148)	9.2	(207)
Total	100.0	(927)	100.0	(2,231)

Table 9. Contacted persons' responses to reminder call, by mode (in percent)

	Mail	E-mail
Response to reminder call:		
Already returned	24.0	22.7
Will return	53.1	39.9
Refused, other	22.9	37.3
Total	100.0	100.0
Among those who refused, reasons given for nonreturn:		
Did not receive	3.9	8.0
Lost, deleted	3.7	9.1
Could not edit	0.0	2.7
No time	4.2	4.4
Confidentiality	1.4	1.2
Other, no reason	9.7	12.0
Total refused, other	22.9	37.3

e-mail or had lost or deleted the message, but did not want to be sent another. This appears to be less of a problem with the mail questionnaire, again suggesting that delivery of an e-mail instrument may be more problematic than that of a mail instrument. Second, 3% of the e-mail contacts (or 7% of those who said they would not respond) reported difficulties editing the instrument as a reason for nonreturn; as one might expect, no mail contacts reported this reason. Third, the most interesting finding from this table is that the proportion of contacts mentioning confidentiality as a reason for nonreturn does not differ by mode. In fact, 6% of the mail contacts who did not intend to respond mentioned confidentiality concerns, compared to 3% of e-mail contacts who did not intend to respond. These findings again suggest that technical difficulties, rather than confidentiality concerns, may account in large part for the lower e-mail response rates obtained.

4.5. Data quality

A final source of data for evaluating the mode experiment comes from the completed questionnaires themselves, which allowed us to examine the relative quality of the data obtained by each method. Given random assignment to mode, we would expect the distributions of key variables and the levels of item missing data to be similar. Table 10 contains item missing data rates by mode, for the 81 attitude items and eight of the background items. The response options for the attitude items were presented in slightly different format across the two modes: horizontally to the right of each question in mail, and vertically below the question in e-mail. The format of the eight background items we examined was the same across modes.

We see from Table 10 that the overall rates of missing data are low for both modes (on average less than one of the 81 attitude items are missing per respondent). There are no significant ($t = 0.16$, $p > 0.05$) mode differences in item missing data on the attitude items. However, mail had a significantly higher ($t = 2.94$, $p < 0.01$) rate of missing data on the background measures. Inspection of the individual items revealed that all eight background items had higher rates of missing data for mail than for e-mail, with these differences being significant ($p < 0.05$) for six of the eight items. One possible

Table 10. Mean item missing data rates by mode

	Mail		E-mail	
	mean	(s.e.)	mean	(s.e.)
81 attitude items	0.63	(0.054)	0.64	(0.059)
8 background items	0.24	(0.018)	0.16	(0.019)
(*n*)	(2,969)		(1,724)	

explanation may be the differential effect of nonresponse – those who did make the effort to complete the e-mail questionnaire may have been more motivated to provide complete information.

We also examined the distributions of both demographic and substantive variables across modes. As we only have these data for respondents (not all employees), we can only examine the relative effects of nonresponse. We assume that those who use computers more routinely in their work (e.g., those in higher grades) would be more likely to return the e-mail questionnaire. We find significant differences ($p < 0.01$ in each case) in the distributions of respondents in terms of grade level, managerial and supervisory status. These results are presented in Table 11. Overall, the direction of the effect is as expected: higher status employees appear to be over-represented in the e-mail survey. These differences are striking, and suggest differential access to, or use of, e-mail among employees of different status. We also find significant differences by race and sex ($p < 0.01$), with non-minorities and males being more likely to respond by e-mail than by mail. These results are also presented in Table 11. We did not find significant differences by mode with respect to years of service in the agency or in the federal government.

Table 11. Distributions of respondent demographic characteristics, by mode (in percent)

	Mail	E-mail
Grade level:		
Grades 1–4	20.2	2.8
Grades 5–11	32.6	25.5
Grades 12–13	34.9	53.1
Grades 14+	12.4	18.7
Managerial status:		
Yes	14.7	22.6
No	85.3	77.4
Supervisory status:		
Yes	23.5	31.2
No	76.5	68.8
Sex:		
Male	40.0	47.6
Female	60.0	52.4
Race:		
White	77.3	82.8
Black	17.3	11.0
Other	5.4	6.2

In terms of substantive differences on the organizational climate items themselves, we hypothesized that nonrespondents would hold more negative attitudes toward their agency than respondents. Thus, with the higher nonresponse rates for e-mail, we expected more positive attitudes among those who did respond, relative to mail. We compared the mean scores between the two groups on each of the 13 organizational climate subscales, as well as the overall mean climate score. We found significant differences by mode on five of the 13 subscales, with mail having a higher (more positive) mean score on three of the five, and e-mail on the remaining two. Overall, mail respondents were more positive on seven of the 13 subscales, and e-mail on six.

However, the differences in the sample compositions may affect the comparative attitude measures. We thus ran a series of regression analyses with the total climate scores and each of the subscale scores in turn as the dependent variable, and examined the effect of mode of administration, controlling for the background variables in Table 11. Controlling for such compositional differences in the two samples, we found e-mail respondents to be significantly more positive than mail respondents on the total climate score, as well as on nine of the 13 subscales. There were no significance differences on the remaining four subscales, but in each case the mean was higher for e-mail than mail.

In summary, when comparing marginal distributions of the two modes we found that they appeared similar, but when controlling for differences in characteristics of the two samples (due to nonresponse) we found significant differences in attitudes toward agencies and organizational climate. This suggests that those who hold more positive attitudes toward the agency are more likely to respond to a survey about organizational climate. This effect may be more prominent in the e-mail mode because of the larger nonresponse rate.

4.6. Costs

Finally, while we do not have a detailed cost breakdown for the two modes, we offer a few observations on the cost implications from our study. The task of evaluating and testing e-mail software took over 150 hours of staff time, or almost 4 times what was budgeted. This suggests that the start-up costs associated with an e-mail survey may not be trivial. Printing and postage costs were approximately 1.60 USD per sample case for mail and 0 USD for e-mail. Keying the completed mail questionnaires cost about 1.81 USD per completed case, whereas managing the e-mail sample (including the clerical actions mentioned earlier) cost about 1.74 USD per completed case. In addition, the SRC staff handled over 900 incoming toll-free calls regarding the survey, most of which were technical questions about e-mail. Given the relatively large start-up costs, technical problems associated with implementing the e-mail survey, the high level of clerical action required, and the low response rate relative to mail, in this study we did not experience the cost savings expected from e-mail.

5. Conclusions

While electronic mail potentially offers savings in time and money over mail for organizational surveys, our results suggest that such benefits will not always be realized, for similar levels of data quality and response rates. Despite pretesting the survey in each

of the agencies in the study, we were unable to anticipate the problems caused by message size limitations on certain servers within some agencies. Thus, the burden placed on sample persons in terms of what actions they had to take to read the questionnaire and respond to the survey varied across hardware and/or software platforms. Our survey was longer than most others reported in the literature. However, the response rates we obtained for the mail portion were respectable, suggesting it was the mode of delivery and return that caused problems, rather than the content of the survey. It may be that length and/or the sensitive nature of the content may interact with mode in affecting nonresponse, but we cannot test this with our data.

It is also clear that simply because every sample person has an e-mail address, does not mean that they will receive the survey or be able to respond in the manner intended. Finding a way to validate all addresses prior to mailing (as Schaefer and Dillman did) may help reduce the proportion of dormant e-mail addresses (assigned but never accessed). Verifying receipt of e-mail messages is another approach to identifying such cases, as long as this does not produce additional technical problems or unduly increase the volume of message traffic. E-mail clearly offers a lot of promise, but the technical limitations need to be overcome before e-mail can be routinely used for surveys of large and diverse populations across multiple organizations.

Despite the response rate differences we observed, we remain optimistic about the potential for e-mail as an alternative to the traditional mail survey. It is clear that a large number of persons could and did respond to an e-mail survey. The nonresponse differentials between mail and e-mail are due in large part, we believe, to technical problems affecting access, rather than to respondent unwillingness to participate in an e-mail survey. This is supported by the large variation in e-mail response rates across agencies. Our experience suggests caution in administering an e-mail survey, but many of the technical problems can be overcome with time and effort.

While our findings add to the growing body of literature on e-mail surveys, there is still much we do not know about this method of data collection. The success of an e-mail survey may depend on many factors: the quality of the e-mail addresses, technical issues related to specific e-mail software or servers, the length and content of the survey, the nature of the target population, and so on. The variation in the results obtained so far suggests caution in generalizing from the findings of any one study, and argues for a need to explore in greater detail the factors that affect nonresponse and measurement errors in e-mail surveys relative to mail surveys.

6. References

Bachman, D., Elfrink, J., and Vazzana, G. (1996). Tracking the Progress of E-mail vs. Snail-mail. Marketing Research, 8 (2), 31–35.

Batagelj, Z. and Vehovar, V. (1998). Technical and Methodological Issues in WWW Surveys. Paper presented at the annual meeting of the American Association for Public Opinion Research, St. Louis, May 1998.

Carlson, L.T. and Rivers, E.B. (1997). Origins of the Organizational Climate Survey of Federal Statistical Agencies. Proceedings of the American Statistical Association, Survey Research Methods Section, 594–599.

Clayton, R.L. and Werking, G.S. (1998). Business Surveys of the Future: The World Wide Web as a Data Collection Methodology. Chapter 27 in Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls, W.L. II, and O'Reilly, J. (eds.), Computer Assisted Survey Information Collection. New York: Wiley.

Comley, P. (1997). The Use of the Internet as a Data Collection Tool. Paper presented at the ESOMAR Annual Conference, Edinburgh, September 1997, <http://www.sga.co.uk/esomar.html>.

Couper, M.P. and Nicholls II, W.L. (1988). The History and Development of Computer Assisted Survey Information Collection. In Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls II, W.L., and O'Reilly, J. (eds.), Computer Assisted Survey Information Collection. New York: Wiley, 1–21.

Dillman, D.A. (1978). Mail and Telephone Surveys: The Total Design Method. New York: Wiley.

Dillman, D.A. (1991). The Design and Administration of Mail Surveys. Annual Review of Sociology, 17, 225–249.

Kiesler, S. and Sproull, L.S. (1986). Response Effects in the Electronic Survey. Public Opinion Quarterly, 50, 402–413.

Kittleson, M.J. (1995). An Assessment of the Response Rate via the Postal Service and E-mail. Health Values, 18 (2), 27–29.

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.

Oppermann, M. (1995). E-mail Surveys – Potentials and Pitfalls. Marketing Research, 7 (3), 29–33.

Parker, L. (1992). Collecting Data the E-mail Way. Training and Development, July, 52–54.

Ramos, M., Sedivi, B.M., and Sweet, E.M. (1998). Computerized Self-Administered Questionnaires (CSAQ). In Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls II, W.L., and O'Reilly, J. (eds.). Computer Assisted Survey Information Collection. New York: Wiley, 389–408.

Schaefer, D. and Dillman, D. (1998). Development of a Standard E-mail Methodology: Results of an Experiment. Public Opinion Quarterly, 62, 378–397.

Schuldt, B.A. and Totten, J.W. (1994). Electronic Mail vs. Mail Survey Response Rates. Marketing Research, 6 (1), 36–44.

Smith, C.B. (1997). Casting the Net: Surveying an Internet Population. Journal of Computer-Mediated Communication, 3 (1), <http://www.ascusc.org/jcmc/vol3/issue1/smith.html>.

Sosdian, C.P. and Sharp, L.M. (1980). Nonresponse in Mail Surveys: Access Failure or Respondent Resistance. Public Opinion Quarterly, 44, 396–402.

Tse, A.C.B. (1996). Comparing the Response Rate, Response Speed and Response Quality of Two Methods of Sending out Questionnaires: E-mail vs Mail. Paper presented at the International Conference on Social Science Methodology, Essex, July 1996.

Tse, A.C.B., Tse, K.C., Yin, C.H., Ting, C.B., Yi, K.W., Yee, K.P., and Hong, W.C.

(1995). Comparing Two Methods of Sending out Questionnaires: E-mail versus Mail. Journal of the Market Research Society, 37 (4), 441–446.

Swoboda, W.J., Mühlberger, N., Weitkunat, R., and Schneeweiß, S. (1997). Internet Surveys by Direct Mailing. Social Science Computer Review, 15 (3), 242–255.

Received August 1998
Revised December 1998