

A Comparison of Mail and E-Mail for a Survey of Employees in Federal Statistical Agencies

Mick P. Couper¹ (Joint Program in Survey Methodology and University of Michigan), Johnny Blair (Survey Research Center, University of Maryland) and Timothy Triplett (Survey Research Center, University of Maryland)

Abstract

This paper 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 organization climate, employees in five agencies were randomly assigned to a mail or e-mail mode of data collection. Similar procedures were used for advance contact and followup of subjects across modes. This paper describes the procedures used to implement the e-mail survey, and discusses the results of the mode experiment.

Key Words: electronic mail; mode of data collection; organizational surveys

1. Introduction

With the proliferation of electronic communications in the last several years, electronic mail (e-mail) is an increasingly attractive alternative to mail for surveys of employees in organizations with high penetration of e-mail technology. The major advantages claimed for e-mail over mail are reduced costs and quick turnaround. However, concerns have been raised about issues such as coverage, nonresponse, and measurement error effects of e-mail data collection.

With this 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 evaluate the relative quality of the two methods (mail and e-mail) for surveys of federal employees. In this paper we describe the steps taken to implement the survey, and discuss the results of the mode comparison.

2. Design and Administration of the Survey

The mode experiment was embedded in an organizational climate survey conducted on behalf of a consortium of federal statistical agencies in the U.S. The study was designed by graduate students in the Joint Program in Survey Methodology (JPSM) as part of the survey practicum. Data collection was undertaken by the students in conjunction with 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. The instrument, and the data collected, would be used to benchmark climate within and between agencies of similar composition and function. Details of the survey are reported elsewhere (see Carlson and Rivers, 1997). We focus here on the mode experiment.

The survey instrument 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 and 10 background items. Nine agencies participated in the larger climate study. The sample 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.

One of the problems of doing organizational studies is the relatively large sampling fractions required for subgroup analysis. This leads to potential contamination issues (with some employees in an

¹Survey Research Center, P.O. Box 1248, Ann Arbor, MI 48106

office getting a questionnaire and others not), as well as concerns about providing all employees with an opportunity to voice their opinions about the organization. Thus, we hoped to do a census of all employees in the participating agencies. We could not afford to do this using traditional methods of data collection, which led to the decision to consider e-mail (considered to be much cheaper than mail at the time of our decision). The cost savings would obviously be greater in the large agencies. At the same time, using both mail and e-mail would allow us to test the efficacy of the alternative method of data collection for a survey such as this.

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. Table 1 lists the number of employees in each of the five agencies assigned to each mode of administration. The imbalance for Agency B is due to the fact that the mail mode was further split between an anonymous and an identifiable group.

Table 1: 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

3. Data Collection Procedures

In this section we discuss the acquisition and evaluation of software for conducting the e-mail surveys, and describe the general data collection procedures used for both modes. 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.

We first examined Raosoft's "EZSurvey" (see <<http://www.raosoft.com>>). Advantages of this product included easy development of instruments, an auto-reply feature that facilitated the return of completed questionnaires, the ability to handle various question types, skip functions, and the use of a graphical user interface (GUI). In addition, the software is available for a fixed price, leading to economies of scale and possible use in other studies. However, EZSurvey creates a DOS or Windows executable file, which means the user's operating system needs to be known in advance. 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 (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 (using Pegasus Mail); 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 be done by e-mailing all sample persons an executable file to automatically update their system configuration. This was deemed unacceptable. Raosoft then offered to fix the program and have us do the testing. Given that this was three days prior to the start of data collection (which could not be postponed because of external constraints), we had to seek an alternative solution.

We switched to Decisive Survey, a product of Decisive Technology Corporation (see <<http://www.decisive.com>>). Decisive is a text-based system, so would work on all operating systems. However, the interface is less appealing than a GUI system, and the system does not accommodate embedded logic such as skips. The product is priced on a sliding scale depending on sample size. In addition, no unique identifiers are attached to messages; however, different surveys could be identified with unique authentication markers. This meant that we had to rely on e-mail addresses to match returns back to the frame, something which proved to be quite difficult in practice. Still, we were able to successfully test Decisive Survey with persons at each agency, so decided to proceed with the mode experiment.

The mail survey materials were printed in booklet form, on 8½ 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 in bulk for distribution using the internal mail system. Returns were mailed directly to SRC where they were processed and responses keyed.

Similar strategies were used for the e-mail version. However, whereas the items in the paper version were grouped into 14 sections, the e-mail software required all 94 items and sub-items to be numbered consecutively. 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. 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. The survey mailing was also preceded by an advance letter or message from the head of each agency informing their staff of the upcoming survey and encouraging participation.

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, but replacement questionnaires were offered, and reasons for nonresponse (when provided) were recorded.

A final source of data on the process came from a set of debriefing calls, conducted among respondents from both mail and e-mail treatments. We solicited respondent reactions to the content of the questionnaire and (in the case of e-mail) to the mode of data collection.

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 discuss each of these in turn.

First, we examine the response rates by agency and mode. These are presented in Table 2. For each of the five agencies, e-mail produced a significantly ($p < .01$) lower response rate than mail. This finding is consistent with that of most other tests of e-mail versus mail (Bachman, Elfrink, and Vazzana, 1996; Mehta and Sivadas, 1995; Schuldt and Totten, 1994; Sproull, 1996; Tse, 1996). The largest differences in response rate are found for agencies A, D, and E. There are several possible explanations. For Agency A and 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.

Table 2: 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

Another source of the difference may be technical problems with different e-mail systems at each agency. We discovered that Lotus CC:Mail can be set to automatically convert e-mail messages over a certain size (e.g., 20 Kb) into attachments. Both Agency A and D use CC:Mail. However, so does Agency B, which had the highest e-mail response rate. We received several reports from employees from Agency A and D that they received attachments, and didn't 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 D. As soon as we learned of this, we sent an additional e-mail message to sample cases in these two agencies with updated instructions on how to deal with attachments. Similar problems were not experienced at Agency C (Novell GroupWise) or Agency E (GroupWise or WPMail).

Further evidence for the technical problems caused by the size of the e-mail survey (23 Kb) can be found in the supplement response rates. 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. The main and supplement response rates for these two agencies are presented in Table 3. 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 19.5% (to 56.3%), while that for Agency D would increase by 16.3% (to 69.3%), 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 questionnaire.

Table 3: 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

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. It is thus important for us to explore what reasons there may be for the response rate differential, and what effect this may have on the quality of the data obtained.

In a **tracking database** of all returns, SRC staff noted which cases required special attention for

a variety of reasons. For an e-mail survey to be cost-effective, the goal is to minimize clerical activity required. Table 4 shows the various types of clerical action that were required for those e-mail questionnaires that were returned. These may also be indicators of the types of difficulty experienced by e-mail sample persons. The first column shows the percentage of returns that were 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 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. Overall, about 3.9% 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 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.

Table 4: Types of Clerical Action Required for E-Mail Returns (in percent)

Agency	Attach- ment	File coded	WP file	Edit	Needs action	Any clerical (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)

As noted earlier, we also conducted a set of telephone **debriefing interviews** with those who returned their questionnaires. A total of 694 sample cases were selected from among the respondents, using several replicates to include both early and late returns. The sample was evenly split between modes, and Agency A was undersampled because of its relatively large size.

Interviews were conducted by JPSM students, ensuring that no student called a respondent from their own agency or known to them. A small portion of the calls were conducted by members of an undergraduate survey methods class. An overall response rate (complete/eligibles) of 77.2% was obtained. The cooperation rate (complete/contacted) was 90.5% (including callbacks) or 98.3% (excluding callbacks). This yielded a total of 244 mail and 256 e-mail respondents who completed the debriefing. 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 5. These findings parallel those shown in Table 4, and suggest that the difficulty of replying to the survey differed across agencies. In Agency A and D, about two-thirds of respondents used a text editor or word processor to complete the survey, whereas the survey was

designed to be completed using a reply function within e-mail.

Table 5: 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)

We asked both sets of debriefing respondents (mail and e-mail) to estimate how long they took to complete the survey: e-mail respondents reported taking significantly longer ($p < .01$) than mail respondents (28.3 minutes versus 22.5 minutes). 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 agency. 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.

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 non-anonymity of e-mail may contribute to a reluctance to complete the survey in this mode. 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 6. Neither of these differ significantly ($p > .05$) 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.

Table 6: Mean Response to Two Questions About Access to Survey Responses by Mode, Debriefing Respondents

Question	Mail	E-mail
Supervisor access	6.15	6.61
Access by others in agency	6.32	6.36
(n)	(244)	(256)

The **reminder calls** may give us further insight into the reasons for nonreturn of the surveys. Toward the end of the study, we attempted to contact all remaining nonrespondents to encourage participation. 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 7. The "other" category includes wrong numbers, sample persons who had left the agency, and so on.

In Table 8 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 were not going to return the questionnaire (37.3% versus 22.9%). Among these, almost half (45.8% of the refusers and 17.1% of all those contacted) claimed that they did not receive the questionnaire by 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, suggesting that delivery of an e-mail instrument may be more problematic than that of a mail instrument. Second,

2.7% of the e-mail contacts (or 7.2% of those who said they would not respond) reported difficulties editing the instrument as a reason for nonreturn; as one would 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, 6.1% of the mail contacts who do not plan to respond mentioned confidentiality concerns, compared to 3.2% of e-mail contacts who planned not to respond. These findings again suggest that technical difficulties, rather than confidentiality concerns, largely account for the lower e-mail response rate.

Table 7: Outcome of Reminder Calls by Mode

Outcome (n)	Mail percent	(n)	E-mail percent	
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 8: 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	<u>22.9</u>	<u>37.3</u>
Total	100.0	10.0
Among those who refused, reasons given for nonreturn:		
Did not receive	3.9	8.0
Lost, deleted	3.7	9.1
Couldn't edit	0.0	2.7
No time	4.2	4.4
Confidentiality	1.4	1.2
Other, no reason	<u>9.7</u>	<u>12.0</u>
Total refused, other	22.9	37.3

A final source of data for evaluating the mode experiment comes from the **substantive responses** themselves. Given random assignment to mode, we would expect the distributions of key variables and the levels of item missing data to be similar. Table 9 contains item missing data rates by mode, for the 81 attitude items and 8 of the background items.

Table 9: Item Missing Data Rates by Mode

	Mail	E-mail
81 attitude items	0.63	0.64
8 background items	0.24	0.16
(n)	(2,969)	(1,724)

We see from Table 9 that the overall rates of missing data are low for both modes (on average less than 1 of the 81 attitude items missing per case). There are no significant ($p > .05$) differences in item missing data on the attitude items. Contrary to expectation, the mail mode has a significantly higher ($p < .01$) rate of missing data on the background measures. Inspection of the individual items suggests that several (e.g., years of service, grade level, managerial and supervisory status, and race) are susceptible to higher missing data rates on the mail questionnaire. One possible 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. This again suggests that confidentiality may not have been a major factor in noncompletion.

We also examined the distributions of both demographic and substantive variables across mode. 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 ($p < .01$ in each case) in the distributions of respondents in terms of grade level, managerial and supervisory status. These results are presented in Table 10. Overall, the direction of the effect is as expected: higher status employees appear to be over-represented in e-mail. These differences are striking, and suggest differential access to, or use of, e-mail. We also find significant differences by race and gender ($p < .01$), with non-minorities and males being more likely to respond by e-mail than by mail. These results are also presented in Table 10.

Table 10: 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
Gender:		
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 climate items, we assumed that nonrespondents may hold more negative attitudes toward their agency. Thus, with the higher nonresponse rate 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 5 of the 13 subscales, with mail having a higher (more positive) mean score on 3 of the 5, and e-mail on the remaining 2. Overall, mail respondents were more positive on 7 of the 13 subscales, and e-mail on 6. Thus, we find little support for our expectation on attitude differences. We also found no significant difference in the overall climate score, or in various other key single-item indicators such as satisfaction with the agency or employee morale. Thus, despite the differential nonresponse, the substantive responses of the mail and e-mail samples appear similar.

Finally, while we do not have a detailed cost breakdown for the two modes, we can offer a few observations on the cost implications of 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. Printing and postage costs were \$13,600 for mail and \$0 for e-mail. Keying the completed mail questionnaires cost about \$5,400 (about \$1.81 per completed case), whereas managing the e-mail sample (including the clerical action mentioned earlier) cost about \$3,000 (or \$1.74 per completed case). 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 e-mail, 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 e-mail offers potential savings in time and money over mail for organizational surveys, it seems clear that such benefits will not always be realized. Most other studies of mail versus e-mail have been conducted in relatively closed settings (e.g., within one organization), thus minimizing the technical difficulties we experienced. Despite pretesting the survey in each of the agencies, we did not anticipate the problems caused by size of message limitations on certain platforms. These problems suggest 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. One advantage of mail in such cases is that it ensures a standard treatment for all sample persons. 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.

Acknowledgments

The students in the JPSM practicum played an important 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. Without agency willingness to participate in this risky venture, the mode experiment could not have taken place. We also thank Skip Camp who, as SRC study manager, invested considerable time and effort in this project.

References

- Bachman, D., Elfrink, J., and Vazzana, G. (1996), "Tracking the progress of e-mail vs. snail-mail." *Marketing Research*, 8 (2): 31-35.
- Carlson, L.T., and Rivers, E.B. (1997), "Origins of the organizational climate survey of federal statistical agencies." Paper presented at the Joint Statistical Meetings of the American Statistical Association, Anaheim, CA, August.
- 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.
- Schuldt, B.A., and Totten, J.W. (1994), "Electronic mail vs. mail survey response rates." *Marketing Research*, 6 (1): 36-44.
- Sproull, L.S. (1986), "Using electronic mail for data collection in organizational research." *Academy of Management Journal*, 29 (1): 159-169.
- 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.