

**Survey Data Collection Strategies:
Response Differences Between Web-based and Paper-based Methods**

Thomas R. Lang

Richard A. Raver

James A. White

Kristine Y. Hogarty

Jeffrey D. Kromrey

University of South Florida

Paper presented at the annual meeting of the Florida Educational Research Association,
Tallahassee, Florida, November 8-10, 2000.

INTRODUCTION

The increasing popularity and wide availability of the World Wide Web in schools provides educational researchers with a vehicle for lowering the cost of and easing the effort required to collect and analyze data. Increasingly, research is being conducted that depends upon web-based collection of educational survey data. However, it is not well known how data collected in this fashion compare to data collected via traditional means. Differences have been reported based upon what an instrument measures, whether respondents are tested alone or in groups, and the manner of group assignment (Synodinos, Papacostas, & Okimoto, 1994; Richman, Kiesler, Weisband & Drasgow, 1999). As early as 1986, Kiesler and Sproull cautioned that, “electronic survey results should be calibrated against those obtained through more conventional means” (p. 300).

A key issue, according to Dillman (2000), is that survey populations accessed through the Web are often invalid samples of any general population because very few, relatively privileged people have Web access. There are signs that this status may change dramatically and soon. According to Stross (2000), it took 46 years for a quarter of the Nation’s population to get electricity; 35 years for the telephone; 16 for the personal computer; but only 7 for Internet access. The rate at which Internet access is becoming common suggests that the time may now be ripe to pursue “calibration” of web instrumentation. This paper describes the web-based data collection efforts that were conducted as a component of a larger research project on teachers’ perceptions of computers and technology, and how data collected in this manner compared to data collected via traditional means.

RESEARCH QUESTIONS

The researchers examined the following questions:

1. Are participants who completed and submitted the survey online representative of the same population as those who completed the pencil-and-paper survey?
2. Were the online survey responses biased in any way in comparison to the responses in pencil-and-paper mode?
3. Were return rates comparable across the two modes?

METHOD

Instrumentation

The purposes of the larger research project were to develop and validate an instrument focusing on technology in the classroom and to gather information regarding the level of technology use in classrooms by teachers and students. Initially, a comprehensive paper and pencil survey was developed containing individual sections addressing level of integration, support, preparation, confidence and comfort and attitude toward computer use. The web version of the survey was developed to investigate whether online data collection might be a viable alternative to more traditional data collection methods.

The web-based survey was developed to parallel the paper version as closely as possible, while still realizing the benefits made available through the use of online technology. The development and implementation of the online survey was conducted following the guidelines identified by White, Carey, & Dailey (2000), with the exception that an active server page (ASP) processed data submissions as opposed to a Perl script, and the data were written to a formatted database. As in White, Carey, & Dailey's efforts, JavaScript was used on the client to validate participants' data entry before submission. Dillman's (2000) 14 web survey design principles were also consulted and generally followed, with two notable variations. First, Dillman suggests that PIN numbers always be used to limit access. Given the nature and population of this study, the researchers thought the disadvantages of this procedure (confusion, lost data sets) outweighed the potential benefits (blocking invalid participants).

A second variation was made from Dillman's principle, "Do not require respondents to provide an answer to each question ... " (p. 394). This maxim was followed, but was implemented in a more refined manner. Data entry errors (incomplete or incorrect entries) were brought to the participants' attention and it was suggested that any errors be corrected, but no participant was forced to complete any field as a condition of submission. In this manner, some of the technological advantage may have been lost (more complete data sets), but the gain was that the process was made more comparable to the paper-based process and participants' rights were respected. Even at this level, however, it is possible that the interactive validation process may be a source of variance in participants' responses. A functional mock-up of the instrument is available online at <http://fcit.usf.edu/jwhite/ptsurvey.html>.

Sample Selection

The sample of schools was drawn from a population of 116 public schools in a large Florida school district. The district is composed of 16 high schools, 23 middle schools, and 82 elementary schools. Once the schools were sorted into groups according to grade level (i.e. elementary, middle, and high) they were matched based on school size, percent of students receiving free or reduced lunch, non-white representation and mobility rate. Approximately 20% of the schools at each grade level were selected for this study. The resulting 28 matched pairs, included 19 matched pairs of elementary schools, 5 pairs of middle schools, and 4 pairs of high schools. One member of each matched pair was randomly assigned to either the web or paper group. Information used to match the schools was obtained from the Florida Department of Education's web site and was based on the 1999-2000 school year.

For each group, letters were sent to the school Principals and to individual teachers, providing a rationale for the study along with detailed instructions. A copy of the paper survey was included with the letters sent to teachers selected for the paper and pencil administration. The teachers selected to respond via the Web received a letter directing them to the web site (however, principals in these schools were also provided paper copies of the survey for those teachers who wished to participate, but preferred a paper version; these respondents were included in the total sample, but were not used in the matched sample analysis). In order to increase the response rate, incentives were offered at both the school and individual levels. A website was provided to facilitate registration and to allow participants to keep track of the percent of teachers who responded from their school and from other participating schools.

Sample

The sample consisted of 770 respondents, 68% were from the paper group and 32% from the web group. Lack of response from three of the web schools resulted in a total of 25 pairs of schools. Eighty-three percent of the participants were female and 17% were male. Ninety-three percent classified themselves as White/non-Hispanic, just over 4% were African American, and approximately 1% were Hispanic, with the remainder split equally between American Indian and Other. Sixty percent reported their highest degree as a bachelor's, 37% held master's degrees, and less than one percent reported holding a doctorate. The remaining 3% was composed of Education Specialist or other degrees. The respondents' years of teaching experience ranged from 'first year' to more than twenty years experience, with a mean of 13 years. Sixty percent of the respondents reported teaching classes of 21 to 30 students.

Data Analyses

In order to establish the comparability of the two modes of survey administration, it was necessary to examine whether there were differences in the response patterns of those teachers who respond to the survey online and those who responded using a paper version of the survey. Specifically, we sought to examine a) the extent to which participants who completed and submitted online surveys were representative of the same population as those who completed pencil-and-paper surveys, b) the extent to which the nature of online survey technology changed the survey completion event in such a way that responses were biased in comparison to pencil-and-paper surveys, and c) the extent to which return rates across the two modes were comparable.

The researchers examined differences in the proportion of males and females who completed web versus paper surveys and examined the ethnic composition of the respondents for the two modes. The proportion of responses by mode was also examined to rule out potential differences in return rates.

Because the results from the factor analysis of those responding to the paper survey supported, in large part, the anticipated structure of the instrument, eleven composite variables were created based on the proposed classification scheme: *Confidence and Comfort*, *General School Support*, *Teacher Application Software Use*, *Teacher Instructional Software Use*, *Student Application Software Use*, *Student Instructional Software Use*, *Integration of Technology into Teaching*, *Personal Use*, *Technical Support*, *Technological Affinity* and *Technological Aversion*. For more information on the development and validation of the instrument see Hogarty and Kromrey (2000).

Potential differences in mean scores on these subscales were used for further exploration of the question of congruence between paper and web responses. Cronbach's alpha was computed for each of the subscales by mode to compare internal consistency. The internal reliability estimates ranged from .67 to .90 (see Table 2). Reliability estimates for the Confidence and Comfort subscale and the Technological Aversion subscale were the same between modes (.75 and .90, respectively). The largest discrepancy between the two modes, a difference of only .07, was evidenced for the Student Application Software Use subscale with a reliability estimate of .74 for the paper mode, and .67 for the web mode.

RESULTS

Based on a review of related literature and consequent efforts to make both survey instruments as similar as possible, the researchers expected to see no statistically significant differences in the responses between the two groups. However, a difference in response rate was anticipated due to manner in which the surveys were distributed. As stated earlier, for the

schools that were chosen to receive the paper version, teachers received the actual instrument to complete, whereas the web participants received a letter that directed them to a web site to complete the survey. This additional step may have influenced the rate of return for the web group. Therefore, analyses focused on the examination of group differences according to demographic classifications, differential response rates between modes, and differences in perceptions and attitudes toward the use of technology.

Gender

There was no statistically significant difference between the proportion of females and males responding to the web and paper versions of the survey ($t(24) = 0.03$, $p > .05$). There was less variability for the paper respondents than for the web respondents. The median proportion of female respondents was .88 with a semi inter-quartile range of .12. For the web administration, the median proportion of females responding was .86, with a semi inter-quartile range of .25. The distribution of the proportion of female respondents within each school is illustrated in Figure 1.

Race

There was no statistically significant difference between the proportion of white and non-white teachers responding to the web and paper versions of the survey ($t(24) = 2.07$, $p > .05$). The paper version had more variability, with the proportion of white respondents ranging from .83 to 1.0 (with a single outlier of .67). The median proportion of white respondents was .91, with a semi inter-quartile range of .04. The web respondents showed much less variability, with the proportion of white respondents ranging from .96 to 1.0 (with several outliers $< .90$). For the web schools, the median proportion of white respondents was approximately 1.0, with a semi inter-quartile range of .02. The distribution of the proportion of white respondents within each school is illustrated in Figure 2.

Return Rates by Mode

The response rate between modes showed a statistically significant difference, with high response rates associated with the paper mode ($t(24) = -12.78$, $p < .05$). There was more variability within the set of schools that responded to the paper version with the proportion of responses ranging from approximately .05 to .80 with an outlier approaching 1.00. The median response rate for the paper version was .39 with a semi inter-quartile range of .10. The web respondents showed much less variability with a range of 0 to .25 with several outliers greater

than .60. The median response rate for the web respondents was .10 with a semi inter-quartile range of .07. The distribution of the within school response rates are illustrated in Figure 3.

Subscale comparisons

The means and standard deviations for each of the eleven subscales by mode are presented in Table 1. A series of ANOVAs were conducted to test for group differences on the eleven subscales. Holm's modified Bonferroni procedure was used to control the familywise alpha for the eleven tests at the .05 level. There were no statistically significant differences between the two modes for any of the subscales. Effect sizes were computed and are presented along with the *F* statistics and probabilities in Table 1. The effect sizes range from -0.12 to 1.03 for the eleven composite variables. Examination of the effect sizes suggests slightly higher means for the web schools, (approximately three-quarters of a standard deviation or larger) for 5 of the subscales: Confidence & Comfort, Teacher Application Software Use, Personal Use of Computers, Technological Affinity and Technological Aversion. Cohen would suggest that an effect of .2 is large enough to be of interest. However, when interpreting effect sizes we need to think about the practical significance of the differences. For example if we look at the composite variable 'Technological Aversion' we see an effect size of 1.03 (a large effect by Cohen's criteria). This represents a 1 standard deviation difference between sample means. However, the pooled standard deviation of the school means is only .20, so this effect size represents a mean difference of approximately 1/5 of a point on a five-point scale. From a practical point of view, this would suggest no practical difference between the groups.

CONCLUSIONS

Our findings suggest that a difference exists in the rates of response between the two modes (.39 for the paper version versus .10 for the web version). This difference suggests that teachers who were given the paper version were more likely to return the survey than those given the web version. However, no statistically significant differences were evidenced between the paper mode and web mode regarding either gender or racial representation in responding. Additionally, an examination of responses on the eleven subscales suggests no statistically significant differences in teachers' perceptions and attitudes by mode. Although the potential exists for the identification of differences in these samples on additional factors, these data support the notion that the web sample and the paper sample are representative of the same general population.

In summary, this investigation of comparability between data collection methods involved many participants and a variety of variables. The potential for differentiation by data

collection mode was large, yet in most respects, it did not occur. The results of the study provide additional evidence to add to the growing body of evidence that, with well-designed instruments and data collection procedures, valid survey data may be collected via the World Wide Web.

References

Cohen, J., (1988). *Statistical Power analysis for the behavioral sciences*. (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

Dillman, D. (2000). *Mail and internet surveys: The tailored design method*. (2nd ed.). New York: John Wiley & Sons.

Hogarty, K. Y., & Kromrey, J. D. (2000, November). *The Nature of Technology use in Classrooms: The Development and Validation of an Instrument to Measure Teacher's Perceptions*. Paper presented at the annual meeting of the Florida Educational Research Association, Tallahassee, Florida.

Stross, Randall E. (2000). Digital divide hooey: Access to the internet is no black and white issue. *U.S. News Online*. Retrieved May 22, 2000 from the World Wide Web: <http://www.usnews.com/usnews/issue/000417/digital.htm>

White, J., Carey, L., & Dailey, K. (in press). Web-based instrumentation in educational survey research. *WebNet Journal: Internet Technologies, Applications and Issues*.

Table 1
Survey Scale Scores by Response Mode.

Scale		Paper Respondents	Web Respondents	Cohen's d	F	p
Confidence and Comfort	M	3.30	3.61	0.81	6.05	0.02
	SD	0.29	0.47			
General School Support	M	3.43	3.38	-0.12	.20	0.66
	SD	0.36	0.56			
Teacher Application Software Use	M	2.60	2.88	0.86	5.60	0.03
	SD	0.32	0.34			
Teacher Instructional Software Use	M	1.54	1.62	0.20	0.56	0.46
	SD	0.36	0.46			
Student Application Software Use	M	1.90	2.11	0.47	2.34	0.14
	SD	0.30	0.59			
Student Instructional Software Use	M	2.37	2.64	0.43	3.13	0.09
	SD	0.52	0.75			
Integration	M	2.57	2.79	0.40	2.34	0.14
	SD	0.50	0.63			
Personal Use	M	3.21	3.61	0.88	9.72	0.01
	SD	0.38	0.51			
Technical Support **	M	3.55	3.47	-0.13	0.17	0.69
	SD	0.40	0.77			
Technological Affinity	M	4.07	4.24	0.71	5.23	0.03
	SD	0.16	0.32			
Technological Aversion	M	3.98	4.19	1.03	6.88	0.02
	SD	0.16	0.24			

Note: ** Responses from teachers in 23 matched pairs of schools

Table 2
Cronbach's Alphas by Mode.

Composite variable	Paper	Web
Confidence and Comfort	.90	.90
General School Support	.82	.84
Teacher Instructional Software Use	.78	.81
Teacher Application Software Use	.77	.74
Student Instructional Software Use	.78	.77
Student Application Software Use	.74	.67
Integration	.89	.87
Personal Use	.72	.76
Technical Support	.83	.85
Technological Aversion	.79	.77
Technological Affinity	.75	.75

Figure 1

Distribution of Respondent Gender by Survey Mode

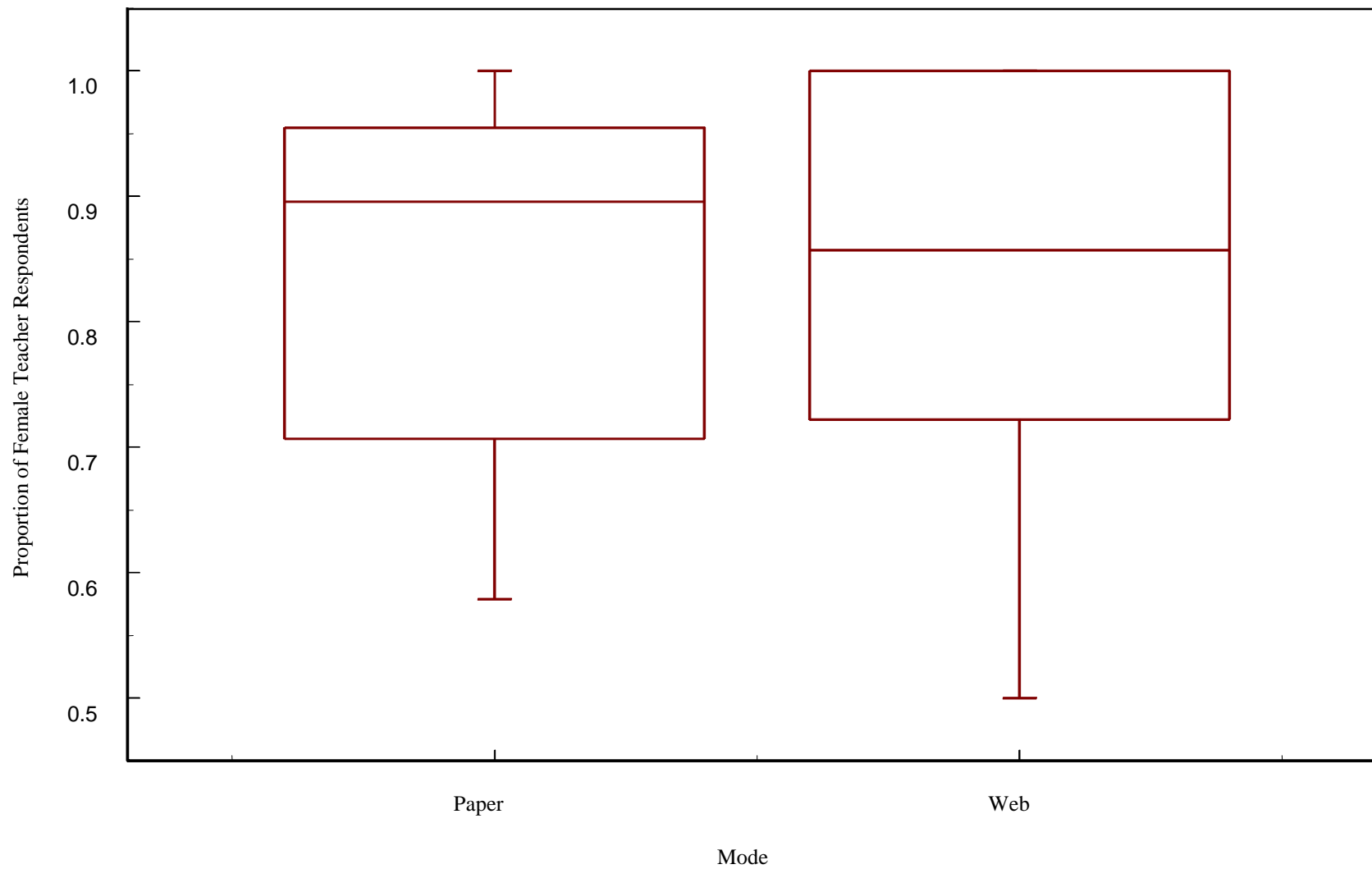


Figure 2

Distribution of Respondent Race by Survey Mode

