

RESPONDENT "HONESTY": A COMPARISON OF COMPUTER VERSUS PAPER AND PENCIL QUESTIONNAIRE ADMINISTRATION

DAVID MEINERT, *Southwest Missouri State University*

SCOTT J. VITELL, *University of Mississippi*

RAY BLANKENSHIP, *Marshall University*

This article examines the differences between interactive computer-based questionnaires and traditional paper and pencil questionnaires. Respondents were asked, using two types of instruments, to respond to a variety of established research scales, including social desirability, locus of control, the ethical position questionnaire and life optimism. Comparisons were then made between the results for these two different modes of questionnaire administration. Results indicate that respondents, when using a computerized format, may respond more "honestly" for some "sensitive" scales such as the ethical idealism scale. For less sensitive questions, there appears to be no difference between computerized and paper and pencil formats in terms of respondent candor.

INTRODUCTION

The use of computer-administered questionnaires for data collection is a rather recent development in the field of marketing (i.e., survey) research (O'Brien and Dugdale, 1978; Higgins, Dimnik and Greenwood, 1987; Liefeld, 1988). The inherent economic advantages of computerized methods, along with their ability to automate the more mundane aspects of data collection, leads us to expect that an ever-growing percentage of self-report survey data will be collected via an electronic mode. In light of the espoused benefits (e.g., Newsted, 1985; Rafael, 1984; and Whalen, 1984), the considerable interest computerized methods for data collection have generated is not surprising. What is surprising is the lack of research focusing on issues arising from this new computer-based method for collecting self-report survey data from respondents. While Fellers, Anson and Anson (1990) note the literature is "supportive" of the

expanded use of computerized surveys for collecting research data, they fail to note the paucity of empirical research evaluating the benefits envisioned by such proponents.

The possibilities for computers to improve the efficiency and effectiveness of self-report surveys is exciting. The elimination of duplication costs, replacement of professional survey administrators/interviewers with less costly facilitators, and elimination of subsequent data entry associated with computerized questionnaires are frequently cited as improvements in efficiency (e.g., Perreault, 1992; Burke and Normand, 1987; Erdman, Klein, and Greist, 1985; Merrell, 1985; Perlman, 1985). Well-documented improvements in effectiveness include: improved standardization, support for adaptive questionnaire strategies, flexibility in scheduling and improved data validity resulting from on-line validation/consistency checks as well as the elimination of subsequent data entry (Higgins, Dimnik and Greenwood, 1987; Merrell, 1985; Perlman, 1985; Wood, 1984).

While it is self-evident that researchers should capitalize on these inherent advantages, it should be equally clear that computerized administration should never sacrifice the *sine qua non* of data collection — quality information. As Klingler, Johnson, and Williams (1976) note, too often computerized methods are "blindly accepted as superior to traditional non-computerized procedures." While the efficiency and effectiveness gains noted above are well documented, the accuracy or quality of self-reported information collected via computerized questionnaires has not been satisfactorily established (Skinner and Allen, 1983; Higgins, Dimnik and Greenwood, 1987; Couper and Rowe, 1996).

To address some of the above concerns about computerized questionnaires, this article reports the results of an experiment designed to investigate the effect of computerized questionnaire administration on one aspect of the quality (i.e., "honesty") of information obtained. The experiment compares the responses from computerized and written modes of data collection for several well-established research scales to assess whether computerized formats lead to more "honest" responses (i.e., quality information).

LITERATURE REVIEW

Despite the recency of computerized formats for marketing research surveys, several related studies have attempted to measure the impact of administration mode on quality or accuracy of data collection. Our review of the literature has necessarily included such diverse fields as psychology (Evan and Miller, 1969; Koson, Kitchen, Kochen and Stodolosky, 1970; Allen and Skinner, 1983), medicine (Lucas, Mullins, Luna and McInroy, 1977; Waterton and Duffy, 1984), and psychiatry (Carr, Ghosh, and Ancill, 1983; Greist and Klein, 1980; Klingler, Johnson and Williams, 1976). These early adopters utilized computerized administration in clinical settings for comprehensive physical, social, behavioral and psychiatric assessments. These included special issue questionnaires such as suicidal threat, substance abuse, and sexual dysfunction. Studies by Higgins, Dimnik, and Greenwood (1987), O'Brien and Dugdale (1978) and Kiesler and Sproull (1986) involved more

traditional market research and/or public opinion surveys.

In a pioneering study, Evan and Miller (1969) found that experimental subjects responded with "greater honesty and candor" under computer questionnaire administration as opposed to self-administered paper and pencil questionnaires. Koson, Kitchen, Kochen and Stodolosky (1970) developed a "Threat Scale" similar to a social desirability scale, and found that "threatening and disturbing" items generated no more "response bias" or "defensiveness" in a computerized format than in a self-administered paper or a personal interview mode. Skinner and Allen (1983), in one of the few experiments designed specifically to investigate the accuracy of information issue, compared traditional face-to-face interviews, self-report questionnaires, and on-line computer interviews for collecting substance abuse data. No significant differences were found in the level of substance abuse reported across the three assessment formats.

In a comparison with traditional personal interviews, Lucas, Mullins, Luna, and McInroy (1977) and Waterton and Duffy (1984) found that patients in alcohol treatment centers report significantly higher levels of alcohol consumption with a computerized format.

Carr, Ghosh and Ancill (1983) found that a computerized psychiatric history resulted in patients divulging more sensitive information such as criminal behavior, attempted suicides, repeated job dismissals and sexual dysfunctions. Greist and Klein (1980) reported that individuals were more likely to report sexual problems to a computer than to a personal interviewer. In Klingler, Johnson and Williams' (1976) comparison of face-to-face interviews, approximately one half of the psychiatric patients reported that they tended to be "more truthful" when responding to a computer.

In an early marketing study, O'Brien and Dugdale (1978) compared responses concerning bathing/hygiene habits by computer to that of an orally administered questionnaire. After noting that differences were small in absolute terms and not

statistically significant on an individual basis, O'Brien and Dugdale (1978) went on to describe a "marginal tendency" for respondents to be more "frank" and willing to use extreme scale values when responding to a computer. A more recent marketing study by Higgins, Dimnik and Greenwood (1987) compared diskette and paper questionnaires. No significant differences in responses to "disquieting" attitudinal questions, hypothesized to induce more socially acceptable answers, were found.

Kiesler and Sproull (1986) investigated the hypothesis that an electronic survey would reduce social context information, which is believed to invoke more socially desirable responses. Responses to computer and paper versions of a health attitudes, behaviors and personal traits survey were compared. Based on differences in responses to selected items borrowed from the "Need for Approval Scale" (Crowne and Marlowe, 1964), Kiesler and Sproull concluded that the computer mode resulted in less socially desirable responses.

Similarly, Couper and Rowe (1996), using four items from the Crowne and Marlow scale, found that computer assisted self-administration led to reduced overreporting of socially desirable responses. In this study the mode of administration was also found to affect substantive responses and data quality. Expressing concern about possible nonresponse error, this same study examined the decision to participate in a computer survey. The decision to complete a computer survey was found to be related to factors such as respondents' age, education and computer experience.

More recently, researchers have examined the implications of utilizing audio and video in computer-assisted self-interviews (CASI) and computer-assisted personal interviewing (CAPI) (Boekeloo, Schiavo, Rabin, Conlon, Jordan and Mundt, 1994; Johnston and Walton, 1995; O'Reilly, Hubbard, Lessler, Biemer and Turner, 1994). Inclusion of audio and video is intended to overcome in part physical and cognitive demands of self-administration.

Generalization of past evaluative efforts to assess the impact of computerized questionnaire administration

on the "honesty" of information obtained is difficult. On the whole, studies to date reveal somewhat inconclusive evidence of systematic differences in the quality of data collected. In addition to inconsistent findings, the range in quality of experimental designs, including survey instruments, is considerable. Subjects are frequently from treatment-seeking populations (e.g., Boekeloo, Schiavo, Rabin, Conlon, Jordan and Mundt, 1994), with findings partially anecdotal and statistical results more often than not unreported. Thus, further research is needed to establish whether the limited findings, which suggest computerized administration improves the "honesty" of information, generalize to other settings and populations such as traditional marketing and consumer research. This paper describes an experimental study designed to yield results that may overcome many of these limitations.

RESEARCH HYPOTHESES

The following research hypotheses were tested in this study:

H₁: Computerized questionnaires will lead to more "honest" (i.e., less socially desirable) responses than "paper and pencil" questionnaires.

H₂: The difference between computerized questionnaires and "paper and pencil" questionnaires will be the greatest for more "sensitive" scales.

Hypothesis 1 is based on the fact, partially supported by previous research, that respondents will answer more candidly when they perceive that their responses are more anonymous, as should be the case with the computerized questionnaire. This latter assumption is based upon the above literature review. The second hypothesis is based on the assumption that the difference between these two methods of questionnaire administration will manifest itself only when the questions are of a sensitive nature. That is, for relatively innocuous questions there will be no significant difference between computerized and "paper and pencil" questionnaires. The methodology employed and the data collected to test these research hypotheses are described in the following sections.

METHODOLOGY

Sample and Sample Characteristics

The sample consisted of undergraduate business students at a major southeastern university. Of a total of 172 students who participated in the experiment, a group of 80 students was first administered a computerized version of the questionnaire followed by a paper and pencil version four weeks later. A second group consisting of 92 students was first given the paper and pencil questionnaire followed by the computerized version in four weeks. Students were asked *not* to identify themselves on either the computerized or the paper and pencil questionnaires. However, they were asked to sign a separate sheet so that the researchers would know which students had, and had not, completed both formats of the questionnaire.

A limited number of demographic questions were asked. The sample consisted of 66 percent males, with most students being either marketing majors or general business majors. Almost all respondents were either juniors or seniors (79 percent). Finally, over 75 percent had completed a course which used computers extensively (i.e., an introductory course in information systems) or else had prior work experience using computers (i.e., at least three months).

Measures

Several well-established research scales were included in the questionnaire. The first part included twelve questions measured on a five-point Likert scale with (1) being "strongly agree" and (5) being "strongly disagree." Eight of these make up the life optimism scale, while the other four were filler questions that usually accompany this scale (Scheier, Weintraub and Carver 1986). Examples of the eight relevant items are "I'm always optimistic about my future" and "In uncertain times, I usually expect the best." Cronbach's coefficient alpha, measuring reliability, for the single dimension of this scale is .956 for the computerized version and .945 for the paper and pencil version.

The second part of the questionnaire consisted of 33 questions measuring social desirability (Crowne and Marlowe 1964). These were measured by a dichotomous scale (i.e., true/false), and form a single dimension. Questions vary from, "I like to gossip at times" to "I have never intensely disliked anyone." Since this is a dichotomous scale, the appropriate measure of reliability is Kuder-Richardson-20 which is .989 for the computer version and .985 for paper and pencil.

Part three was the 20 item consumer ethics scale developed by Muncy and Vitell (1992). This scale consists of 20 potentially "unethical" consumer practices measured on a five-point scale with anchor points (1) "strongly believe it is wrong" and (5) "strongly believe it is not wrong." This scale has four distinct dimensions. The first might be labeled as "actively benefitting from an illegal activity." The most distinguishing characteristics of these actions are that they are all initiated by the consumer (i.e., changing price tags on merchandise in a store) and they are all likely to be perceived as illegal by most consumers. Five questions make up this dimension, with the coefficient alpha for the computer version being .666 and for the paper version .720.

The second dimension can be labeled as "passively benefitting at the expense of others." Here the consumer benefits from the seller's mistake (i.e., getting too much change and not saying anything) rather than his own actions. The coefficient alpha for the four items of this dimension is .804 for the computer version and .835 for paper and pencil. The third dimension has been labeled as "actively benefitting from a questionable, but not necessarily illegal, action." Here, as in the first factor, the consumer initiates the action, but these are not as likely to be perceived as illegal (i.e., not telling the truth when negotiating the price of a new automobile). The four items of this dimension yield a coefficient alpha of .647 for computer and .715 for paper and pencil.

The fourth and final dimension can be labeled as "no harm/no foul." These are actions that most consumers perceive as not resulting in any harm (i.e., "using computer software or games you did not buy")

and, therefore, at least some consumers perceive them as acceptable actions. The five questions comprising this factor have a coefficient alpha of .569 for computer versions and .630 for paper and pencil versions.

The fourth part of the questionnaire consisted of the ethics position questionnaire (EPQ) developed by Forsyth (1980). It was measured on a nine-point Likert scale from (1) "completely disagree" to (9) "completely agree." The EPQ has two separate dimensions, an idealism scale and a relativism scale, each containing ten items. One is designed to measure ethical idealism, or the acceptance of moral absolutes, and the second is designed to measure ethical relativism, or the rejection of universal moral principles. Cronbach's coefficient alpha for the idealism scale is .789 for the computerized version and .799 for the paper and pencil version while for the relativism scale it is .770 for the computer version and .777 for paper and pencil.

Part five consisted of a short form (13 items) of Rotter's (1966) Locus of Control Scale. This forced-choice scale asks respondents to choose which of two statements is representative of the real-world situation. An example of one pair of such statements is, "One of the major reasons why we have wars is because people don't take enough interest in politics" and "There will always be wars, no matter how hard people try to prevent them." The theoretical range of scores on this scale is from 1 to 13, with a higher score indicating a more external locus of control. The Kuder-Richardson reliability for this scale is .974 for the computer version and .965 for paper and pencil. The last part of the questionnaire contained a series of demographic questions.

Sensitivity of Scales

Since one of our hypotheses is dependent upon the perceived "sensitivity" of the different scales, a separate sample of undergraduate business students was used to examine this issue. A separate sample was used so as not to risk alerting the main sample to the true nature of the study. This secondary sample of 88 students was asked the following question for each scale, "I would feel comfortable answering most of the

questions on page #___" (each scale appeared on a separate page). A 7-point agree-disagree scale (with 1 being strongly agree and 7 strongly disagree) was used.

Results indicate that respondents were most comfortable with the life optimism and locus of control scales and least comfortable with the consumer ethics scales, social desirability and the ethics position questionnaire. Additionally, individual 1-tail t-tests indicate that there was a significant difference in terms of "comfortableness" between the life optimism and locus of control scales and each of the other three scales (i.e., consumer ethics, social desirability and the ethics position questionnaire). No significant differences were found between locus of control and life optimism nor were there any significant differences among the other three scales.

Thus, our results indicated that the consumer ethics scales, social desirability and the ethics position questionnaire were the "sensitive" scales in our study whereas life optimism and locus of control were the "non-sensitive" ones. However, since all four consumer ethics scales had been grouped together for purposes of determining sensitivity, no data was available on the sensitivity of the four individual dimensions. However, based upon the nature of the questions and the fact that respondents in previous studies (e.g., Muncy and Vitell 1992) had perceived these practices to be acceptable, we judged that the "no harm/no foul" dimension of the consumer ethics scale was also likely to be "non-sensitive."

FINDINGS

Our research hypotheses posit that computerized questionnaires will lead to more "honest" (i.e., less socially desirable) responses, and that differences with respect to candor will be the greatest for more "sensitive" scales. Based upon the discussion in the previous section, this suggests that differences would be the greatest for the social desirability scale, three of the four dimensions of the consumer ethics scale (the exception being the "no harm/no foul" dimension), and the two dimensions of the EPQ scale since these scales contain the most "sensitive"

Respondent "Honesty": A Comparison . . .

Meinert, Vitell and Blankenship

questions. Due to a lack of "sensitive" items, the remaining scales were not expected to elicit any differences. Hence, these "non-sensitive" scales served as controls. The mean values for each scale for the computerized versus the paper and pencil version, by test group, appear in Table 1.

TABLE 1
Means for Scales by
Questionnaire Method and Group

Scale	Group 1 ^a Paper	Com-puter	Group 2 ^b Paper	Com-puter	Deemed "Sensitive"
Locus of Control	16.95	17.60	17.59	17.36	No
Social Desirability	14.39	14.94	14.60	14.18	Yes
Life Optimism	19.51	19.34	19.05	18.88	No
Ethics Position (EPQ):					
Idealism	6.34	6.16	6.67	6.50	Yes
Relativism	5.98	5.76	5.32	5.48	Yes
Consumer Ethics ^c					
Active, Illegal	1.99	1.92	1.85	1.81	Yes
Passive	2.86	2.85	2.82	2.78	Yes
Active, Questionable	2.76	2.64	2.74	2.61	Yes
No Harm	4.14	4.11	4.15	4.20	No

^aGroup 1 was administered the computerized version first.

^bGroup 2 was administered the paper and pencil version first.

^cConsumer Ethics Scale:

Active, Illegal = Actively benefiting from an illegal act.

Passive = Passively benefiting at the expense of others.

Active, Questionable = Actively benefiting from a questionable, but not necessarily illegal act.

No Harm = No harm/no foul.

Before examining the results of these hypotheses, however, it was necessary to first verify that there was no order effect, in terms of which version was administered first, nor any gender bias. Thus, those who completed the paper and pencil questionnaire first and those who completed the computerized questionnaire first, as well as males and females, could be aggregated for comparison purposes.

In order to accomplish this, nine separate analyses of variance were performed (one for each scale) with the difference between the computer version and the paper

and pencil version of a specific scale as the dependent variable. The independent variables were the group that a student was in (to test for order effect) and gender (in order to test for gender effect).

The results showed no significant "order" differences (.05 level) between groups for seven of the nine scales, indicating that the data could be aggregated for these scales. The two exceptions were locus of control and relativism. The latter was also the only scale that exhibited any significant "gender" differences. Although these results may be due to a statistical "quirk," because of possible order bias, any subsequent, aggregated findings for the locus of control and relativism scales must be viewed with some caution. Additionally, the social desirability scale showed significant differences between groups at the .056 level of significance, indicating that, perhaps, data for this scale should not be aggregated either. The remaining six scales showed levels of significance of .250 or higher, indicating that the two groups could be combined without concern.

In order to test the hypotheses, groups 1 and 2 were combined and individual t-tests were performed on the means of each scale, comparing the computer and paper and pencil versions. Because of the possibility of an order effect, no aggregate analyses were performed on the locus of control and the relativism scales. The results for the remaining scales appear in Table 2 and indicate that there were no significant differences between the two versions for the following scales: life optimism, "no harm/no foul," social desirability, "actively benefitting from an illegal act," and "passively benefitting at the expense of others." The first two of these scales were not expected to exhibit any differences and, thus, these results support our hypotheses.

However, the latter three scales were considered "sensitive," and, thus, differences were expected. The problem of possible order effect for the social desirability scale may explain why it showed no differences between the computer and paper and pencil versions, but the same cannot be said of the other two scales. Perhaps, the "actively benefitting from an illegal activity" dimension of the consumer ethics scale included such extreme items (i.e.,

Respondent "Honesty": A Comparison . . .

"changing price-tags on merchandise in a store") that students had no problem responding "honestly" irrespective of the method of questionnaire administration. The authors, however, can offer no explanation as to why the "passively benefitting" scale showed no differences due to administration method. Perhaps, the students involved in this research simply did not perceive these questions as being very "sensitive." This finding suggests future research would benefit from more extensive pretesting to differentiate between "sensitive" and "non-sensitive" questions. The relative "sensitivity" of specific questions is likely a function of the questions themselves, the sample frame and context of the survey.

TABLE 2
Results of t-Tests for Differences
in Means for Combined Sample

Scale	Means		t-Value	Deemed "Sensitive"
	Paper	Computer		
Social Desirability	14.53	14.53	0.02(.980)	Yes
Life Optimism	19.26	19.09	-0.78(.438)	No
Ethics Position (EPQ):				
Idealism	6.51	6.34	-2.37(.019)	Yes
Consumer Ethics ^a				
Active, Illegal	1.91	1.86	-1.34(.183)	Yes
Passive	2.84	2.81	-0.57(.570)	Yes
Active, Questionable	2.75	2.62	-3.34(.001)	Yes
No Harm	4.14	4.16	0.40(.692)	No

^aGroup 1 was administered the computerized version first.

^bGroup 2 was administered the paper and pencil version first.

^cConsumer Ethics Scale:

Active, Illegal = Actively benefitting from an illegal act.

Passive = Passively benefitting at the expense of others.

Active, Questionable = Actively benefitting from a questionable, but not necessarily illegal act.

No Harm = No harm/no foul.

As a final comment, the "actively benefitting" and "passively benefitting" scales, while not significant, did show differences in the hypothesized direction with the computer version yielding more "honest" results. That is, the computer version tended to show a stronger belief that these behaviors were not wrong.

Meinert, Vitell and Blankenship

Additionally, the computer versions tended to result in less socially desirable (i.e., more honest) responses as anticipated.

As posited, the idealism scale and the "actively benefitting from a questionable, but not necessarily illegal, activity" scale both showed differences due to administration method. In addition, these differences were in the expected direction with the computer version yielding more "honest" answers. The computer version tended to show that students were less idealistic and that they were less inclined to believe that "actively benefitting from questionable acts" was wrong. Thus, there is some support for our hypothesized position. Apparently, for most questions there is little or no difference between computer and paper and pencil methods, but for select "sensitive" questions the computer-based mode may be a better approach.

In light of possible order differences for the locus of control and relativism scales as well as the social desirability scale, t-tests were performed for these scales for each individual group. The results appear in Table 3. As can be seen, the only significant difference between computer and paper and pencil versions was for group 1 for the locus of control scale. The computer version, which was administered first, tended to show that respondents were more "external." This was contrary to expectations of no difference. However, there was no significant difference between the two versions for the other group tested.

Additionally, there were no differences between the two versions for either the social desirability or the relativism scale. Nevertheless, the direction of differences for both groups indicates that the more "honest" responses were obtained for the version that was administered first regardless of whether it was computer or paper and pencil. That is, the version administered first resulted in a lower social desirability score and a higher relativism score. Thus, there appears to be an order effect for some items when a questionnaire is administered twice to the same individuals within a short period of time. This result is not surprising. However, the fact that the initial administration seems to be the one yielding

Respondent "Honesty": A Comparison

Meinert, Vitell and Blankenship

the more "honest" answers is both a fortuitous and a significant finding.

TABLE 3
Results of t-Tests for Differences
in Means for Individual Groups

Scale	Means		t-Value	Deemed “Sensitive”
	Paper	Computer		
Locus of Control			No	
Group 1 ^a	16.96	17.59	3.23(.002)	
Group 2 ^b	17.59	17.36	-1.14(.256)	
Social Desirability			Yes	
Group 1 ^a	14.44	14.94	1.31(.194)	
Group 2 ^b	14.60	14.18	-1.42(.158)	
Relativism			Yes	
Group 1 ^a	5.98	5.76	-1.71(.091)	
Group 2 ^b	5.32	5.48	1.56(.123)	

^aGroup 1 was administered the computerized version first.

^bGroup 2 was administered the paper and pencil version first.

There are limitations to all experimental research, and it is important to note those that may affect the results of this study. Clearly, the results of this investigation are bound by the nature of the subjects employed and the experimental design, as well as the particular scales examined and the characteristics of the computer software utilized. First, one could question how typical student subjects are in relation to the general population. Their familiarity and/or proficiency with computer technology may have influenced their responses, and hence our findings.

Nevertheless, as Hunt (1990, p. 173) states, "Using students as subjects is entirely appropriate in many instances." While he does not explicitly mention the specific instances in which student samples are appropriate, the authors believe that the current study is an appropriate context for using a student sample. Students represent a segment of the population that is very familiar with computers (Zimmerman 1993), perhaps more so than any other segment of the population, and "computer literacy" is clearly a requirement of the population sampled in this study. The study could not have been conducted in the manner in which it was if the sample used was not familiar with using the computer. Additionally, since

we are not trying to generalize our results to a larger population, but are simply attempting to examine the differences between survey techniques for a "computer literate" population such as the one we did use, a student sample is appropriate.

Secondly, the results may have been distorted by the experimental design, most notably the order of presentation of the scales within the questionnaire and the short length of time between repeated measures. Thirdly, the "sensitivity" of the research scales selected for this study may have been insufficient to produce significant differences based upon the mode of administration.

Finally, the results may have been influenced by the nature of the computer software utilized. Alternative programming options such as menu selection versus point-and-click, monochrome versus color displays, or the inclusion of audio and/or video, may also function as confounding variables. The extent to which such limitations affected the results is unknown; however, the findings clearly demonstrate the capacity for the mode of administration to have some influence on respondent "honesty."

IMPLICATIONS AND CONCLUSIONS

A computerized format may be capable of overcoming the problem of unconscious feedback by personal interviewers and reduce the embarrassment of recording intimate data on paper to be reviewed by other individuals. The impartiality and anonymity of computerized questionnaires is believed to foster greater acceptance by respondents (Erdman, Klein and Greist 1985; Turkington 1984). Computerized questionnaires may reduce, at least some, respondent anxiety or embarrassment when sensitive data are sought. The embarrassment associated with sharing data of a personal nature directly with another individual via a personal interview or written questionnaire may be reduced with a computerized format. For example, following the response to a potentially embarrassing or sensitive item, the item and the response can be quickly "whisked" from the screen and replaced by the next item.

An order effect also seems to exist for some scales. That is, regardless of the method used initially, more "honest" responses may be obtained in that initial administration than in a subsequent one. Since most marketing research involves only a single administration of an instrument, this is a reassuring finding. Our only explanation for this is that respondents may tend to "analyze" the questions too much when seeing them twice within a short period of time, perhaps leading to less candid responses. While the results of this study are not definitive, they indicate that a single administration of a computerized version of an instrument may be preferred under certain conditions, namely, where questions may be "sensitive."

While computer-based questionnaire administration has attracted widespread interest, few studies have attempted to discern its ability to collect higher quality data than traditional non-computerized methods. Although this study found only moderate differences in the "honesty" of responses for selected scales, computer-based questionnaire administration may influence responses to specific types of items or specific question formats (i.e., semantic differential). Further examination of response differences and alternative computer software design attributes is needed to provide researchers with more definitive evidence concerning the effects, if any, on the quality and honesty of data collected.

REFERENCES

- Boekeloo, B.O., L. Schiavo, D.L. Rabin, R.T. Conlon, C.S. Jordon and D.J. Mundt. (1994) "Self-reports of HIV Risk Factors by Patients at a Sexually Transmitted Disease Clinic: Audio vs Written Questionnaires," *American Journal of Public Health*, 84, No. 5, 754-760.
- Burke, Michael J. and Jacques Normand (1987). "Computerized Psychological Testing: Overview and Critique," *Professional Psychology: Research and Practice*, 18, No. 1, 42-51.
- Carr, A.C., A. Ghosh, and R.J. Ancill (1983). "Can a Computer Take a Psychiatric History?," *Psychological Medicine*, 13, 151-158.
- Couper, M.P. and G. Burt. (1994) "Interviewer Attitudes Toward Computer-Assisted Personal Interviewing (CAPI)," *Social Science Computer Review*, 12, No. 1, 38-54.
- Couper, Mick P. and Benjamin Rowe. (1996) "Evaluation of a Computer-Assisted Self-Interview Component in a Computer-Assisted Personal Interview Survey," *Public Opinion Quarterly*, 60, 89-105.
- Crowne, D. and D. Marlowe (1964). *The Approval Motive*, Wiley and Sons: New York.
- Erdman, Harold P., Marjorie H. Klein and John H. Greist (1985). "Direct Patient Computer Interviewing," *Journal of Consulting and Clinical Psychology*, 53, No. 6, 760-773.
- Evan, William M. and James R. Miller, III (1969). "Differential Effects on Response Bias of Computer Versus Conventional Administration of a Social Science Questionnaire: An Exploratory Methodological Experiment," *Behavioral Science*, 14, 216-227.
- Fellers, Jack W., Robert G. Anson, and Patricia Harvey Anson (1990). "An Experimental Evaluation of Response Bias and Respondent Preference for a Computerized Survey as Compared to a Paper and Pencil Survey," *Proceedings of HICSS-23*, R. Sprague, ed., 488-495.
- Forsyth, Donelson R. (1980). "A Taxonomy of Ethical Ideologies," *Journal of Personality and Social Psychology*, 39, No. 1, 175-184.
- Greist, J.H. and M.H. Klein (1980). "Computer Programs for Patients, Clinicians, and Researchers in Psychiatry," In *Technology in Mental Health Care Delivery Systems*, J.B. Sidowski, J.H. Johnson, and T. A. Williams, eds., Norwood, N.J: Ablex, 161-182.
- Higgins, C.A., T.P. Dimnik, and H.P. Greenwood (1987). "The DISQ Survey Method," *Journal of Marketing Research Society*, 29, No. 4, 437-445.
- Hunt, S.D. (1990) "Commentary on an Empirical Investigation of A General Theory of Marketing Ethics," *Journal of Academy of Marketing Science*, 18 (Spring), 173-177.
- Johnston, J., and C. Walton. (1995) "Reducing Response Effects for Sensitive Questions: A Computer-Assisted Self Interview with Audio," *Social Science Computer Review*, 13, No. 3, 304-319.

- Kiesler, Sara and Sproull, Lee S. (1986). "Response Effects in the Electronic Survey," *Public Opinion Quarterly*, 50, 402-413.
- Klingler, D.E., J.H. Johnson, and T.A. Williams (1976). "Strategies in the Evolution of an On-Line Computer-Assisted Unit for Intake Assessment of Mental Health Patients," *Behavior Research Methods and Instrumentation*, 8, 95-100.
- Koson, D., C. Kitchen, M. Kochen, and D. Stodolosky (1970). "Psychological Testing By Computer: Effect on Response Bias," *Educational and Psychological Measurement*, 30, 803-810.
- Liefeld, John P. (1988). "Response Effects in Computer-Administered Questioning," *Journal of Marketing Research*, 25, 405-409.
- Lucas, R.W., P.J. Mullins, C.B. Luna, and D.C. McNroy (1977). "Psychiatrists and a Computer as Interrogators of Patients with Alcohol-Related Illnesses: A Comparison," *British Journal of Psychiatry*, 131, 160-167.
- Merrell, Kenneth W. (1985). "Computer Use in Psychometric Assessment: Evaluating Benefits and Potential Problems," *Computers in Human Services*, 1, No. 3, 59-67.
- Muncy, James A. and Scott J. Vitell (1992). "Consumer Ethics: An Investigation of the Ethical Beliefs of the Final Consumer," *Journal of Business Research*, 24, No. 4, 297-311.
- Newsted, Peter R. (1985). "Paper Versus Online Presentations of Subjective Questionnaires," *International Journal of Man-Machine Studies*, 23, 231-47.
- O'Brien, Terry and Valerie Dugdale (1978). "Questionnaire Administration by Computer," *Journal of the Market Research Society*, 20, 228-237.
- O'Reilly, J.M., M. Hubbard, J. Lessler, P.P. Biemer, and C.F. Turner. (1994) "Audio and Video Computer Assisted Self-Interviewing: Preliminary Tests of New Technologies for Data Collection," *Journal of Official Statistics*, 10, No. 2, 197-214.
- Perlman, Gary (1985). "Electronic Surveys," *Behavior Research Methods, Instruments, and Computers*, 17, No. 2, 203-205.
- Perreault, William D., Jr. (1992). "The Shifting Paradigm in Marketing Research," *Journal of the Academy of Marketing Research*, 20, 367-375.
- Rafael, Joseph E. (1984). "Self-Administered CRT Interviews: Benefits Far Outweigh the Problems," *Marketing News*, 23 (November 9), 16.
- Rotter, J.B. (1966). "Generalized Expectancies for Internal vs. External Control of Reinforcement," *Psychological Monographs*, 80, whole no. 609.
- Scheier, Michael, Jagdish Weintraub and Charles Carver (1986). "Coping with Stress: Divergent Strategies of Optimists and Pessimists," *Journal of Personality and Social Psychology*, 51, No. 6, 1257-1264.
- Skinner, Harvey A., and Barbara A. Allen (1983). "Does the Computer Make a Difference? Computerized Versus Face-to-Face Versus Self-Report Assessment of Alcohol, Drug and Tobacco Use," *Journal of Consulting and Clinical Psychology*, 51, 267-275.
- Turkington, Carol (1984). "The Growing Use, and Abuse, of Computer Testing," *APA Monitor*, 7, 26.
- Waterton, J.J. and Duffy, J.C. (1984). "A Comparison of Computer Interviewing Techniques and Traditional Methods in the Collection of Self-Report Alcohol Consumption Data in a Field Survey," *International Statistical Review*, 2, 173-182.
- Whalen, Bernie (1984). "On-Site Computer Interviews Yields Research Data Instantly," *Marketing News*, 18 (November 9), 1, 17.
- Wood, S. (1984). "Computer Use in Testing and Assessment," *Journal of Counseling and Development*, 63, 177-179.
- Zimmerman, D.P. (1993) "The Psychological Effects of Computers Upon Youth: A Selective Review," *Advances in Social Science and Computers*, 3, 99-138.