

HOW COMPUTERIZED INTERVIEWING ELIMINATES THE SCREEN-IN BIAS OF FOLLOW-UP QUESTIONS

Jon Pinnell, IntelliQuest Inc.
Valentine Appel, Consultant

Abstract

Because computer assisted self-interviewing prevents the respondent from knowing the consequences of screening-in, the result is higher screen-in levels and higher average issue audience estimates.

Virtually all print audience research employs a two-stage approach to estimate audience size. In the first stage, respondents are presented with a list of publications and asked to indicate those they might have read or looked into in the last six months. The process is called screening, and those publications selected are said to have "screened-in." The next stage employs a second question the response to which defines the size of the average issue audience. The average issue audience question is asked for each and every title screened-in. The wording of the question may differ from one supplier to the next.

Additionally, a number of questions are frequently asked to measure the quality of reading. We refer to this series of questions, the frequency of reading and the quality of reading questions, as follow-up questions, in that they follow the six month screen.

The screening process serves two purposes: (1) The first purpose is to permit the respondent to satisfy whatever need there may be to appear well read without necessarily impacting the readership levels which are determined by the answers to the second question for each title screened-in. (2) The second purpose of the screening process is simply to keep the interview to manageable length since only those titles which screen-in are subjected to further questioning.

The process of screening has a major problem associated with it, however: To the extent that the respondent and/or interviewer are aware that screening-in to a particular title will result in additional questions being asked, fewer titles will be screened-in, and readership levels will be lower than if the respondents and/or interviewers are not so aware.

Survey researchers have been aware of this problem for some time, and Sudman and Bradburn (1991) have described it quite well. (page 145) Although these authors specifically mention reports of magazine readership, no references are cited to that effect, and we have not been able to identify any published studies which have specifically studied the effects of follow-up questions upon the screen-in levels generated earlier.

There have been indications of late that the computer may eventually replace the paper and pencil questionnaire using this two stage method of questioning. Computer assisted telephone interviewing (CATI) is already an established method. Computer assisted personal interviewing (CAPI) is under development in the U. K. (Meier and Finch 1993), and in the U. S. MRI (Baim 1994) and others are experimenting with the use of computer assisted self-interviewing (CASI). Michael Weeks (1992) has recently published the results of a detailed literature survey on the subject of computer assisted interviewing, and IntelliQuest is experimenting with the use of computer interviewing via a computer disk sent through the mail.

Computer assisted self-interviewing has a number of advantages over traditional paper and pencil surveys. These include: controlled skip patterns, better ability to randomize presentation order and no need for data entry. Another advantage of computer assisted self-interviewing is that neither the interviewer or the respondent is in a position to exert any influence, however subtle, upon the respondent's screen-in responses in an effort to reduce the interview burden. It is the last point that is of most interest here.

Recent research has demonstrated the importance of the six month screen question, and its impact on audience size (Appel 1993a, 1993b, 1994, Mallett 1993, and Joyce 1992). Using paper and pencil mail questionnaires, unlike self-administered computer-based interviews, the respondent soon learns that each publication screened into results in having to answer several additional questions and will screen into fewer publications as a result.

The purpose of the present article is to explore the magnitude of this effect on both screen-in rates and audience size estimates and to demonstrate how the use of computer assisted self-interviewing can mitigate it. The paper is based on two pilot study controlled experiments which have recently been completed in the U. S.

The First Pilot

The objectives of the first pilot were to compare the response rates, screen-in and rating levels achieved when the mail survey respondent completed the readership questionnaire in the conventional way using paper and pencil and when it was completed using a computer disk sent through the mail. The rating is sometimes referred to as a coverage percentage and defines the size of the average issue audience.

The first pilot was based on a sample of 302 respondents who had been identified by telephone as individuals having influence in the purchase of a computer at work. Over 99% of the influencers had access to a personal computer. Half of the sample so identified was mailed the paper questionnaire, and the other half was mailed the questionnaire on a computer disk.

The paper questionnaire was printed on legal size paper containing 66 black and white magazine and newspaper logos printed down the left side of the page, 22 logos for each of three pages. The top portion of one of the three pages is shown in Figure 1. Note that for each title screening-in three additional questions were asked: frequency of reading, readership of advertising and percent of pages opened.

Figure 1

SECTION 1																	
<p>Directions: This section is about some of the publications that you, yourself, may read or look into. When answering whether you have read or looked into each of the publications, please include any issues you may have read at work, at home, at school or elsewhere, as well as those you happen to glance through. Beginning with the first publication in the list, please fill out the entire row of questions from left to right before moving to the next publication in the list.</p>																	
<p>Questions About Filling Out This Survey Call 1-800-627-8839 ext. 975</p>	<p><i>Have you read or looked into any issue of this publication in the past six months? (If NO, proceed to the next publication.)</i></p>	<p><i>How many issues (of the publications listed below) do you usually read or look into out of every four (or five in the case of daily newspapers) that are published? (Check the appropriate box.)</i></p>	<p><i>How closely do you read or examine the advertising for computer hardware, software, and communications related products and services in this publication? (Check the appropriate box: 1= not at all closely 2= somewhat closely 3= very closely 4= extremely closely.)</i></p>				<p><i>What percentage of the pages do you usually look at or read in the course of your reading a typical issue? (Check the appropriate box: 1= just a few 2= about 25% 3= about 50% 4= about 75% 5= all or most.)</i></p>										
			Less than One	One	Two	Three	Four	Five	1	2	3	4	1	2	3	4	5
(Weekly)	COMMUNICATIONSWEEK <small>The Magazine for Business Electronics</small>	<input type="checkbox"/> YES <input type="checkbox"/> NO	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(18 Times a Year)	Data Communications	<input type="checkbox"/> YES <input type="checkbox"/> NO	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(Monthly)	DEC Professional	<input type="checkbox"/> YES <input type="checkbox"/> NO	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(Twice a Month)	digital news&review	<input type="checkbox"/> YES	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

There were eight additional pages of questions having to do with other media behavior, computer usage and purchase and demography. The questionnaire was saddle stitched and including the cover was 12 pages in length.

The study was done using a single mailing with no follow-up of non-respondents. The response rate was 47% for the disk version and 43% for the paper one - not a significant difference.

From Table 1 it is clear that the disk version produced a screen-in rate and a mean rating which was about 50% higher than that produced by the paper and pencil questionnaire. Both of these differences were statistically significant well beyond the .001 level.

Table 1

First Pilot: Screen-Ins And Ratings

	Disk	Paper
(No. of respondent)	(158) %	(144) %
Mean Screen-in (standard error)	19.5 (1.08)	13.1* (0.81)
Mean Rating (standard error)	9.0 (0.63)	5.7* (0.44)

*Significantly different from the disk ($P < .001$)

The result of this study was that the computer disk-based questionnaire produced higher screen-in and audience levels than did the paper and pencil version. But why? Three hypotheses suggested themselves: (1) physical form, (2) absence of logos and (3) question avoidance.

(1) Physical Form:

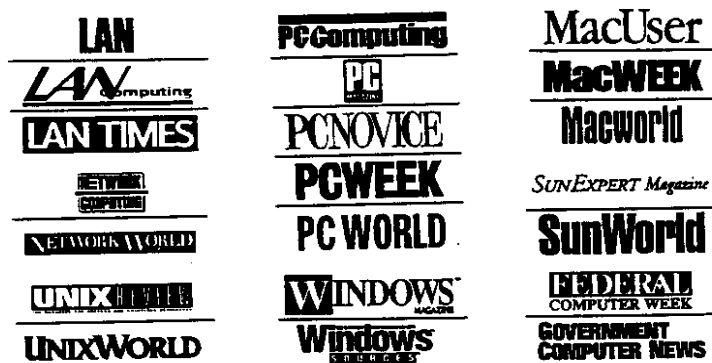
One hypothesis was that perhaps the computer format itself, i.e., the difference in data collection modality, was the cause of the higher screen-in and read levels. This was a realistic possibility, and would cast some doubt on the validity of readership studies using computer assisted self-interviewing.

(2) Absence of logos:

The paper and pencil questionnaire presented the list of publications using their logos, but for the disk-based questionnaire it was not possible at the time to present the logos on the computer screen.

In this regard it should be noted that especially among the trade publications studied, the potential for name confusion was large. Examples of some of the likely to be confused publication titles and their logos, as presented on the paper questionnaire, are shown in Figure 2

Figure 2



For the disk version only the names of the magazines were shown. To partially overcome this problem a printed sheet that contained the logos for all of the publications studied was included in the mail package. The logo sheet was printed on two sides with all of the computer publications' logos on one side and all of the non-trade publications' logos on the other. The logos were produced in black and white and in the same size as on the paper version of the questionnaire. It is not clear, though, whether the device of a separate printed sheet was as effective as actually seeing the logos on the questionnaire, as was possible with the paper and pencil version.

(3) Question Avoidance:

For respondents filling out the paper and pencil questionnaire it was obvious that for each title they claimed they might have read in the past six months, they would have to answer three additional questions which they would not otherwise have to answer. Respondents completing the disk based-questionnaire, however, had no way of knowing this.

For each of the 66 publications the disk version of the questionnaire first asked whether the respondent had read or looked into any issue in the last six months. Then only after all 66 titles had been screened, for those titles that had previously screened-in, were the three additional questions asked.

Perhaps this questionnaire difference was the cause of the difference in screen-in rates -- a conscious or unconscious desire on the part of the paper and pencil respondents to avoid the additional questioning required for those titles screening-in.

The Second Pilot

In order to test these three hypotheses a second pilot was conducted. Although the sample composition of the second pilot was different from the first, resulting in somewhat lower screen-in levels and mean ratings, the second pilot was conducted identically with the first using four, rather than two, different questionnaires. The results of this second pilot along with capsule descriptions of the four questionnaire versions are shown in Table 2.

Table 2
Second Pilot - Screen-Ins And Ratings
Questionnaire Version

	(1)	(2)	(3)	(4)
Format:	Disk	Paper	Paper	Paper
Logos vs. Names:	Names	Names	Logos	Logos
Follow-Up Questions:	None*	None	None	Three
(No. of respondents)	(76)	(94)	(87)	(83)
	%	%	%	%
Mean Screen-in (standard error)	16.6 (1.00)	15.2 (1.12)	13.2** (0.94)	9.6*** (0.65)
Mean Rating (standard error)	7.6 (0.58)	NA "	NA "	4.0*** (0.38)

*None that were obvious until all titles had been screened

**Significantly smaller than cell 1 ($P < .01$, one tailed), not significantly different from cell 2 ($t = 1.37$, $P < .10$, one tailed)

***Significantly fewer than for any of the other cells ($P < .005$, one tailed)

The first and fourth questionnaire versions were the same as had been used in the first pilot. Questionnaire versions 2 and 3 were both paper, and neither asked any readership questions other than the screening question. The three additional questions for each screened-in title were not asked. The difference between versions 2 and 3 was that version 2 identified the publications by name only and version 3 identified the publications using their logos.

The response rates for the four questionnaire versions were: (1) 45%, (2) 52%, (3) 48% and (4) 46%. A chi-square test showed these differences not to be statistically significant.

An analysis of variance of the differences in screen-in levels shown in Table 2 was highly significant ($P < .001$), confirming that differences in questionnaire version was the cause of significant differences in screen-in rate. A comparison of the results obtained with version 1 with those obtained in version 4 - the two questionnaires employed in the first pilot - confirm the findings of the first pilot. The disk version showing names only, and no way for the respondents to know that they would have to answer three additional questions for each title screened-in, produced much higher mean screen-ins and ratings than did the paper version which showed the publication's logos and made clear that three more questions had to be asked for all titles screened in.

Hypothesis #1, Physical Form

Comparing versions 1 and 2 where the only difference that was apparent to the respondent was the physical form of the questionnaire - disk vs paper - there was very little difference in the mean screen-in rate ($t < 1.0$). Since the frequency of reading question was not asked in version 2 no comparison of mean ratings is possible (NA).

Hypothesis #2, Absence of Logos

The absence of logos hypothesis can be tested by comparing version 3 which contained the logos with versions 1 and 2 which did not. None of these three versions made the respondent aware of any additional questions to be asked of all titles screened in. Here the results are somewhat mixed. Version 3 generated a significantly lower mean screen-in than did version 1 suggesting that the use of logos reduced the problem of name confusion, but a comparison of the two paper versions 2 and 3 did not. The two percentage point difference is in the expected direction, however, and the t-ratio (1.37) is approaching the 0.05 level of significance ($p < .10$ -- one tailed).

Hypothesis #3, Question Avoidance

Comparing versions 3 and 4, where the only difference between them was the presence or absence of the these additional questions, it is clear that the presence of the three additional questions had a significant depressing effect on both the screen-in and rating levels. This finding strongly confirms the question avoidance hypothesis.

Reading Frequency

The fact that the higher screen-in rate resulted in a higher mean rating is completely to be expected because each title which the respondent screened-in received a weight equal to the decimal equivalent of the number of issues claimed read out of every four published. Those indicating "less than one" were arbitrarily assigned a weight of 0.10 - the lowest weight allowed.

What might not have been expected was the fact that the percent increase in the mean rating was so large relative to the increase in the mean screen-in percentage. The conventional wisdom (Mallett, 1993) is that changes in screen-in levels caused by changes in procedure tend disproportionately to come from infrequent readers who are less certain of their past six months reading behavior.

If this were true one would expect to find larger percentages claiming to read less than one out of four issues among those completing the questionnaire on computer disk which generated the higher screen-in rate; but this was not the case. As can be seen from table 3, if anything, in both the first pilot and in the second the reverse was found.

Table 3

(Number of responses)	Frequency Of Reading			
	First Pilot		Second Pilot	
	Disk (1981) %	Paper (1242) %	Disk (818) %	Paper (513) %
Less than one in four	25	29	23	31
One in four	28	22	28	25
Two in four	17	17	21	18
Three in four	10	11	11	6
Four in four	21	22	18	20

Discussion

The implication of these findings is that the use of a paper and pencil questionnaire where respondents are aware that claims of past six month readership will require them to answer additional follow-up questions for each title screened-in spuriously lowers the audience estimate. The evidence supporting the conclusion that the paper and pencil estimates are spuriously low, rather than the disk estimates are spuriously high, is the fact that those screening-in via the disk have about the same frequency of reading distribution as do those screening-in via paper and pencil. It is not the less frequent, less certain reader which the paper and pencil questionnaire is more likely to screen out.

Although the use of disk-based mail questionnaires can eliminate this problem, because of the still low level of computer literacy its application using the mail will be limited until such time as computer literacy reaches acceptable levels. The inability to project magazine logos using a mailed disk is at worst a temporary problem which has already been solved using lap top computers carried by door to door interviewers. For this latter application, no computer literacy is required at all.

The present paper adds one more to the rapidly growing number of examples in which changes in questioning procedure, including the use of follow-up questions, have affected screen-in levels which in turn have affected average issue audience estimates. This last phenomenon has implications beyond the use of mail questionnaires and partially explains the activity currently underway exploring the broader use of computer assisted self-interviewing.

References

- Appel, Valentine. "Anatomy of a Magazine Audience Estimate: The ARF Comparability Study Revisited." *Journal of Advertising Research* 33, 1 (1993 a): 11-17.
- "How Changes in Screen-Ins Affect Reads" *Proceedings: Readership Research Symposium 6*, San Francisco (1993b): 117-123.
- "Length of Screening Interval and Print Media Audience Estimates." *Journal of Advertising Research* (1994):
34, 5 (1994) 22-26
- Baim, Julian. "The Testing of Computer Assisted Personal Interviews in Media Research" *Paper presented at the 40th Annual Conference of the Advertising Research Foundation* (1994).
- Joyce, Timothy. "Screening and Reading." *Paper presented at the MPA Seminar on Screening and Audience Levels* (1992).
- Mallet, Daniel T., Jr. "The Relationship of Screen-in Rates and Readership Levels in MRI and SMRB." *Journal of Advertising Research* 33, 1 (1993): 18-22.
- Meier, Erhard and Steven Finch. "CAPI and the NRS" *Proceedings: Readership Research Symposium 6*, San Francisco (1993): 129-137.
- Sudman, Seymore and Norman N. Bradburn. "Asking Questions." *San Francisco* (1991): *Jossey-Bass Publishers*.
- Weeks, M. W. "Computer Assisted Survey Information Collective: A Review of CASIC Methods and Their Implications for Survey Operations." *Journal of Official Statistics* 3, 4 (1992): 443-465

