ON THE RELATIVE THREATS FROM SAMPLING AND NON - SAMPLING ERRORS IN PRINT AUDIENCE RESEARCH

Jon Pinnell and Sherry Englert, MarketVision Research, Ohio, USA

As a standard practice, print audience estimates are accompanied by a measure of the error attributed to sampling – commonly reported in a technical guide as two sigma tolerances. Frequently derived through simulation or replication methods, estimates of sampling error include many other sources of error such as systematic measurement error and bias (Lessler and Kalsbeek 1992). In one sense, this is necessary because the sampling designs and measurements used in print audience estimation are more sophisticated than what the common formulae of sampling theory allow. In another sense, it is unfortunate that systematic measurement errors and biases are blamed on sampling. The purpose of the current work is to report the findings of a controlled experiment which seeks to minimize some sources of measurement error common in print audience estimates.

Background

Print audience research in the U.S. relies on one of two common methods:

- Recent Reading, which is also used in Europe, and
- Frequency of Reading for self-administered paper and pencil questionnaires.

Both methods commonly employ a screening mechanism before asking specific readership questions. The screening mechanism, commonly an interval of six months in the U.S., serves three specific purposes:

- eliminates non-readers of a particular publication, preventing claims of readership,
- reduces the number of questions respondents are required to answer for non-applicable publications, and
- provides an opportunity for socially desirable claims of readership – hopefully getting any obligatory *yes*'s out of the way.

Traditionally, much attention has been focused on the actual "readership" question, and in the U.S. specifically, the difference between Through the Book and Recent Reading (Chook, 1981). In recent years, a great deal of attention has focused on the differences between Recent Reading and Frequency of Reading (Appel 1996, Douglas and Jones 1995, McGlathery 1996) and other assorted methods (Joyce and Napior 1995). Many discussions have centered around periodicity-based differences and concerns about title confusion.

As interest in Through the Book was waning, more attention was focused on the role of the "screening" question (Appel 1993a, Appel 1993b, Joyce 1992, Mallet 1993, Pinnell and Appel 1994, Pinnell and Appel 1995). Unfortunately, relatively little attention has been focused on the administration of the screening question – only its importance has been determined. Since the size of the average issue audience is determined largely from the screen-in rate (and not the actual readership question), title confusion at the screen-in stage is troublesome. Some methods of determining readership levels might be more apt to subsequently eliminate confused titles, but published evidence of that is slight.

Several documented investigations have focused on the screening process. In the 1980s, the British experimented with and adopted "grouped titles" as a modified screening mechanism (Cornish and Meier, Meier). The grouped titles are presented in typeset on a card, and a respondent (informant) indicates reading any publication on the card. Since the titles are arranged by potentially confused titles, respondents are actually screening-in to a category of publications. The card is then turned over to reveal mastheads which are used to screen specific publications. The British, drawing on experience from the Danish Disk, opt to have the initial screen be stimuli neutral – not drawing attention to specific publications with the use of logos.

At the other extreme of the screening spectrum is personal interviewing common in the U.S. The logo for a single publication is shown on a card, and a respondent indicates reading the publication. In this way, respondents evaluate only one title at a time in random order.

In both of these administrations, titles (or groups of titles) for which the respondent screens-in are processed a second time. In the British case, by reversing the card to show mastheads which are used for title specific claims, and in the U.S. case, by sorting the publications by periodicity and asking specific frequency and recency questions.

It is possible, though not entirely clear, that this second stage of processing reduces title confusion. The concern should be more marked in the U.S. system where likely confused titles are potentially never grouped together. Still, eliminating title confusion at the screening level must be more preferred than eliminating confusion at a later stage in the process.

The U.S. and British examples illustrate two of the many issues in determining the administration of the screening question: the stimuli used (logos versus typeset titles) and the order and proximity in which publications are shown (grouped versus random and order rotations).

Both examples above, and many other experiments such as the work of the Canadians with cover reproductions, are in the realm of personal interviewing. Very little research has been published as it relates to self-completed interviews – especially with the Frequency of Reading questionnaire. In many ways the issues are the same, such as stimuli and presentation order. In some ways, though, the issues in the self-completed questionnaire are more critical. In the personal interview, the entire screening phase is completed before any specific readership questions are asked. The directions for the self-completed interview typically instruct respondents to process one title at a time (screen, readership, and qualitative) before proceeding to the next title.

Concerns about title confusion should be heightened because of this single stage processing. We are concerned that the past research (in the realm of a two stage processing) might not be directly translatable to single stage self-completed questionnaires.

Also, in self-completed questionnaires, the publications are typically not grouped. Rather, they are commonly presented in alphabetical order – juxtaposing some titles: PC Magazine, PC World; Men's Health, Men's Fitness; Parents, Parenting; and doing very little for other titles: Travel and Leisure, National Geographic Traveler, Conde Nast Traveler; Bon Appetite, Gourmet, Food and Wine; Cooking Light, Eating Well.

When the titles are grouped, they are generally grouped by periodicity. The practice of grouping by periodicity may be more dangerous than no grouping at all. When grouped by periodicity, some publication groups that would be accidentally presented together, as above, would be split apart. To wit, PC Magazine, a bi-weekly and PC World, a monthly. The other publications listed above would still be grouped or separated as shown – as most measured media are monthlies.

We hypothesize the prospect of random grouping of publications represents the most dangerous practice, followed closely by presenting titles grouped by periodicity. Grouping titles alphabetically has some positive (though accidental) consequences, and finally, we believe least dangerous is the practice of grouping likely confused titles.

While we believe that past research might not be directly translatable to single stage self-completed questionnaires, we would be surprised if findings from a single stage process would not benefit even multiple stage mechanisms. To that end, we investigate the impact of title confusion as it relates to self-completed paper questionnaires.

Method

In any investigation such as this, one is always troubled to find a criterion measure for determining the best method. The best method is the most valid method, but a measure of validity will never be formulated based on self-reported data. However, any measure of validity must be reliable. Reliability refers to a measure's ability to produce the same results under different questioning conditions. One well documented source of error in print audience measures, and survey research in general, is position or order bias (Whitley, Schuman and Presser). Position or order bias decreases the reliability of a measure by adding systematic measurement error. With large samples and multiple versions of the questionnaire, the errors might cancel each other out, but the bias does not go away. That is, the reported number of readers might be correct, but it is likely the wrong readers.

Position or order bias is indicative of title confusion. Therefore, a method of questionnaire administration that reduces the impact of the order in which publications are presented to respondents should have two benefits:

- increased measurement reliability and
- decreased title confusion.

We investigate the stability of results (invariance due to presentation order) from alternative questionnaire administrations.

We approach the analysis of the measurement error somewhat differently than previous works in the area of screener administration and title confusion. It has been common to analyze these data in one of two ways:

- post hoc analyses of audiences estimates or its components relative to circulation, such as readers per copy (RPC) or screens per copy (SPC), or
- repeated measures experiments of claimed readership.

While not wanting to dismiss either approach, we consider a different set of measures to determine the relative goodness of design alternatives.

Assume that we have two presentation orders for each of two methods of questionnaire administration. The resulting screen-in rates for 6 publications are shown in the following table.

Screen-in Rates for 6 Publications
Two Presentation Orders for Each of Two Methods

Publication	Method One A-Z	Z-A	Method Two A-Z	Z-A
A	0.35	0.51	0.25	0.22
B	0.25	0.48	0.15	0.18
C	0.20	0.22	0.08	0.06
D	0.18	0.12	0.05	0.05
E	0.12	0.07	0.03	0.02
F	0.09	0.04	0.02	0.01

These results can be shown graphically for each of the two methods. Two lines have been superimposed on each of the graphs below. In both instances, the steeper line represents the least squares line had the presentation orders produced identical results. The other line (the flatter line in both instances) represents the actual least squares fit between the two presentation orders.

Impact of Presentation Order

Method One

Method Two
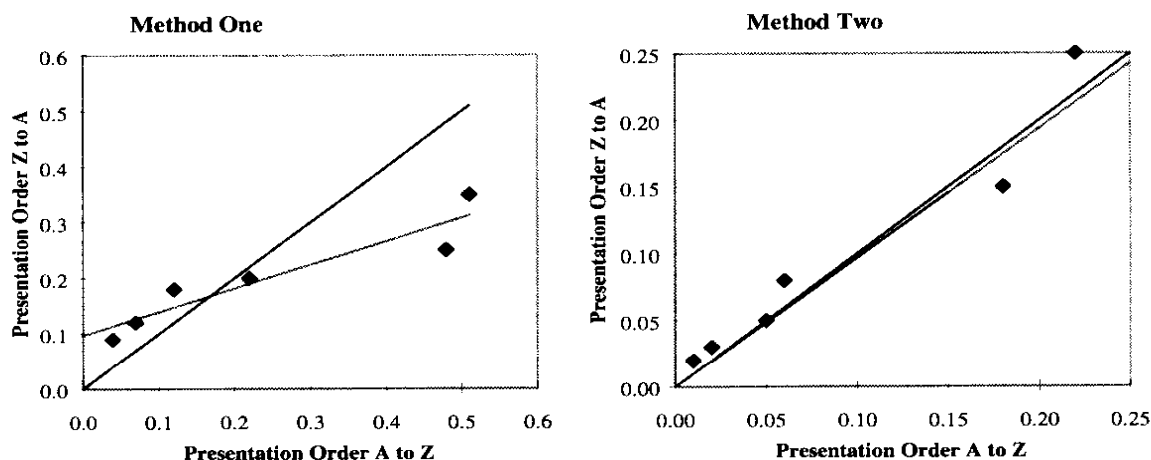
We evaluate specific criteria for the alternative methods. These criteria are:

- Reliability – how closely the two presentation orders' results fall in a straight line,
- Slope and Intercept – how similar the actual results are to a prior expectation of no presentation order effect, and
- Number of Potentially Confused Titles Screened Into – how similar are number of potentially confused titles actually screened into.

Each is discussed below:

Reliability
First, we evaluate the similarity of the screen-in rates between the two different methods. We quantify the similarity of the results based on a reliability correlation that can range from 0.0 to 1.0[1]. The higher the correlation between two methods the more similar the results and the less affected the results are by the presentation order of publications. Recall that a method that is less affected by presentation order should be more preferred.

Correlation assumes a linear relationship between the two methods, so a method that disproportionately impacts large or small publications based on presentation order will have a lower correlation. Also, note that this correlation includes all error, including sampling error. Therefore, we should not seek to interpret variance explained (as is frequently done with squared reliability correlations). The reliabilities would increase with increased sample sizes. Our interest is in the relative differences between the methods.

In the hypothetical example above, we calculate the following reliability correlations:

Reliability Correlations

Method	Reliability
One	0.933
Two	0.973

[1] Though the correlations can technically range from -1 to +1, correlations below zero are unlikely in this application.

In this example, we can see that both methods produce a strong linear relationship between screen-in rates for different presentation orders. From these reliability correlations, though, we can see that method two produces results that are more linearly similar. Put another way, method two produces a more reliable measure – with a 60% reduction in error as shown below:

Method	Maximum Reliability	Actual Reliability	Difference (Error)	Ratio	
One	1.00	0.933	0.067		
Two	1.00	0.973	0.027	40%	(0.027/0.067)

Even though the reliability correlation shows a meaningful reduction in error with method two relative to method one, reliability correlations do not tell the whole story. Reliability correlations express the strength of a linear relationship -- they do not express what the relationship is. In this investigation, not only do we care about the strength of the linear relationship, but we also care about the actual relationship.

Slope and Intercept
Consider the previous example and graphs. It is clear from both the table of screen-in rates and the graphs that method two produces very similar results even with different rotational orders, while method one produces results which vary a great deal with different presentation orders. Recall that the steep line was what we would expect if a method were not affected by presentation order. The other line was the actual relationship between screen-in rates for different presentation orders. If a method is unaffected by presentation order, then it should produce a slope of 1.0 (like the steep line). Note how similar method two's line is to the theoretical expectation and how far method one's line is from our expectation of no presentation order effect.

For a method to produce the same results regardless of presentation order, the intercept must be zero. Title confusion is frequently evidenced by smaller circ publications realizing unrealistically high readers per copy when they appear early in the questionnaire. This title confusion will manifest itself in our investigation as flatter lines that don't go through the origin.
The slope and intercepts for the two hypothetical methods are shown in the following table:

	Method One	Method Two
Slope	0.423	0.970
Intercept	0.097	-0.004

It is important to point out that two effects are impacting the slope. The first effect is title confusion, as discussed above. Second, and maybe less obvious, is respondent fatigue. Whitley (1981) distinguishes between presentation order effect (what we refer to as fatigue) and order effect (our title confusion). That research from the BNRS showed a fatigue effect that benefits early monthly publications and hurts late monthly publications. Therefore, a method that reduces title confusion but not fatigue would be expected to have an intercept of 0 and a slope less than 1.0. A slope of 1.0 would indicate substantial reduction in fatigue effects as well.

Number of Potentially Confused Titles Screened Into
The third criterion to measure if the impact of presentation order and fatigue differs with alternative methods is the number of potentially confused titles that respondents screen into. For example, consider three titles which are likely to be confused: A, B, and C. A method that produces a similar number of these three publications screened into regardless of presentation order is a more reliable method. Specifically, we consider the absolute difference between presentation orders as a percent of the mean of the number in the two presentation orders. As this measure becomes small, we can say a method is less susceptible to presentation order. A measure of zero would indicate no impact from presentation order, but since this measure includes sampling error as well, measures of zero are unlikely. We conduct this analysis separately for groups of likely confused titles.

Empirical Data

The following analyses are based on 446 self-completed personal-interviews completed during the summer of 1997. Respondents were randomly assigned to one of eight questionnaire versions. The questionnaire versions represent all combinations of the following treatments:

- Elicitation Stimuli (logos versus typeset titles)
- Grouping of Titles (titles likely to be confused either shown together or in random order)
- Presentation Order (A to Z and Z to A)

The eight versions can most concretely be thought of as four screening mechanisms, each with two presentation orders. The four mechanisms are:

- Logos/Grouped Titles
- Logos/Random Titles
- Typeset Titles/Grouped Titles
- Typeset Titles/Random Titles

Each of the four screening mechanisms had two presentation orders, with publications appearing early in one version appearing late in the other. The logos were reproduced in black and white, and were judgmentally sized to produce equal prompting. The grouping was also completed judgmentally. In total, 63 publications were measured, most of which were monthlies. Studied titles were consumer in nature, with most being commonly available on the newsstand.

Empirical Findings

We investigate the effect of presentation order using the criteria presented above:

- Reliability – how closely the two presentation orders' results fall in a straight line,
- Slope and Intercept – how similar the actual results are to a prior expectation of no presentation order effect, and
- Number of Potentially Confused Titles Screened Into – how similar are number of potentially confused titles actually screened into.

Reliability

Initially, we evaluate the reliability of these two presentation orders through the reliability correlation. Methods that reduce the impact of the presentation order of magazines should be preferred to methods that are highly dependent on presentation order. That is, methods that reduce order bias are better. The reduction in order bias can be identified by increases in the reliability measure. The reliability measures between the four screening mechanisms are shown below.

Impact of Order of Presentation
Reliability Correlations

Logos/Grouped	Logos/Random	Titles/Grouped	Titles/Random
0.732	0.738	0.689	0.725

Reviewing these correlations, two weak findings emerge.

- Publications grouped by potentially confused titles are *slightly* more affected by presentation order than publications presented in random order.
- Publications presented with logos are *slightly* less affected by presentation order than publications presented by typeset titles.

Neither finding is significant, though the effect for logos is close ($t = 1.91$).

From this, it appears that the Titles/Grouped is the least reliable, but that the others are about equally reliable.

One question that frequently arises when conducting between cell analyses such as these is: How reliable are the data themselves? If we are investigating the reliability of two noisy measures, we might conclude that the two methods produce different results when in fact they produce equal but very noisy results. To explore this further, we can investigate the stability of our initial conclusion that most methods produce equally reliable results.

We conducted within cell tests of reliability. Frequently, tests of reliability such as this are conducted by dividing the respondent sample into replicates and conducting the analysis for each independent replicate – hopefully confirming the same effect in each replicate. Since we are dealing with relatively small sample sizes in this instance, we created replicate measures by randomly dividing the publications into replicates.

Impact of Presentation Order
Reliability Correlations

	Logos/Grouped	Logos/Random	Titles/Grouped	Titles/Random
Replicate 1	0.7576	0.6481	0.6541	0.6439
Replicate 2	0.6192	0.3920	0.4168	0.4782
Replicate 3	0.8255	0.8624	0.7327	0.8792
Replicate 4	0.7993	0.9007	0.8121	0.8881
MEAN	0.7504	0.7008	0.6539	0.7223
STD DEV.	0.0918	0.2340	0.1707	0.1982
MIN	0.6192	0.3920	0.4168	0.4782
MAX	0.7993	0.9007	0.8121	0.8881
RANGE	0.1801	0.5087	0.3953	0.4099

Here we have a finding at odds with the one above. In two ways we would conclude that the most reliable measure (least affected by presentation order) is Logos/Grouped:

- First, the mean reliability across the four replicates is highest for the Logos/Grouped cell.
- Second, the dispersion between replicates is smallest for the Logos/Grouped cell, using either a standard deviation of the mean or the range of observations.

We confirm the finding that Titles/Grouped is the least reliable. We were somewhat surprised that the measure with the least stability (based on dispersion) is Logos/Random – especially given its popularity in The States.

Tests of statistical significance in this application are particularly tricky. Despite the fact that we have an average of only 55 respondents in each cell, we have screener responses for each of 63 publications, producing over 3000 observations per cell. It is not clear, though, that these should be considered independent observations.

Even without tests of statistical significance, the results for Logos/Grouped relative to Logos/Random are striking. The mean reliability of Logos/Grouped reduces the error of Logos/Random by nearly a fifth [83% = (1-.7504)/(1-.7008)]. Also, Logos/Grouped is less than half as variable as Logos/Random.

Slope and Intercept

As discussed above, reliability measured in this way is not a wholly conclusive measure. The reliability correlation measures the relative linear congruence between the two presentation orders but does not indicate whether the two orders are producing the same results. A biased method might produce very reliable but unequal results. For a method to be unbiased we expect the intercept to be 0.0 and the slope to be 1.0. A good method, then, must be both reliable and unbiased.

Slope and intercept estimates are shown in the following table for the four treatment cells.

Impact of Presentation Order
Slope and Intercept

	Logos/Grouped	Logos/Random	Titles/Grouped	Titles/Random
Intercept	0.0195	0.0194*	0.0084	0.0263*
Slope	0.7516**	0.6617**	0.9296	0.6649**

* Significantly different than 0.0
** Significantly different than 1.0

Our results are somewhat mixed. Here we find what would appear to be the method least affected by position bias is Titles/Grouped. We showed above, however that it was the noisiest measure. The next best method would be Logos/Grouped. Interestingly, this suggests that grouping titles removes systematic bias.

Logos/Random and Titles/Random seem to be most affected by presentation order. Both have intercepts larger than zero (indicating beneficial effects to the smallest publications with some rotations) and nearly equal slopes of approximately 0.66.

As with the reliability correlations, we explore this effect over replicates. The slope for each of four replicates is shown below for each treatment.

Impact of Presentation Order
Slope

	Logos/Grouped	Logos/Random	Titles/Grouped	Titles/Random
Replicate 1	0.7501	0.4574	1.2606	0.3446
Replicate 2	0.9665	0.2599	0.6887	0.5677
Replicate 3	1.1478	0.7452	0.9975	1.1638
Replicate 4	0.5593	0.9949	0.8474	0.7367
MEAN	0.8559	0.6144	0.9486	0.7032
MAX	1.1478	0.9949	1.2606	1.1638
MIN	0.5593	0.2599	0.6887	0.3446
RANGE	0.5885	0.7350	0.5719	0.8192

With the replicated measure, we again see both grouped methods perform well and don't see either random method improve.

Once again, the appropriate test of statistical significance is not clear. It is clear, though, that means closer to one are better and smaller ranges are better.

In addition to the results across replicates, we were somewhat concerned about the effect of influential data points on the slope and intercept estimates. In particular, the three largest publications had a sizable impact[2] on all four regression runs. Since extreme observations can unduly influence slope coefficients, we opted to drop all three publications from each of the four models and reestimate the parameters. The resulting estimates are shown in the table below.

Impact of Presentation Order
Slope and Intercept - Excluding Three Largest Publications

	Logos/Grouped	Logos/Random	Titles/Grouped	Titles/Random
Intercept	0.0188	0.0264*	0.0192	0.0323*
(std. error)	(0.0103)	(0.0086)	(0.0111)	(0.0097)
t-ratio	1.82	3.07	1.73	3.33
(parameter = 0)				
Slope	0.7028**	0.5356**	0.6930**	0.5376**
(std. error)	(0.1068)	(0.0989)	(0.1260)	(0.0833)
t-ratio	2.78	4.69	2.43	5.55
(parameter = 1)				

* Significantly different than 0.0
** Significantly different than 1.0

We do not believe it is coincidence that the two grouped methods have intercepts not significantly different than zero and slopes that are nearly equal (and 30% larger than the random methods). We were also quite impressed by the similarity of the slope coefficients within the two grouped methods (.7028 and .6930) and the two random methods (.5356 and .5376).

Number of Screened Publications
Finally, we investigate each of the four treatments' ability to produce similar screen-in rates among likely confused titles with different presentation orders. We divided the 63 studied media into 15 unique groups of likely confused titles. For example, one group included PC Magazine, PC World, and PC Computing. Another group included Food and Wine, Bon Appetit, Cooking Light, Gourmet, and Eating Well.

First, we calculated the difference between the number of publications screened into for the two presentation orders. Then, we divided this difference by the mean of the number of publications screened with the two presentation orders. We refer to this measure as a coefficient of deviation.

$$coefficient\ of\ deviation = \left| \frac{\#screened\ in\ A\ to\ Z - \#screened\ in\ Z\ to\ A}{mean(\#screened\ in\ A\ to\ Z, \#screened\ in\ Z\ to\ A)} \right|$$

Across the 15 groups of publications we calculate a sum of the coefficients of deviation and a standard deviation of the coefficients of deviation. Measures closer to zero or with small standard deviations indicate a method less affected by presentation order. The findings are shown in the following table:

Coefficient of Deviation

	Logos/Grouped	Logos/Random	Titles/Grouped	Titles/Random
Sum	7.01	8.42	11.01	8.34
Std. Dev.	0.30	0.50	0.54	0.50

Examining the sums and standard deviations, Logos/Grouped appears the least affected by presentation order.

---

[2] We interpret impact using the hat matrix and a liberal critical value of 3p/n. Even with this criterion, each of the three largest publications were influential observations in at least two of the four models.

Summary of Empirical Findings

The empirical findings are summarized in the following table.

	Logos/Grouped	Logos/Random	Titles/Grouped	Titles/Random
Reliability				
Average of 4 replicates	0.7504	0.7008	0.6539	0.7223
Dispersion across replicates	0.0918	0.2340	0.1707	0.1982
Slope – All Publications				
Average of 4 replicates	0.8559	0.6144	0.9486	0.7032
Dispersion across replicates	0.5885	0.7350	0.5719	0.8192
Slope – Excluding Influential Obs.	0.7028	0.5356	0.6930	0.5376
Intercept – Excluding Infl. Obs.	0.0188	0.0264	0.0192	0.0323
Coefficient of Deviation				
Sum	7.01	8.42	11.01	8.34
Standard deviation of CDs	0.30	0.50	0.54	0.50

- Titles/Grouped appears to be the least reliable method and produces the largest coefficients of deviation but also appears to be the least impacted by respondent fatigue.

- Logos/Grouped appears to be the most reliable method and also appears to have the most stable reliability (based on dispersion of reliabilities). Logos/Grouped nearly equals Titles/Grouped in terms of susceptibility to respondent fatigue. The two (Logos/Grouped and Titles/Grouped) are equal when influential (large circulation) publications are removed from the analysis. Neither Logos/Grouped nor Titles/Grouped appear to suffer from title confusion (based on intercept) to the extent of the random methods. Also, Logos/Grouped produces the most similar number of potentially confused title screens.

- Logos/Random and Titles/Random have similar reliabilities and appear most affected by title confusion. Both methods have significant intercept terms and very flat slopes.

Discussion

This study, as many do, creates as many questions as answers. While the findings are directionally impressive, our limited sample size prevents broad generalizations. We can conclude, based on this data though, that the method of grouped logos should receive much greater attention for self-completed questionnaires. These findings reproduce well across sample replicates and produce more stable findings than the other methods tested.

As suggested above, we don't understand how a method that would be beneficial in a single-stage self completed questionnaire could be deleterious to a multi-stage interviewer assisted questionnaire. As such, we question the use of monadic logo cards presented in random order, as is common in the U.S. – at least additional research is justified.

References

Appel, Valentine (1996), "Frequency vs. Recency: Frequency of Reading Estimates and Their Relationship to Syndicated Study Estimates." ARF Print Readership Research Workshop, New York.

Appel, Valentine (1993a), "How Changes in Screen-Ins Affect Reads." Worldwide Readership Research Symposium, San Francisco.

Appel, Valentine (1993b), "Anatomy of a Magazine Audience Estimate: The ARF Comparability Study Revisited." Journal of Advertising Research 33, 1 pp. 11-7.

Chook, Paul (1981), "The ARF Comparability Study." Worldwide Readership Research Symposium, New Orleans.

Cornish, Pym and Erhard Meier (1983), "An experiment with grouped titles." Worldwide Readership Research Symposium, Montreal.

Douglas, Steven and Richard Jones (1995), "Frequency of reading for AIA: Under researched?" Worldwide Readership Research Symposium, Berlin.

Joyce, Timothy (1992), "Screening and Reading." Presentation to MPA Seminar on Screening and Audience Levels.

Joyce, Timothy and David Napior (1995), "Alternative uses of magazine covers as prompt aides." Worldwide Readership Research Symposium, Berlin.

Lessler, Judith and William Kalsbeek (1992), Nonsampling Error in Surveys. New York: John Wiley & Sons.

McGlathery, Donald (1996), "Top Line on Frequency of Reading Issues Discussed at the 1995 Worldwide Readership Research Symposium." ARF Print Readership Research Workshop, New York.

Mallet, Daniel (1993), "The Relationship of Screen-in Rates and Readership Levels in MRI and SMRB." Journal of Advertising Research 33, 1 pp.18-22.

Meier, Erhard (1985), "The extended media list – grouped titles, no mastheads." Worldwide Readership Research Symposium, Salzburg.

Pinnell, Jonathan and Valentine Appel (1995), "How computerised interviewing eliminates the screen-in bias of follow-up questions." Worldwide Readership Research Symposium, Berlin.

Pinnell, Jonathan and Valentine Appel (1994), "How Follow-Up Questions Can Affect Print Audience Estimates." ARF Media Research Workshop, New York.

Schuman, Howard and Stanley Presser (1981), Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Content. New York: Academic Press.

Whitley, Edward (1981), "Some rotation effects in the British survey." Worldwide Readership Research Symposium, New Orleans.