

1.7

DEVELOPMENT IN MAGAZINE AUDIENCE MEASUREMENTS IN THE US SINCE MONTREAL

Since the Montreal conference there have been numerous developments and a wide variety of studies in the United States. Work has been completed covering four issues which are at the heart of the past and current Readership Symposia.

- Validity
- Reliability
- Cross media comparisons
- Qualitative measures

In order to present a complete picture of all these activities, I shall comment briefly about those studies which will be reported here in Salzburg, give a synopsis of the findings from studies already published elsewhere and discuss projects under way but not yet completed.

As is painfully well known, in the United States there are two sets of average issue audience estimates which do not come close to agreeing with each other. Each has its advocates and its users. Parenthetically, differences in audience estimates between the two have diminished across time. However, there is still strong sentiment among many Americans that we should determine which audience estimates, if either, reflect actual readership levels for American magazines.

As a result the ARF has provided a vehicle for bringing together representatives of all parties with an important interest in trying to establish actual readership levels.

Previously, as reported in New Orleans, the ARF sponsored some small scale Certitude Studies. The results of these studies showed that underclaiming is a severe problem. Subsequently interest turned to measuring the comparability of the Through-the-Book and the Recent Reading technique. We found they were not comparable. They yielded systematic and non-systematic differences in audience estimates.

Now we have returned to the issue of validity and our current goal is an ambitious one. Our efforts are designed to develop an audience measurement system which we can use as a 'Gold Standard' - a measure against which all other estimates can be compared and judged as correct or incorrect. We felt this task to be sufficiently difficult not to be restricted to approaches which are economically feasible on a syndicated basis.

We have assembled a knowledgeable and unusual group of participants. It included representatives of publishers, research companies and agencies. People from extremely competitive companies such as MRI and SMRB and such as *Time*, *Newsweek* and *US News* have worked together with the rest of us in a highly co-operative effort towards establishing a 'Gold Standard'. To date, we have decided on a measurement model, developed a specific methodology, pilot tested it and designed a sequence of five tests to measure its validity.

In selecting our approach we attempted to address all known sources of measurement error and minimise them.

Here are the nine elements of our method and the reasons for their inclusion.

(1) The first time-read yesterday model was selected to minimise the effects of memory decay between the reading event and its measurement.

(2) A filter question was avoided as numerous studies have shown that it filters out readers. For example the ARF certitude study showed that a filter question used in connection with the Recognition method would have filtered out 12% of the readers.

(3) The Through-the-Book method was selected since extensive research has

shown that retrieval of past events is much more facilitated by recognition than by simple recall.

(4) We are using full-issues with exposure to all items in the magazine. Obviously, this precludes missing readers who happen to have read only parts of the magazine.

(5) The number of titles and issues included in the interview has been limited to about 12 to minimise respondent fatigue.

(6) The titles chosen include pairs of magazines which might well be confused because of similarity in content or appearance. This is the same idea as that behind the grouped titles work completed in the UK and elsewhere.

(7) Multiple issue of each title are included to minimise confusion between readership of different issues of the same publication.

(8) The procedure includes defining readership for the respondent. This should minimise variations in respondents' interpretation of the meaning of the readership question. Readership is defined for respondents by having the interviewer say, "The next question about this particular issue deals with whether or not you have looked into it before now. When I say looking into the issue, this includes reading, looking into or paging through, or opening."

(9) We have avoided asking respondents if "Yesterday, was the first time" etc. This question has a bias in favour of first-time reading - as any 'Yes-No' question has a bias in favour of the 'Yes' response. Similarly we have not asked if the issue has been read before yesterday, as in this case a 'Yes' answer is biased *against* first-time reading. The actual question to determine first-time reading is "Not counting today, on how many different days did you happen to look into this particular issue?" A yesterday reader who has read the issue on one day is

the only one who contributes to average issue audience.

This method, as described, has been tested for workability including: video-taped sessions, a few Belson-type follow-up interrogations; and some actual field trials. On this basis we have a method which is now ready to face the cold reality of validation tests. Most of these tests will contrast observed reading with readership claims.

The grid which follows outlines all possible combinations of observed readership and readership claims. It also shows the effect of each combination on readership estimates.

This grid lists all possible readership claims a respondent could claim to have:

- Read yesterday for the first time or
- Read yesterday not for the first time or
- Not to have read yesterday.

The grid also lists all possible observation results as a respondent/could have:

- Read yesterday for the first time
- Read yesterday not for the first time
- Did not read yesterday but read before
- Did not read yesterday or before.

Under four conditions out of 12, there is complete agreement between observations and readership. These combinations are labelled 'Correct', when for example a first-time reader claims first-time reading; or non-reader claims to be a non-reader.

With this model three incorrect combinations of claims versus observed readership can occur without causing incorrect readership estimates. They are noted in the Grid as 'OK'. For example, if a respondent read yesterday but not for the first time and he failed to recall that reading, we would

Overclaiming/underclaiming analysis grid

Results of observation

Readership claim	Read yesterday		Did not read yesterday	
	First time	Not first time	Read before	Not read before
Read yesterday for first time	Correct	Overclaim	Overclaim	Overclaim
Read yesterday not for first time	Underclaim	Correct	OK	OK
Did not read yesterday	Underclaim	OK	Correct	Correct

Correct = Agreement between observation and claim

OK = Disagreement between observation and claim does not affect readership estimate.

not be missing a first-time reader.

According to the model, underclaiming happens when first-time readers either fail to claim yesterday readership or claim yesterday readership was not for the first time. Overclaiming occurs when a respondent did not read for the first time yesterday but claims he did. This can occur under three conditions: in two circumstances readers who read before can claim first-time readership; and non-readers can claim such readership.

The proposed testing for the Gold Standard method involves testing it against all types of overclaiming and underclaiming as outlined in this grid.

There are five tests in the validation programme:

Two are studies of over and underclaiming of readership of young issues, and

Two are studies of overclaiming based on non-first-time reading being claimed as first-time reading.

Finally there is a study of first-time reading claims of aged issues where we attempted to locate the issues themselves.

Study I measures overclaiming and

underclaiming using pre-publication and post-publication issues among public place readers.

Respondents will be observed reading as they spend time in a public place under completely normal conditions. Interviews conducted one day after the observation will be completed among some respondents who read target, pre-publication issues and among some respondents who did not. Readership of non-pre-publication target magazines will also be observed and included for questioning. The total number of interviews will be 300.

Study II is a mirror image of Study I, except that the observation of reading will be surreptitiously made by the respondent's spouse and will be limited to adults living in subscriber households.

Studies III and IV are studies of overclaiming when time has elapsed between multiple readings of a given issue. They include both readers whose first reading occurs in a public place or at home. Overclaiming occurs if a reader claims first-time readership after the second reading.

For readers whose first reading occurs in a public place, the observation will be made by interviewers. For those whose first reading occurs at home the

observation will be made by the reader's spouse. Both types of readers will be later observed in a waiting-room where they have come to 'participate in a soft drink taste test'. A qualified respondent is one who has been actually observed reading the same issue on two separate occasions. Half of the sample will consist of these whose second reading occurred two weeks after the first observations. For the other half elapsed time between readings will be four weeks.

Study V measures the extent to which respondents who claim first-time yesterday readership of aged issues could have done so based on copy availability. Step one is an actual interview using the 'Gold Standard' method plus detailed questions on place of reading and issue read. Step two includes same day or next day follow-up to determine if issues respondents claim to have read yesterday (for the first time) are or were available at the claimed place of reading.

Total cost for the validation experiments is \$280,000.

We plan to proceed through the five experiments sequentially, with success at each stage being a pre-requisite for proceeding to the next stage.

If all five validation tests are successful, the industry will then be approached for the massive funding required to execute the Gold Standard Method on a large enough scale to obtain meaningful audience estimates against which to compare the results of the two existing, syndicated services.

In Montreal two papers were presented on the results of developmental and initial validation studies using a telephone interview application of the first-time read yesterday technique. *Newsweek* has sponsored some additional validation research on this technique and at the same time provided very useful pre-testing for tests II and V in the ARF validation programme. These

results are reported by Wayne Eadie in Session 2. And in Session 5 he reports on another study which is related to the validity issue. This study examines the extent to which there are significant numbers of magazines in public places - which are older than the age at which they are tested with the Through-the-Book technique.

Under the heading of reliability there have been five studies or developments. Two are covered in Session 2, where Timothy Joyce reports on a comparability study conducted by MRI which contrasts the results of the Recognition & Recent Reading methods when magazine issues ages are varied for the recognition technique, and Val Appel of SMRB reports on an analysis he has done which relates differences in audiences estimates for the two US services to differences in turnover rates.

Other work of SMRB on measuring newspaper audiences was presented at the 1984 ARF Research Quality Conference and is available from the ARF. The study was conducted because measurements using telephone rather than personal interviews were found to generate 15% higher newspaper audience estimates. An experimental study was conducted by selecting respondents who had been interviewed in person two to nine weeks earlier and then re-interviewing the same people by telephone. The experiment uncovered some very major upward shifts in claimed readership levels with the telephone interviews. The much higher readership claims occurred among lower income and lower education segments. SMRB also found these segments claimed higher income and higher education when they were interviewed by telephone than when they were interviewed in person. For this reason, they concluded that the telephone estimates overstated readership. This certainly is one reasonable interpretation of the results and as such represents a caveat to anyone who plans to switch from personal to telephone interviewing in measuring newspaper audiences.

DEVELOPMENTS IN MAGAZINE AUDIENCE MEASUREMENTS IN THE US SINCE MONTREAL

1.7

Prior to 1985, there were two syndicated services measuring newspaper audiences on a market-by-market basis: SMRB and Scarborough. They used identical readership questions and were both audited by the ARF and the Newspaper Research Council. It may still come as a surprise, but they developed very similar audience estimates. For the top 20 newspapers in the US based on circulation, the average readers per copy for both services was 2.9. Clearly here is an instance where we did not have a contractor effect. As a result of the similarities in estimates it was possible for the two companies to join forces and more efficiently serve the newspaper industry.

The final item under reliability is a methodological study to compare audience estimates for farm publications using two different measurement techniques. One approach, currently being used by a syndicated service called FARMS, conducted by Starch, uses a cover recognition technique and mail interviews. The other approach being tested is a full Through-the-Book, personal interview. This work is sponsored by the Agricultural Publishers Association who feel the mail interview using black and white cover reproductions seriously underestimate their audiences. The results of this study should be published early in 1986.

Under the heading of Cross Media Comparisons there have been two studies. One was a study called *Changing channels*. It was designed by Audits & Surveys for the Magazine Publishers Association. This study measured magazine readership and television viewing among a single sample and went beyond audience measurement to measures of advertising exposure.

A single telephone interview was used to gather data on both magazine and television audiences. In the magazine portion of the study, the yesterday recall method was used to determine

readership and MPX for a representative sample of 31 magazines. The questionnaire and estimating techniques were identical to those reported in the MPX paper delivered in Montreal.

For television, audience was measured by the near coincidental method. Programme audience included individuals who were in a room which had a television set on for five or more minutes during the 15 minutes prior to the telephone interview. The audience for commercials is based on individuals who remained in the room with the television set on and did not switch channels during the last commercial break. All interviewing was completed during a 30-day. The total sample was 3,756 adults.

Sampling and weighting to develop universe estimates for yesterday reading involved sampling at-home adults and weighting them to reflect the two variables which determined their probability of selection. One is the number of adults in the household who were at home at the time of the call. The other is the 'nights at home' weight which reflects the probability of finding the respondent at home. Sampling and weighting to develop universe estimates of television viewing involved sampling at-home adults as well as determining the incidence of not-at-home adults.

The key results of the study deal with estimates of advertising exposure. For magazines the measure is MPX, magazine page exposure, which is an estimate of magazine page traffic. This estimate has three components:

$$\begin{array}{r} \text{Average Issue Readers} \\ \times \\ \text{Average reading days per reader} \\ \times \\ \text{Average \% pages opened per reading day} \end{array}$$

In this study, like in the MPX study reported in Montreal, the overall level was 1.7. This means that the average page in the magazines studied is exposed to an average reader 1.7 times.

Television commercial audiences were determined by starting with programme audience and then subtracting those who at commercial time either left the room or switched channels or turned off the set. Half of those who were in the room for part of the commercial were counted towards the commercial audience.

In this study an average prime time network commercial had an audience equal to about 60% of the audience generated by the programme in which it was found. The major loss of audience stems from people leaving the room at the time of a commercial break. This phenomenon alone reduced commercial audience by 30%. The remaining loss comes from channel switchers plus those turning off the set before the commercial is shown.

At the time of the last commercial break

	Average for cable and non-cable homes %
Programme audience	100
Less	
Those who left room	
All the time	21
50% of those leaving part-time	9
Total	30
Those in room at commercial time	70
Less	
Channel switchers	8
Those turning off set	4
Total	12
Equals	
Commercial audience	58

In sum this study showed that in magazines ad page exposure is 1.7 times greater than average issue audience, while for television commercial

exposure is only about 60% of average programme audience.

A recent development may result in American advertisers being able to make joint media or cross media comparisons for magazines and television. SMRB used a diary to measure television viewing among the sample that constitutes the basis for their syndicated magazine measurement service.

The results indicate one important finding. The diary estimates of viewing levels for television are somewhat lower but similar to those generated by the Nielsen Meters - which provide the accepted national estimates of television viewing. In fact, for 66 prime time programmes the correlation between the two sets of estimates is .95. Viewing levels from the two sources differ more markedly after 11pm when diary levels become even lower compared with meter levels. Presumably, some of this difference, if not the majority of it, is due to people going to sleep with the television set on - as the diary measures viewing while the meter measures electric current.

Cross media comparisons Television viewing measured by meters versus dairies

	Average ratings		
	Nielsen Meter	SMRB Diary	Index SMRB/ Nielsen
Weekday daytime	2.85	2.75	98
Weekend daytime	3.29	2.77	84
Evening	9.98	9.31	93

In summary, we have had an active two years - but major work on the validity front is just beginning.