

2.14 The ARF certitude tests

Earlier in this session we were shown the results of an experiment which demonstrated the unusual ability of some readership methods to capture reading which never occurred. In this paper I review the results of some small scale experiments which indicate the extent to which some current readership methods fail to capture reading which *does* occur.

The ARF Certitude Studies came about because of the work of the ARF's Magazine Research Development Council. These investigations were prompted by chronic inconsistencies in syndicated audience estimates for US magazines. Firstly, the reader recognition and the recent reading methods both presume to measure total average issue audience, but often yield very different audience estimates for the same magazine. Furthermore, the two techniques exhibit systematic patterns of differences in estimates for weekly vs monthly publications. Secondly, trend data generated by either method have often shown unexpected audience gains and losses: that is, from time to time both methods have yielded major gains and losses which were much greater than random variations but which were not associated with changes in circulation, competition or any other identifiable variable.

As these situations undermined industry confidence, the ARF began a programme to determine whether or not any current techniques provide accurate estimates of magazine audiences. This of course required establishing a 'Standard of Truth'. The 'Certitude Tests' reviewed here were seen as part of a first step towards developing such a 'Standard of Truth', a standard which in turn could be used to validate current commercial techniques.

The approach chosen to develop such a standard was to conduct a series of small-scale tightly controlled experiments, each exploring our ability to correctly identify magazine readers produced by different kinds of reading situations. The three different reading situations included in the testing to date are: **Test I:** Laboratory Waiting Room Reading (1977). **Test II:** Subscriber at Home Reading (1978) and **Test III:** Natural Waiting Room Reading (1979).

It was decided that the testing would start by dealing with reading which might be difficult to measure. I refer to:

- (a) reading away from home.
- (b) reading of short duration.
- (c) reading occurring primarily among non-subscribers.
- (d) one-time reading.

We anticipated that a technique that could accurately capture reading behaviour which occurred in

this setting should certainly perform well among primary readers and for at home reading in general. Therefore, we conducted **Test I** in a laboratory waiting room.

TEST I: LABORATORY WAITING ROOM

The purpose of this test was to obtain a preliminary assessment of the extent to which the recognition and recall methods accurately capture away-from-home magazine reading of short duration.

In this experiment we had highly controlled magazine exposure conditions. Respondents were recruited ostensibly to participate in a central-location soft drink taste test. When they arrived they were escorted to a special waiting room where for 15 minutes their magazine reading behaviour was surreptitiously observed both by a receptionist in the room and by another observer behind a one-way mirror.

The test magazines located in the waiting room on a table immediately adjacent to the seating area, consisted of prepublication issues of six monthly magazines: *Esquire*; *Harpers*; *Ladies Home Journal*; *Reader's Digest*; *Redbook*; and *Sport*. In this experiment we used these five different testing methods (none being exact duplicates of existing techniques):

Magazine readership measurement techniques

<i>Technique</i>	<i>Elapsed time</i>
Reader recognition	One day
Reader recognition	One week
Reader recognition	Two weeks
Recall	One day
Recall	One week

The sample for each group consisted of approximately 40 respondents who were matched by age, sex and occupation.

In all instances, the respondents were interviewed in their own homes with the appropriate questionnaire and after the appropriate elapsed time between reading and measurement. The interviewers knew nothing, of course, of the prior part of the experiment.

As each of the 208 respondents could have read the six magazines, there were 1,248 reading opportunities. In 272 instances reading occurred and in 976 instances reading did not occur.

2.14 The ARF certitude tests

TABLE 1

<i>What was observed</i>	<i>Number of events</i>	<i>What respondents reported</i>		
		<i>Reading</i>	<i>Non-reading</i>	<i>Total</i>
Reading	272	56%	44%	100%
Non-reading	976	5%	95%	100%
	1248			

Across all techniques, among the 272 reading events only 56% of the cases resulted in respondents claiming they had read the issue in question. In other words, only 56% of all observed reading events were captured by the interview techniques tested, as is shown in **Table 1**.

It appears, however, that the techniques can much more accurately capture non-reading. But this is misleading, for if a magazine is read by 10% of the population and 5% of the non-readers falsely report reading, this error is equivalent to a 45% error among the 10% who actually read the magazine. It should be noted, then, that when a small percentage of non-readers claim to read, this overclaiming inflates audience estimates. (The three studies I report on here do not properly measure overclaiming, for reasons which are outside the scope of this discussion.)

The recognition technique, applied one day after reading occurred, captured 85% of the readings that occurred. For the recognition interviews, conducted one or two weeks later, the technique captured about half of the readers and lost about half.

The recall technique captured 61% of those who read a magazine one day before the interview – while the one day plus the balance of the week method was an unmitigated disaster, capturing only 36% of the observed readings.

TABLE 2
Capture rates by technique

	<i>Observed readings</i>	<i>Capture rate</i>
Recognition		
One day	48	85%
One week	61	49%
Two weeks	56	52%
Recall		
One day	54	61%
One week	53	36%
Total	272	56%

Next we look at some variables which were related to successfully capturing reading. The first one is reading time.

With the Reader Recognition Technique, we found a positive correlation between reading time and capture rates. Not so for the Recall Technique, where the two variables appear unrelated.

As can be seen in **Table 4**, the use of the typical filter question in conjunction with the recognition technique results in losing actual readers.

Of all reported readings (this does not include observed readings which were not reported by respondents) 12% were captured from respondents who earlier in the interview responded negatively to the filter

TABLE 3
Capture rates related to reading time

<i>Reading time</i>	<i>Observed readings</i>	<i>Capture rate</i>
Recognition		
1–9 minutes	73	47%
10–15 minutes	92	72%
Recall		
1–9 minutes	50	50%
10–15 minutes	57	47%

TABLE 4
Capture rates and 'The filter question'

	<i>Reported readings</i>	<i>Answer to 'The filter question'</i>	
		<i>Yes</i>	<i>No</i>
Recognition			
One day	41	93%	7%
One week	30	83%	17%
Two weeks	29	81%	14%
Total	100	88%	12%

2.14 The ARF certitude tests

question; that is, they said they had not "read or looked into" any issue of the magazine in question in the past six or seven months.

TEST II: SUBSCRIBER AT-HOME READING

The purpose of this experiment was to obtain a preliminary assessment of the extent to which the yesterday recall technique can accurately capture the natural, at-home reading of a magazine's subscribers.

In this experiment we not only looked at the yesterday recall technique, but we wanted to evaluate one way of measuring first time readers, as yesterday reading provides an estimate of reading days but not of average issue audience. The addition of 'first time' reading to yesterday reading solves this problem. However, by the time many respondents had read the target publications (two to five days after the magazine arrived at their homes) we could not be sure that the reading observed was in fact 'first time' reading.

Magazine exposure conditions consisted of having male respondents engaged in natural at-home reading of magazines to which the household had a subscription. This reading behaviour was surreptitiously recorded by the male's spouse. The test magazines included three weekly magazines and one monthly magazine: *Newsweek*; *Time*; *US News and World Report*; and *Reader's Digest*.

The magazine readership measurement technique employed was yesterday recall, with a day-part examination of reading behaviour. The cover plus table of contents cues were handed to each respondent. The sample consisted of 46 males living with a spouse in a household subscribing to two or more of the four test magazines.

The yesterday recall procedure involves, first, determination of any magazine reading by day-part. Next, magazine titles are presented for each day-part for which magazine reading was claimed. Finally, covers and table of contents of live issues are shown for titles for which

TABLE 5
Capture rates by number of test magazines read

Number of test magazines	Number of respondents	Number of magazines read	Capture rate
1	41	41	95%
2	5	10	80%
Total	46	51	92%

reading was claimed. In this experiment the capture rate was 92%, as shown in **Table 5**.

However, we found five out of 51 instances where a subscriber was observed to read a magazine one day but claimed not to have read it the very next day.

All respondents who properly claimed to have read the magazine in question also properly reported that the reading took place at home. However, there was much less than complete agreement between respondent and spouse on the amount of time the respondent spent reading or the proportion of pages opened.

TABLE 6
Observed vs reported behaviour

Claimed at home as a place of reading	100%
Level of agreement on time spent reading	68%
Level of agreement on proportion of pages opened	75%

TEST III: NATURAL WAITING ROOM READING

The purpose of this experiment was to obtain a preliminary assessment of the extent to which the recognition method accurately captures natural waiting room reading. The conditions of this experiment were similar to those in the first experiment except that the waiting room conditions were all completely natural instead of experimentally controlled.

Normal magazine reading behaviour was observed while potential respondents were sitting in waiting rooms or waiting areas in barber shops, beauty salons or the offices of doctors or dentists. Four of each kind, or a total of 16 such establishments, were involved in the study.

The test magazines consisted of prepublication issues of five monthly and two weekly magazines. Single-sex publications were appropriately placed only in Barber Shops or Beauty Salons.

<i>The monthlies</i>	<i>The weeklies</i>
Harpers	Newsweek
LHJ	Time
Reader's Digest	
Redbook	
Sport	

The magazine readership measurement techniques tested were reader recognition conducted one day vs one week after exposure.

Each of the two test groups consisted of the sample shown in **Table 7**.

2.14 The ARF certitude tests

TABLE 7

	<i>Men</i>	<i>Women</i>	<i>Total</i>
Barber shops	10	—	10
Beauty shops	—	10	10
Doctor's offices	5	5	10
Dentist's offices	5	5	10
Total	20	20	40

TABLE 8
Capture rates by amount of time spent reading

<i>Observed minutes spent reading</i>	<i>Number of observed readings</i>	<i>Capture rates</i>
Under 10	21	62%
10-19	72	75%
20 and over	34	71%
Total	127	72%

The capture rate among those interviewed one day after reading occurred was 80%, while the capture rate among those interviewed one week after reading occurred was 65%. The average capture rate across both groups was 72%.

In **Table 8** we see that the capture rate was 80% among those who were observed to read over half of the

issue in question, while this drops to 65% among those reading less than half of the pages. It is interesting to note that the capture rate was not strongly associated with reading time once reading time exceeded ten minutes.

The reader recognition technique and the particular first time reading question tested did not work well together, as can be seen in **Table 9**.

CONCLUSION

Our work in the Certitude area was interrupted when events in the market place turned attention and funding to the subject of 'Comparability'.

At that time the entire community raised a prayer saying in effect: 'If we can't have truth at least give us comparability', but as has been reported by many of my countrymen, this was not to be. I trust, then, that our attention will soon again turn towards establishing a measure of truth in this complicated world of magazine audience measurement.

TABLE 9
Reported levels of 'First time' reading with day after reader recognition

	<i>Number</i>	<i>Percentage</i>
Observed readings	59	100
Reported readings	47	80
Reported 'First time' readings	32	54